

**ULTRA GÜVENİLİR DÜŞÜK GECİKMELİ SINIR  
BİLİŞİMDE DİNAMİK GÖREV AKTARIMI VE KAYNAK  
TAHSİSİ**

**DYNAMIC TASK OFFLOADING AND RESOURCE  
ALLOCATION FOR ULTRA-RELIABLE LOW-LATENCY  
EDGE COMPUTING**

**ŞEVKET EFE GÜLLERCİ**

**PROF. DR. EMRE AKTAŞ**

**Tez Danışmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Elektrik ve Elektronik Mühendisliği Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2023

# ÖZET

## ULTRA GÜVENİLİR DÜŞÜK GECİKMELİ SINIR BİLİŞİMDE DİNAMİK GÖREV AKTARIMI VE KAYNAK TAHSİSİ

Şevket Efe GÜLLERCİ

**Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü**

**Tez Danışmanı: Prof. Dr. Emre AKTAŞ**

**Nisan 2023, 45 sayfa**

Çevrimiçi video, artırılmış gerçeklik/sanal gerçeklik (AR/VR) ve nesnelerin interneti (IoT) gibi uygulamalar, yüksek boyutlu çok miktarda verinin gerçek zamanlı olarak işlenmesini gerektirmektedir. Bu veriler kaynak-sınırlı aygıtlarda işlenirken performans ve kullanıcı deneyim kalitesi önemli ölçüde etkilenmektedir. Bu bakımdan mobil sınır bilişim (MEC) sistemleri; hesaplamaların ve trafik denetiminin ağ sınırı dahilinde gerçekleştirilmesine olanak sağlamaktadır. Bununla birlikte ortalama performansa dayalı olarak geliştirilen MEC sistemleri, 5G haberleşmede yüksek öneme sahip ultra güvenilirlik ve düşük gecikme gereksinimlerini sağlamakta başarısız olmaktadır. Bu tezde; C. -F. Liu, M. Bennis, M. Debbah ve H. V. Poor tarafından önerilen bir MEC sistem mimarisi referans alınmıştır. Bu MEC mimarisinde, ultra güvenilir ve düşük gecikmeli haberleşme (URLLC) için ekstrem değer kuramı kullanılarak görev sıralarına uygulanan istatistiksel kısıtlamalar, bununla birlikte haberleşme ağı dahilindeki kullanıcılarda güç tüketiminin minimizasyonu hedefiyle önerilen, yerel hesaplama ve görev aktarımı için kaynakların tahsisi yöntemleri incelenmiştir. Buna ek olarak, kablosuz kanal kalitesinin, sunucuların mevcut yüklerinin ve hesaplama kabiliyetlerinin dikkate alındığı bir kullanıcı-sunucu eşleştirme politikası üzerinde çalışılmıştır. Referans alınan

MEC mimarisi, kullanıcı-sunucu eşleřtirmesi ve dinamik görev aktarımı için iki ölçekten oluşan zaman akışı tanımlamaktadır. Kullanıcı-sunucu eşleřtirmesi uzun/yavaş zaman ölçeğinde; dinamik görev aktarımı ise kısa/hızlı zaman ölçeğinde yürütölmektedir. Tez kapsamında bu yöntemler incelenmiş ve benzetimler yapılmıştır. Benzetim sonuçlarına göre, ultra güvenilir görev hesaplama ve düşük gecikmeli haberleşme performansının garanti edilebileceđi görölmüştür.

**Anahtar Kelimeler:** 5G ve ötesi, mobil sınır bilişim (MEC), ultra güvenilir düşük gecikmeli haberleşme, dinamik görev aktarımı ve kaynak tahsisi, ekstrem değer kuramı

## **ABSTRACT**

### **DYNAMIC TASK OFFLOADING AND RESOURCE ALLOCATION FOR ULTRA-RELIABLE LOW-LATENCY EDGE COMPUTING**

**Şevket Efe GÜLLERCI**

**Master of Science, Department of Electrical and Electronics Engineering**

**Supervisor: Prof. Dr. Emre AKTAŞ**

**April 2023, 45 pages**

Applications such as online video, augmented reality/virtual reality (AR/VR) and Internet of Things (IoT) involve real time computation of high-size and large amount of data. This data have impact on users' quality of experience and system performance while being computed on resource-limited devices. In this regard, mobile edge computing (MEC) systems allow computations and traffic control can be performed at the network edge. However, MEC systems developed with respect to average-based performance fail to meet ultra-reliability and low-latency requirements which are receiving significant attention in 5G communication. In this thesis; C. -F. Liu, M. Bennis, M. Debbah ve H. V. Poor proposed a MEC system architecture in which imposed statistical constraints on the queue lengths with using extreme value theory within the context of ultra-reliable low-latency communication (URLLC) and resource allocation methods for local computation and task offloading to minimize users' power consumption in the network are investigated. Furthermore, a user-server association policy is studied by taking into

account wireless channel quality, servers' existing workloads and computation capabilities. Referenced MEC architecture defines two-scale timeline for user-server association and dynamic task offloading. In this context, user-server association is executed in long/slow timescale; while dynamic task offloading is executed in short/fast timescale. In this thesis, proposed methods within the reference work are investigated and simulations are performed. Simulation results corroborate the efficiency of proposed methods providing ultra-reliable task computation and low-latency communication performance.

**Keywords:** 5G and beyond, mobile edge computing (MEC), ultra-reliable low-latency communication, dynamic task offloading and resource allocation, extreme value theory.

# İÇİNDEKİLER

|   |      |
|---|------|
| ÖZET .....  | i    |
| ABSTRACT.....   | iii  |
| İÇİNDEKİLER .....   | v    |
| ŞEKİLLER DİZİNİ .....   | vii  |
| ÇİZELGELER DİZİNİ .....   | viii |
| SİMGELER VE KISALTMALAR .....                                       | ix   |
| 1. GİRİŞ .....  | 1    |
| 1.1. Tezin Amacı ve Kapsamı.....                                    | 3    |
| 1.2. Tezin Organizasyonu .....                                      | 5    |
| 2. GENEL BİLGİLER .....   | 6    |
| 2.1. MCC ve MEC Sistemleri.....                                     | 6    |
| 2.2. MEC Uygulamaları.....  | 8    |
| 2.3. MEC Hesaplama Modelleri .....                                  | 10   |
| 3. ULTRA GÜVENİLİR DÜŞÜK GECİKMELİ HABERLEŞME .....                 | 12   |
| 3.1. Sistem Modeli .....  | 12   |
| 3.1.1. UE Tarafındaki Trafik Modeli.....                            | 13   |
| 3.1.2. Sunucu Tarafındaki Trafik Modeli .....                       | 14   |
| 3.2. Gereksinimlerin ve Problemin Tanımı .....                      | 16   |
| 3.3. Görev Aktarımı ve Kaynak Tahsisi .....                         | 21   |
| 3.3.1. Lyapunov Optimizasyon Problemi .....                         | 21   |
| 3.3.2. UE-Sunucu Eşleştirmesi, Görev Ayırma ve Kaynak Tahsisi ..... | 25   |
| 3.3.3. Sunucu Tarafında Kaynak Planlama.....                        | 30   |
| 4. NÜMERİK SONUÇLAR .....   | 33   |
| 5. SONUÇ .....  | 40   |
| 6. KAYNAKLAR .....  | 41   |
| EKLER.....  | 44   |

|  |    |
|--|----|
| EK 1 - Tez Çalışması Orjinallik Raporu ..... | 44 |
| ÖZGEÇMİŞ .....                               | 45 |

## ŞEKİLLER DİZİNİ

|  |    |
|--|----|
| Şekil 2.1 MCC ile birlikte kullanılan MEC sistem mimarisi.....     | 6  |
| Şekil 2.2 İnsansız hava aracı destekli MEC ağı modeli.....         | 9  |
| Şekil 2.6 Görev aktarımı tercihlerinin sonuçları.....              | 11 |
| Şekil 3.1 MEC sistem mimarisi ve iki ölçekli zaman akışı [14]..... | 16 |
| Şekil 3.2 Farklı şekil parametreleri için GPD fonksiyonları .....  | 18 |
| Şekil 3.3 Algoritma-1 akış şeması .....                            | 28 |
| Şekil 3.4 Algoritma-2 akış şeması .....                            | 31 |
| Şekil 3.5 Algoritma-3 akış şeması .....                            | 32 |
| Şekil 4.1 MEC sunucularının ve UE'lerin konumları-1 .....          | 33 |
| Şekil 4.2 MEC sunucularının ve UE'lerin konumları-2 .....          | 34 |
| Şekil 4.3 İşlem yoğunluğu ve görev ayırma oranı .....              | 35 |
| Şekil 4.4 Fazlalık değerlerin GPD ile yaklaşımı-1 .....            | 36 |
| Şekil 4.5 GPD şekil ve ölçek parametrelerinin kestirimi-1 .....    | 36 |
| Şekil 4.6 Fazlalık değerlerin GPD ile yaklaşımı-2 .....            | 37 |
| Şekil 4.7 GPD şekil ve ölçek parametrelerinin kestirimi-2 .....    | 37 |
| Şekil 4.8 Alınan sinyal gücüne dayalı eşleştirme-1 .....           | 38 |
| Şekil 4.9 Alınan sinyal gücüne dayalı eşleştirme-2.....            | 39 |
| Şekil 4.10 Alınan sinyal gücüne dayalı eşleştirme-3.....           | 39 |



## ÇİZELGELER DİZİNİ

|  |   |
|--|---|
| Çizelge 2.1 MCC ve MEC karşılaştırması [2] ..... | 7 |
|--|---|

## SİMGELER VE KISALTMALAR

### Simgeler

|              |   |
|--------------|---|
| $Av_i$       | UE i'nin gelen görevleri                                      |
| $Av_i^{Loc}$ | UE i tarafında yerel hesaplama için ayrılmış görevler         |
| $Av_i^{Off}$ | UE i tarafında sunucuya aktarım için ayrılmış görevler        |
| $Av_{unit}$  | Birim görev miktarı   |
| $Q_i^{Loc}$  | UE i'nin yerel hesaplanan görev sırasının uzunluğu            |
| $Q_i^{Off}$  | UE i'nin sunucuya aktarılan görev sırasının uzunluğu          |
| $M_{ji}$     | MEC sunucu j tarafındaki görev sırasının uzunluğu             |
| $u_i^{Loc}$  | Yerel hesaplanan görev sırasının uzunluk sınırı               |
| $u_i^{Off}$  | MEC sunucuya aktarılan görev sırasının uzunluk sınırı         |
| $u_{ji}$     | MEC sunucudaki görev sırasının uzunluk sınırı                 |
| $f_i$        | UE i'nin işlemci frekansı                                     |
| $f_{ji}$     | MEC sunucu j tarafında UE i'ye tahsis edilen işlemci frekansı |
| $f_j^{max}$  | MEC sunucu j'ye ait her işlem biriminin maksimum frekansı     |
| $h_{ij}$     | UE i ve MEC sunucu j arasındaki kanal kazancı                 |
| $W$          | Her bir MEC sunucuya atanan bant genişliği                    |
| $N_0$        | AWGN kanalın spektral güç yoğunluğu                           |
| $\eta_{ij}$  | UE i ve MEC sunucu j arasındaki eşleştirme indikatörü         |
| $\tau$       | Zaman bölümü uzunluğu   |
| $T_0$        | Zaman dilimi uzunluğu   |
| $L_i$        | UE i'nin işlem yoğunluğu                                      |
| $\lambda_i$  | UE i'ye gelen görevlerin ortalama hızı                        |
| $P_i$        | UE i'nin iletim gücü  |

|                       |  |
|-----------------------|--|
| $P_i^{max}$           | UE i'nin güç bütçesi   |
| $\kappa$              | Yerel hesaplama için güç parametresi                               |
| $Cn_j$                | MEC sunucu j'nin işlem birimi sayısı                               |
| $\varepsilon_i^{Loc}$ | UE i'nin yerel hesaplanan görev sırasının sınır ihlali olasılığı   |
| $\varepsilon_i^{Off}$ | UE i'nin sunucuya aktarılan görev sırasının sınır ihlali olasılığı |
| $\varepsilon_{ji}$    | MEC sunucudaki görev sırasının sınır ihlali olasılığı              |
| $t$                   | Zaman bölümü indeksi   |
| $n$                   | Zaman dilimi indeksi   |
| $X_i^{Loc}$           | UE i'nin yerel hesaplanan görev sırasının fazlalık değerleri       |
| $X_i^{Off}$           | UE i'nin sunucuya aktarılan görev sırasının fazlalık değerleri     |
| $X_{ji}$              | MEC sunucudaki görev sırasının fazlalık değerleri                  |
| $\xi_i^{Loc}$         | $X_i^{Loc}$ için GPD şekil parametresi                             |
| $\xi_i^{Loc,th}$      | $X_i^{Loc}$ için GPD şekil parametresinin eşik değeri              |
| $\sigma_i^{Loc}$      | $X_i^{Loc}$ için GPD ölçek parametresi                             |
| $\sigma_i^{Loc,th}$   | $X_i^{Loc}$ için GPD ölçek parametresinin eşik değeri              |
| $\xi_i^{Off}$         | $X_i^{Off}$ için GPD şekil parametresi                             |
| $\xi_i^{Off,th}$      | $X_i^{Off}$ için GPD şekil parametresinin eşik değeri              |
| $\sigma_i^{Off}$      | $X_i^{Off}$ için GPD ölçek parametresi                             |
| $\sigma_i^{Off,th}$   | $X_i^{Off}$ için GPD ölçek parametresinin eşik değeri              |
| $\hat{\xi}_i$         | GPD şekil parametresi kestirim sonucu                              |
| $\hat{\sigma}_i$      | GPD ölçek parametresi kestirim sonucu                              |
| $\xi_i$               | GPD şekil parametresi  |
| $\sigma_i$            | GPD ölçek parametresi  |
| $c_i^M$               | Fazlalık değerlerin deneysel ortalaması                            |

|                   |  |
|-------------------|--|
| $c_i^V$           | Fazlalık değerlerin deneysel ikinci momenti                |
| $R_{ij}$          | UE i ve MEC sunucu j arasındaki iletim hızı                |
| $R_{ij}^{max}$    | UE i'nin MEC sunucu j'ye maksimum aktarım hızı             |
| $K_i^{Loc}$       | $X_i^{Loc}$ rastgele değişkeninin karesi                   |
| $K_i^{Off}$       | $X_i^{Off}$ rastgele değişkeninin karesi                   |
| $K_{ji}$          | $X_{ji}$ rastgele değişkeninin karesi                      |
| $Q_i^{Loc,(X)}$   | $X_i^{Loc}$ değerlerinin sanal görev sırası                |
| $Q_i^{Loc,(K)}$   | $K_i^{Loc}$ değerlerinin sanal görev sırası                |
| $Q_i^{Off,(X)}$   | $X_i^{Off}$ değerlerinin sanal görev sırası                |
| $Q_i^{Off,(K)}$   | $K_i^{Off}$ değerlerinin sanal görev sırası                |
| $Q_{ji}^{(X)}$    | $X_{ji}$ değerlerinin sanal görev sırası                   |
| $Q_{ji}^{(K)}$    | $K_{ji}$ değerlerinin sanal görev sırası                   |
| $Q_i^{Loc,(Q)}$   | $Q_i^{Loc}$ sınır ihlali olasılığı için sanal görev sırası |
| $Q_i^{Off,(Q)}$   | $Q_i^{Off}$ sınır ihlali olasılığı için sanal görev sırası |
| $Q_{ji}^{(M)}$    | $M_{ji}$ sınır ihlali olasılığı için sanal görev sırası    |
| $\varphi_i^{Loc}$ | UE i'nin yerel hesaplanan görev sırasının ağırlığı         |
| $\varphi_i^{Off}$ | UE i'nin sunucuya aktarılan görev sırasının ağırlığı       |
| $\varphi_{ji}$    | MEC sunucudaki görev sırasının ağırlığı                    |
| $\Gamma_i$        | UE i'nin eşleştirme fayda fonksiyonu                       |
| $\Gamma_j$        | MEC sunucu j'nin eşleştirme fayda fonksiyonu               |

## **Kısaltmalar**

|       |   |
|-------|---|
| MCC   | Mobile Cloud Computing                          |
| MEC   | Mobile Edge Computing                           |
| 5G    | Fifth Generation                                |
| ETSI  | European Telecommunications Standards Institute |
| URLLC | Ultra-Reliable Low-Latency Communication        |
| eMMB  | Enhanced Mobile Broadband                       |
| AR/VR | Augmented Reality/Virtual Reality               |
| IoT   | Internet of Things                              |
| UE    | User Equipment                                  |
| CPU   | Central Processing Unit                         |
| AWGN  | Additive White Gaussian Noise                   |
| CCDF  | Complementary Cumulative Distribution Function  |
| CDF   | Cumulative Distribution Function                |
| GPD   | Generalized Pareto Distribution                 |
| DVFS  | Dynamic Voltage – Frequency Scaling             |
| RSS   | Received Signal Strength                        |

# 1. GİRİŞ

Günümüzde kablosuz bağı mobil aygıtların sayısının ve mobil internet trafiğinin katlanarak artmasıyla kablosuz haberleşme ağlarındaki geliştirmeler, kullanıcıların ve uygulamaların gereksinimleri doğrultusunda ilerlemektedir. Özellikle küçük hücresel ağlarda, çok antenli ve milimetre dalga haberleşmelerindeki önemli ilerlemeler sayesinde yeni nesil sistemlerde gigabit hızında kablosuz erişim desteği verilmektedir. Yüksek hızlı ve güvenilir hava arayüzü ile mobil aygıtlara ait verilerin, uzak bulut veri merkezinde işlenebilmesine olanak sağlanmasıyla mobil bulut bilişim (MCC) sistemleri ortaya çıkmıştır [1, 18]. Fakat ağdaki kullanıcılar ve bulut sunucuları arasındaki uzun mesafeler nedeniyle uygulamalar çok uzun gecikmelere maruz kalmaktadır. Bu bağlamda MCC sistemleri, gecikme kritik uygulamalarda yetersiz olarak değerlendirilmektedir [2, 3].

Gelişmiş hesaplama ve depolama yeteneklerine sahip; küçük hücre baz istasyonları, kablosuz erişim noktaları, dizüstü bilgisayarlar, tabletler ve akıllı telefonlar gibi aygıtlar 5G haberleşme ağlarında konumlandırılmaktadır. Çok miktarda hesaplama ve depolama kaynağının ağ sınırında toplanması sayesinde mobil bilişime imkan sağlanmaktadır. Uygulamalar bakımından 5G sistemlerde tek başına bulut bilişim, milisaniye ölçeğinde gecikme performansı gerektiren uygulamaları yürütmekte başarısız olmaktadır. Bu durumda, MCC sistemlerinin mobil sınır bilişim (MEC) ile desteklenerek trafiğin, hesaplamaların ve depolamanın ağ sınırına taşınması gerekmektedir.

MEC kavramı ilk önce ETSI tarafından 2014 yılında “bilişim teknolojisi ve bulut bilişim kabiliyetlerinin mobil abonelere yakın radyo erişim ağı dahilinde sağlanması” tanımıyla yeni bir platform olarak önerilmiştir [19]. Asıl tanımı itibariyle MEC sistemlerinde, kablosuz ağa bağı aygıtlar üzerinde icra edilen uygulamalar tarafından yüklenen görevlerin hesaplanmak üzere daha yüksek veri işleme kabiliyetine sahip sunuculara aktarımı mümkün olmaktadır.

5G haberleşme sistemlerinde; ağ kapsamının ve düğümler arasındaki etkileşimin giderek artması, topolojinin rastgele değişiminden kaynaklanan yüksek seviyede belirsizlik, uygulamaların ve aygıtların arasındaki çeşitlilik nedenleriyle ultra güvenilirlik ve düşük gecikme gereksinimleri önem kazanmıştır. Bugüne kadar, haberleşme ağlarında gecikme ve güvenilirlik yeteri kadar dikkate alınmamıştır [21]. URLLC olarak adlandırılan ultra güvenilir ve düşük gecikmeli haberleşme sistemlerinin tasarımında çok miktarda zorluk ortaya çıkmaktadır. Örneğin, eMMB haberleşmede yüksek güvenilirlik sağlamak üzere

veriler yeniden iletilebilir. Fakat bu çözüm, ağır gecikme gereksinimleri nedeniyle URLLC için geçerli değildir. Buna ek olarak haberleşme ağında gecikme minimizasyonu, iletimde kısa veri paketlerinin kullanımını gerektirmektedir. Bu durum, kanal kodlama kazancının ciddi oranda düşmesine neden olmaktadır [22]. Buna karşın, güvenilirlik performansı sağlamak için ağ üzerinden daha fazla verinin iletilmesi (fazlalık veriler, yeniden iletim) gerekmektedir. Fakat bu durum, gecikmenin artmasına neden olmaktadır. Sonuç olarak, haberleşme ağlarında ultra güvenilirliğin ve düşük gecikmenin birlikte sağlanması ihtiyacı, fiziksel katman yönünden URLLC tasarımını zorlaştırmaktadır [7].

Endüstriyel otomasyon, otonom sürüş ve uzaktan sağlık hizmetleri gibi konularda URLLC haberleşme sistemlerinin kullanımı yüksek potansiyele sahiptir. Bu bağlamda, endüstriyel uygulamalar dahilinde çeşitli uçtan uca gecikme, veri iletim hızı, güvenilirlik ve emniyet gereksinimleri mevcuttur. Örneğin hızlı hareket eden cihazların kontrol devreleri, bir milisaniyeden daha düşük zaman hassasiyeti ile çalışmaktadır [38]. Bu durum, haberleşmenin çok düşük gecikme ile gerçekleştirilmesini zorunlu kılmaktadır. Günümüzde kontrol işlemleri, hızlı kablolu endüstriyel Ethernet arayüzleriyle gerçekleştirilmektedir. Gelecekte, bu kablolu sistemler üretimde esnekliği sağlamak üzere kablosuz bağlantılar ile değiştirilecektir. Endüstriyel sistemlerdeki bu geliştirme, yüksek güvenilirlik ve düşük uçtan uca gecikme gerektirmektedir [39]. Tamamen otonom sürüş uygulamalarında URLLC haberleşme sayesinde, taşıtlar arasında veya taşıtlar ile yapılar arasında haberleşme yoluyla trafik kazalarında ve trafik sıkışıklıklarında önemli ve sürdürülebilir bir azalma sağlanabilmektedir. URLLC haberleşmede sağlanan bir milisaniye düzeyinde uçtan uca gecikme, otonom sürüş manevralarının hızlanmasına ve böylelikle sürüş emniyeti kapsamında çarpışma önlemeye destek olmaktadır. Otonom sürüş uygulamalarının, trafik davranışlarını bütünüyle değiştirmesi beklenmektedir. Özellikle, trafikte taşıtlar arası mesafelerin azalması nedeniyle potansiyel emniyet kritik olayların insan sürücülere kıyasla otonom taşıtlar tarafından çok daha hızlı tespit edilebilmesi için gelecekteki haberleşme sistemlerinin ultra güvenilir ve düşük gecikmeli olarak çalışması gerekmektedir. URLLC haberleşme sistemlerinin başka bir potansiyel kullanım alanı uzaktan sağlık hizmetleridir. Gelişmiş araçlar uzaktan kontrol edilerek muayene, cerrahi ve rehabilitasyon işlemlerinin doktorların konumundan bağımsız olarak icra edilebilmesi mümkündür [40]. Muayene ve cerrahi uygulamalarda hastanın konumunda bulunan medikal ekipmanların, doktorlar tarafından uzaktan kontrol edilmesi yoluyla görsel ve işitsel bilgilerin yanı sıra dokunsal geri bildirimlerin alınmasına olanak

sağlanmaktadır. Hastaların hareketleri uzaktan yönlendirilerek ve kontrol edilerek rehabilitasyon hizmetinin uzaktan verilmesi mümkün olmaktadır. Sağlık alanlarında uzaktan verilen hizmetlerin yüksek doğruluk ve hassasiyet ile gerçekleştirilebilmesi için URLLC tabanlı haberleşme sistemlerinin kullanılması gerekmektedir.

### **1.1. Tezin Amacı ve Kapsamı**

Geleneksel haberleşme sistemleri, daha yüksek ağ kapasitesine ulaşmak üzere tasarlanmıştır ve bu sistemlerde güvenilirlik ve gecikme performansına yeterli önem verilmiştir. 5G haberleşmenin önemli bileşenlerinden biri olan URLLC konusu, hem akademi hem de endüstride büyük ölçüde ilgi görmektedir [7-9]. Mevcut MEC çalışmalarının büyük çoğunluğunda ortalama gecikme, performans ölçütü veya servis kalitesi gereksinimi olarak kabul edilerek görev sıralarının ortalama uzunluklarının sınırlandırılması amaçlanmıştır [10-13]. Fakat yalnızca ortalama esaslı performansa dayanan sistem tasarımları, düşük gecikme ve yüksek güvenilirlik gereksinimlerini karşılamayı garanti edemediğinden URLLC uygulamalarında yetersiz kalmaktadır.

Bu tezde, C. -F. Liu, M. Bennis, M. Debbah ve H. V. Poor tarafından önerilen, görev sıralarının ekstrem uzunluk değerlerinin hesaba katıldığı URLLC merkezli görev aktarımı ve kaynak tahsisi şeması referans alınmıştır. Bu yayında, çok sayıda ve ayrışık hesaplama kaynaklarına sahip MEC sunucularının kullanıldığı çok kullanıcı bir MEC mimarisi geliştirilmiştir [14]. UE'lerin sınırlı hesaplama kabiliyeti ve görev aktarımı sırasında meydana gelen ilave gecikme nedenleriyle gecikme toleransı düşük olan uygulamalarda, UE'ler görevlerini yerel hesaplamak ve kablosuz ortamdan sunucuya aktarmak üzere ayırmalı ve buna uygun olarak kaynaklarını tahsis etmelidir.

Hesaplama öncesinde aktarılan veri miktarı ile karşılaştırıldığında, hesaplama sonucunda elde edilen verilerin boyutu genellikle daha küçük olmaktadır [25, 26]. Bu nedenle incelenen sistem modelinde [14], hesaplanan görevlerin sunuculardan indirilmesi ihmal edilerek görevlerin yalnızca UE'lerden sunuculara aktarılması dikkate alınmaktadır.

Görev sıralarının uzunlukları, gecikme ile alakalı olduğundan referans çalışmada [14] gecikme ölçüsü olarak görev sıralarının değerleri esas alınmıştır. Ortalama gecikmeye dayalı tasarımların eksiklerinden hareketle görev sıraları için eşik değer tanımlanmıştır. Bu kapsamda, UE'lere gelen görev miktarının eşik değeri ihlal etme olasılığı ve ekstrem



değer kuramı [15] kullanılarak modellenen eşik değerin aşılması durumu, URLLC kısıtlamaları olarak hesaba katılmıştır.

Bu kısıtlamalar üzerinde referans çalışmada [14] verilen esas problem; görevlerin yerel hesaplanması ve aktarımı için ağ genelinde güç minimizasyonu olarak tanımlanmıştır. UE'ler ile sunucuların eşleştirilmesi ve görevlerin hesaplanması iki zaman ölçeğinde yürütülmüştür. Görev sıralarının durum bilgisi, sunucuların hesaplama kabiliyetleri ve mevcut iş yükleri, kanallar arası girişim ve URLLC kısıtlamaları hesaba katılarak ve eşleştirme kuramının [16] yardımıyla uzun zaman ölçeğinde UE-sunucu eşleştirilmesi yapılmıştır. Eşleştirmenin ardından görev aktarımı ve kaynak tahsisi problemleri, Lyapunov optimizasyonu [17] kullanılarak kısa zaman ölçeğinde çözülmüştür.

Bu tez çalışmasında öncelikle MEC sistemlerinde URLLC probleminin matematiksel olarak modellenmesi ve formülasyonu mevcuttur. Problemin ve buna ilişkin formülasyonun incelemesini takiben, referans çalışmada [14] önerilen çözüm üzerine odaklanılmıştır. Burada verilen çoklu kaynaklara sahip sunucu modelleri, UE'nin güç tüketimi ve görev trafik modelleri incelenmiştir. Bunu takiben güvenilirlik ve gecikme gereksinimleri ve bunların probleme dahil edilmesi konusu ele alınmıştır. Bu problemin çözümü kapsamında görev sıralarının ekstrem uzunluk değerlerinin GPD ile yaklaşımı ve çoktan bire eşleştirme yöntemi üzerinde çalışılmıştır. Tanımlanan ana problemin üç probleme ayrıldığı ve her problemin bağımsız olarak çözüldüğü görülmüştür. Birinci problem, UE-sunucu eşleştirmesini ve UE tarafında kaynak tahsisini içermektedir. İkinci problem, UE'lerde yerel hesaplama ve sunucuya aktarım için görevlerin ayrılmasını ele almaktadır. Üçüncü problem ise, UE'lerden aktarılan görevlerinin ağırlığına göre MEC sunucularında hesaplama kaynaklarının tahsis edilmesi konusuna odaklanmaktadır. Bu tez kapsamında, referans çalışmada [14] önerilen her bir problem, çözümleri ile birlikte incelenmiştir ve şemalar halinde sunulmuştur. Buna ek olarak, nümerik benzetim çalışmaları yapılarak GPD yaklaşımının gerçekçiliğinin ortaya çıkarılmasının detayları incelenmiştir ve bu yaklaşımın gerçekçiliği doğrulanmıştır. Benzetim ortamında UE'ler, sunucular ve görevlerin trafiği benzetilmiş ve referans çalışmada [14] verilen üç algoritmanın çalıştığı tez kapsamında doğrulanmıştır.

## 1.2. Tezin Organizasyonu

Bu tezin organizasyonu ařađıda verilmiřtir:

İkinci bölümde, ilk olarak mobil biliřim alıřmaları kapsamında MCC ve MEC sistemlerinin geliřimi hakkında genel bilgiler verilerek bu sistemler; fiziksel sunucu, iletim mesafesi, sistem mimarisi ve uygulamalar bakımından karřılařtırılmıřtır. Daha sonra, MEC sistemlerinde hesaplama modelleri tanımlanmıřtır. Bu bölümde son olarak, MEC platformunun uygulamalar yönünden performansı ve avantajları açıklanmıřtır.

Üüncü bölümde ilk olarak üzerinde alıřılan MEC sistemi, UE ve sunucu tarafındaki trafik modelleri verilmiřtir. Daha sonra gecikme ve güvenilirlik gereksinimlerini karřılamak üzere görev sıralarına olasılıksal kısıtlamalar uygulanarak tanımlanan esas problem ele alınmıřtır. Bu bölümün sonunda, görev aktarımı ve kaynak tahsisi kapsamında Lyapunov sürüklenme artı ceza ifadesi biçiminde türetilen esas problem için önerilen özüm algoritmaları incelenmiřtir.

Dördüncü bölümde, benzetim yoluyla ulařılan nümerik sonuçlar verilmiřtir. Benzetim adımlarında ilk olarak UE'lerin ve MEC sunucularının konumları belirlenmiřtir. Daha sonra farklı hızlarda gelen birim görevlerin artan veri iřlem yoğunluđuna karřı yerel hesaplama ve MEC sunucularına aktarım için görevlerin ayrılma oranı hesaplanmıřtır. Bu bölümde son olarak, sırasıyla görev sıralarının kuyruk dađılımları için önerilen yaklařım dođrulanmıřtır ve alınan sinyal gücüne dayalı olarak icra edilen UE-sunucu eřleřtirmesinin etkinliđi deđerlendirilmiřtir.

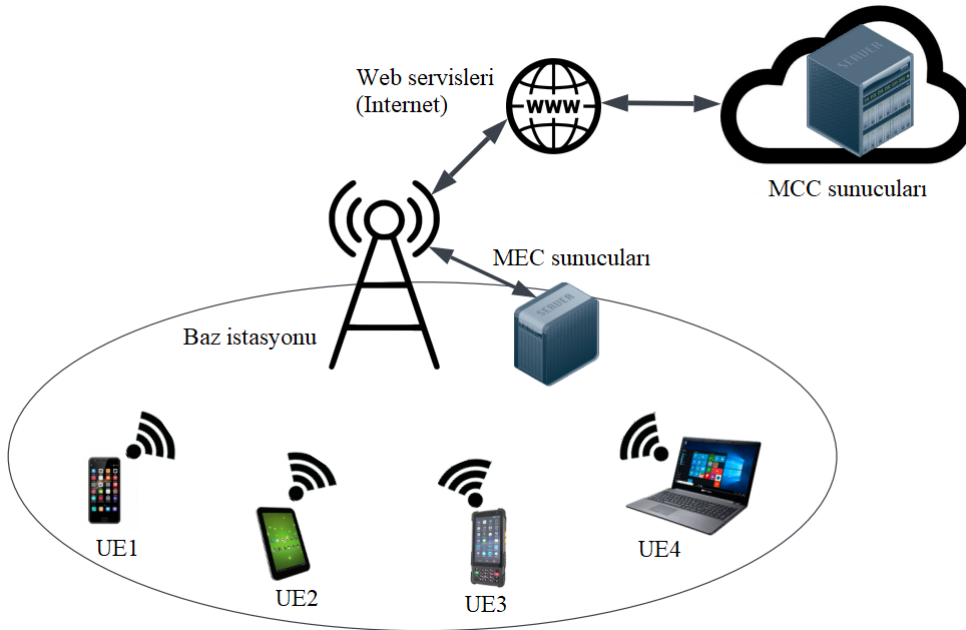
Son olarak beřinci bölümde, incelenen yöntemler ve benzetim sonuçları yorumlanarak gelecek alıřmalara iliřkin görüřler verilmiřtir.

## 2. GENEL BİLGİLER

### 2.1. MCC ve MEC Sistemleri

Mobil bulut bilişim olarak adlandırılan MCC, mobil kullanıcıları bulut bilişim ağına dahil edilmesiyle hesaplama ve depolama servislerinin sunucular üzerinde gerçekleştirilmesine olanak sağlamaktadır. Bu sayede, mobil aygıtların hesaplama ve depolama kabiliyetleri geliştirilmektedir [1-3]. MCC sistem tasarımında, mobil aygıtların konum ve zaman bağımsız olarak web servisleri üzerinden sunucuların kaynaklarını kullanabilmesi hedeflenmiştir. Mobil aygıtlar, yakındaki baz istasyonu ya da erişim noktası üzerinden web servislerine bağlanmaktadır. Web servisleri, mobil aygıtlar ile bulut arasında bir uygulama programlama arayüzü gibi davranarak bulut üzerinde yürütülen uygulamaların mobil aygıtlara ulaştırılmasını sağlamaktadır. Fakat MCC sistemlerinde, uzun iletim mesafeleri nedeniyle uygulamalar ilave gecikmeye maruz kalmaktadır ve bu sistemler gecikme açısından kritik uygulamaların zamanlama kısıtlamalarını karşılamakta başarısız olarak değerlendirilmektedir.

MEC sistemlerinde bulut bilişim servisleri, radyo erişim ağı dahilinde ve mobil aygıtlara yakın konumda sağlanmaktadır. MEC düşüncesiyle arka uç ağda tıkanıklığın azaltılması, düşük gecikme gereksinimi olan uygulamaların desteklenmesi ve kullanıcı deneyiminin geliştirilmesi amaçlanmıştır. Şekil 2.1’de MCC ile birlikte kullanılan MEC sistem mimarisi verilmiştir.



Şekil 2.1 MCC ile birlikte kullanılan MEC sistem mimarisi

MCC sistemlerinde fiziksel sunucular, geniş ölçekte ve büyük yapılardan oluşturulmuş veri merkezlerinde konumlandırılmaktadır. Bu yapılar, yeterli miktarda güç kaynağı ve soğutma ekipmanı ile donatılmıştır. MCC sunucuları, çok yüksek hesaplama ve depolama kabiliyetlerine sahiptir. MEC sistemlerinde fiziksel sunucular, MCC sistemlerine göre daha kısıtlı kabiliyetlere sahiptir.

MCC sunucuları ve mobil aygıtlar arasındaki iletim mesafeleri binlerce kilometreye kadar farklılık gösterebilmektedir ve hatta bu mesafeler, farklı ülkeleri kapsayabilmektedir. Diğer taraftan, MEC sunucuları ve mobil aygıtlar arasındaki mesafeler kısa olmaktadır. MEC sistemlerinde iletim mesafeleri, on ile yüzlerce metre ölçeğinde tanımlanmıştır.

MCC sistemleri genellikle Google ve Amazon gibi bilişim teknolojisi şirketleri tarafından kurulmaktadır ve sistem mimarileri yönünden karmaşık ve yüksek oranda merkezi olarak tasarlanmıştır. MEC sunucuları ise Telekom işletmeleri tarafından kurulmaktadır. Bu sunucular ağ dahilinde yoğun olarak dağıtılmıştır ve basit bir konfigürasyona sahiptir.

MCC sistemlerinde uygulamalar belirli oranda gecikmeye toleranslı olmaktadır. Fakat bu uygulamalar yüksek miktarda hesaplama kaynağı gerektirmektedir. MEC sistemleri, hesap yoğun ve gecikme duyarlı uygulamaları icra etmek üzere tasarlanmıştır. MCC ve MEC arasındaki temel farklar Çizelge 2.1’de verilmiştir.

|                        | <b>MCC</b>  | <b>MEC</b>   |
|------------------------|---|--|
| <b>Fiziksel sunucu</b> | Çok gelişmiş hesaplama ve depolama kabiliyetlerine sahiptir, geniş alanda veri merkezlerinde konumlandırılmıştır. | MCC sunucularına göre daha sınırlı kabiliyetlere sahiptir, mobil kullanıcılara daha yakın konumlandırılmıştır. |
| <b>İletim mesafesi</b> | Kullanıcılara 1-1000+ kilometre mesafede bulunmaktadır.   | Kullanıcılara 10-100+ metre mesafede bulunmaktadır.  |
| <b>Sistem mimarisi</b> | Karmaşık konfigürasyona sahiptir, servisler yüksek oranda merkezi olarak sağlanmaktadır.                          | Basit konfigürasyona sahiptir, servisler ağ sınırında dağıtılmış olarak sağlanmaktadır.                        |
| <b>Uygulamalar</b>     | Gecikme toleranslı ve hesap yoğunluğu yüksek uygulamalarda kullanılmaktadır.                                      | Gecikme duyarlı ve hesap yoğunluğu yüksek uygulamalarda kullanılmaktadır.                                      |

Çizelge 2.1 MCC ve MEC karşılaştırması [2]

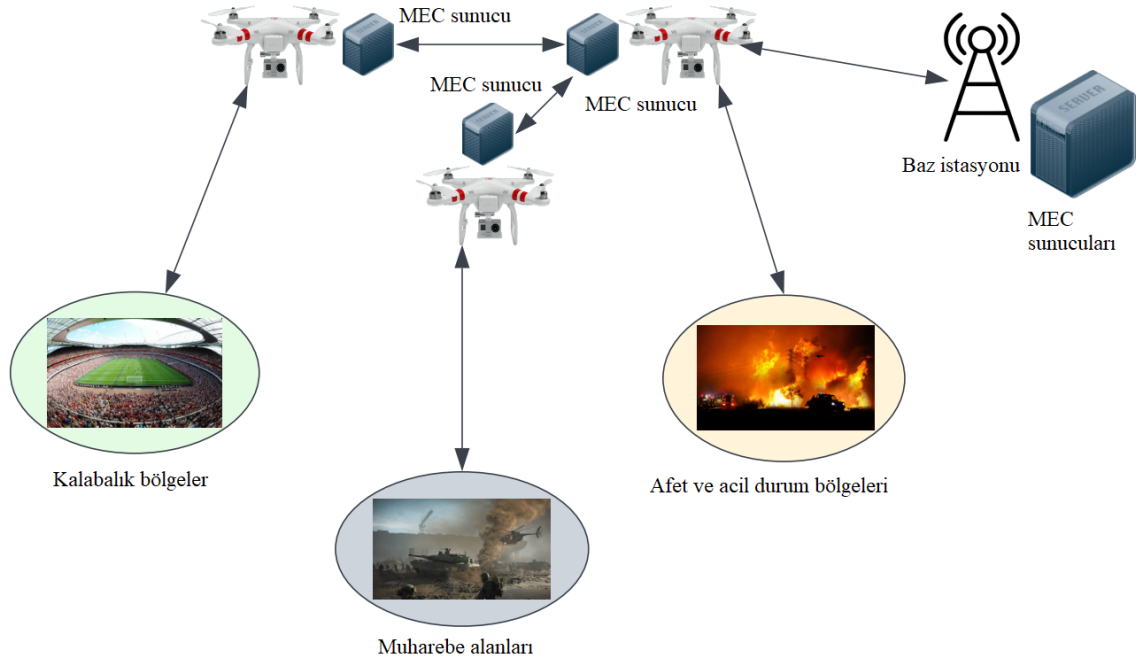
## 2.2. MEC Uygulamaları

AR ve VR uygulamaları; ses, görüntü ve konum girdilerinden oluşturulan verilere dayanan sanal dünya ile kullanıcıların daha doğal etkileşimini mümkün kılmaktadır. AR/VR uygulamaları, kullanıcıların gerçek zamanlı konum ve yön bilgilerine ihtiyaç duymaktadır. Bu uygulamalar, daha iyi kullanıcı deneyim kalitesi sağlanabilmesi adına düşük gecikme ve yoğun veri işleme gerektirmektedir [4]. Bir VR denetim aygıtının MEC sunucusunda uygulanması sayesinde görüntü ve ses verileri kablosuz bağlantı üzerinden kullanıcılara iletilerek bir milisaniye gecikme ve yüksek güvenilirlik ile artırılmış takip tutarlılığı sağlanmaktadır [23]. AR uygulamalarında, hesap yoğun görevler MEC sunucularına aktarılarak aygıtların hesaplama kapasitesi ve pil ömrü artırılmaktadır [24]. Çevrimiçi video uygulamalarında hesaplama yoğunluğu ve kullanılan bant genişliği çok yüksektir. Bu nedenle, video verilerinin mobil aygıtlarda işlenmesi veya bulut sunucularına aktarılması uygulama performansını düşürmektedir. Video verilerinin MEC sunucuları tarafından işlenmesi sayesinde gecikme ve ağdaki trafik azaltılmaktadır [5].

Ağ dahilinde çeşitli görevleri gerçekleştiren çok sayıda ve farklı özelliklere sahip aygıtın dahil olduğu IoT uygulamalarında, hesap yoğun görevlerin IoT aygıtları tarafından MEC sunucularına aktarılması ile bu görevlerin yönetimi ve yürütülmesi işlemleri sunucular tarafından gerçekleştirilmektedir. Bu sayede, gecikme duyarlı uygulamalarda zorunlu olan düşük gecikme gereksinimi karşılanmaktadır. Buna ek olarak, IoT uygulamalarında kullanılan aygıtlar genellikle sınırlı pil kapasitesine sahiptir ve mevcut durumda sürekli ve hızlı bir biçimde pillerinin tekrar doldurulması mümkün değildir. Aygıtlar tarafından MEC sunucularına aktarılan hesap yoğun görevlerin sunuculara icra edilmesi sayesinde güç tüketiminin düşürülmesi sağlanarak ağdaki enerji sorunları çözülmektedir [6].

Muharebe alanları, afet ve acil durum bölgeleri gibi gecikme hassasiyeti yüksek ve görev kritik uygulamalarda karasal ağlar güvenli bir şekilde kurulamamaktadır. İnsansız hava araçları tarafından desteklenen MEC ağlarında; bütünüyle kontrol edilebilir mobilite, sunucuların esnek dağıtımı ve yer cihazları ile güçlü görüş hattı kanalları gibi avantajlar sağlanmaktadır [27]. Bir afet bölgesinde, karasal haberleşme altyapılarının zarar görmesi durumunda görev kritik uygulamaları icra etmek mümkün olmamaktadır. Bu durumda, MEC sunucuları ile donatılmış insansız hava araçları sayesinde görev kritik uygulamaları yürütmek mümkün olmaktadır. Buna ek olarak, çok miktarda mobil aygıtın bulunduğu kalabalık bölgelerde mobil aygıtlar tarafından sunuculara aktarılan hesap yoğun görevler

dolayısıyla yerleşik sunucuların hesaplama kaynakları tükenmektedir ve bu durum uygulamalarda hesaplama gecikmelerinin artmasına ve kullanıcıların deneyim kalitesinin düşmesine neden olmaktadır. Bu tür uygulamalarda, insansız hava araçlarının desteği sayesinde gecikme performansı geliştirilerek kullanıcıların deneyim kalitesinin artması sağlanmaktadır [28]. Şekil 2.5'te insansız hava aracı destekli MEC ağı modeli verilmiştir.



Şekil 2.2 İnsansız hava aracı destekli MEC ağı modeli

MEC sistemleri aracılığıyla verilerin ve uygulamaların bağlı taşıtlara yakın olarak konumlandırılması sayesinde haberleşme gecikmesinin azaltılması sağlanmaktadır. MEC sunucuları, taşıtlarda yürütülen uygulamalardan ve yol kenarındaki sensörlerden iletilen verileri işleyerek gecikme duyarlı mesajların (örneğin, kaza ve tehlike uyarıları) mevcut alanda bulunan diğer taşıtlara çok düşük gecikme ile gönderilmesini sağlamaktadır. Bu durum, taşıtlarda verilerin milisaniye ölçeğinde alınmasına ve sürücülerin hemen tepki vermesine olanak tanımaktadır [4]. MEC sistemlerinin taşıtlar arası haberleşmede kullanımı kapsamında otonom taşıtlar, bir diğer uygulama olarak verilmektedir. MEC sunucuları tarafından, planlanmış veya güncel yol durum bilgilerinin işlenerek bölgeye özel olarak otonom taşıtlara gönderilmesi sayesinde bu taşıtlar, bir hedef istikamete gitmek için kullanılacak en hızlı ve en kısa yolu bulabilmektedir. MEC uygulamalarının yardımıyla, yol durum bilgilerinin taşıtlara çok düşük gecikmelerle gönderilmesi ile otonom taşıtların toplam farkındalığının artması sağlanmaktadır [36].

### 2.3. MEC Hesaplama Modelleri

Mobil aygıtlar tarafından gerçekleştirilen görev aktarımı; hesap yoğun ve gecikme duyarlı uygulamalarda oluşturulan görevlerin, kaynak bakımından zengin sınır bilişim sunucularına aktarılması esasına dayanmaktadır. Bu yaklaşım sayesinde mobil aygıtların pil ömrünün uzaması ve veri işleme kaynaklı gecikmenin düşürülmesi sağlanmaktadır. Görev aktarımı tercihleri, aktarılacak görev miktarının belirlenmesi ve aktarım yapılacak sunucunun seçimi görev aktarımındaki ana problemlerdir. Temel olarak, görev aktarımı kapsamında alınabilecek üç tipte karar mevcuttur [3]:

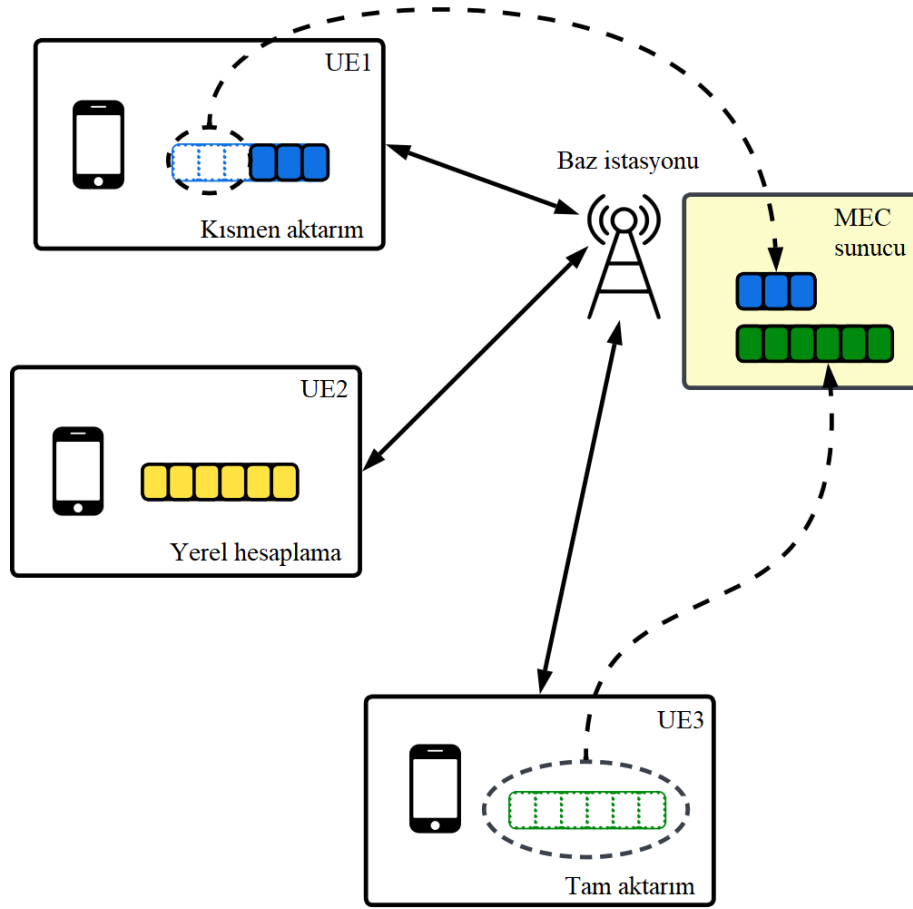
Yerel hesaplama, göreve ilişkin verilerin tamamı mobil aygıtın dahili kaynakları kullanılarak işlenmektedir. MEC sunucularında herhangi bir müsait hesaplama kaynağının bulunmadığı veya kablosuz kanal kalitesinin düşük olması nedeniyle veri iletiminde yüksek gecikmelerin ortaya çıkması durumlarında görev aktarımı yerine yerel hesaplama tercih edilmektedir.

Tam aktarım, mobil aygıtlar tarafından görevlerin tümünün hesaplanmak üzere MEC sunucularına gönderilmesi olarak tanımlanmaktadır. Fakat, mobil aygıtlara gelen görevlerin tümünün sunuculara aktarımı tercih edildiğinde ağda bulunan kullanıcıların artan miktarı ile birlikte sunucu tarafında tüm görevlerin sıralanması kaynaklı olarak ilave hesaplama gecikmeleri ortaya çıkmaktadır.

Kısmen aktarımda ise görevler, bir kısmı yerel olarak hesaplanmak üzere diğer kısmı MEC sunucularına aktarılmak üzere ayrı parçalara bölünmektedir. Uygulama seviyesinde mobil aygıtların ve sunucuların paralel olarak çalışabilmesi, görevlerin aktarım için ayrılabilmesi, ayrılmış görevler arasındaki bağımlılık ve kanal kalitesi bakımından; sunuculara görevlerin kısmen aktarılması karmaşık bir işlemdir. Kısmen aktarımda temel problem, sistem kısıtlarının dikkate alınarak görev ayırma oranının belirlenmesidir. Örneğin, bir aygıtın enerji ve hesaplama kaynaklarının tükendiği durumlarda, görevlerin MEC sunucuya aktarımı tercih edilmektedir. Bunun yanı sıra, kablosuz kanalın kalitesinin düşük olduğu veya sunucuların sınırlı hesaplama kaynağına sahip olduğu durumlarda görevlerin aygıtlarda yerel olarak hesaplanması daha iyi bir tercih olacaktır.

Ayrılan görevlerin tamamının hesaplanmak üzere MEC sunucularına aktarılabilirdiği veya bu parçaların belirli kısmının aktarılamayan birim görevlerden oluştuğu iki tipte görev mevcuttur. Sistemlerde genellikle daha iyi gecikme ve enerji performansı sağlamak üzere

görevlerin kısmen aktarımı dikkate alınmaktadır [29-32]. Şekil 2.6’da görev aktarımı tercihlerinin görev sıralarına etkisi verilmiştir.



Şekil 2.6 Görev aktarımı tercihlerinin sonuçları

MEC sistemlerinde kullanıcı ekipmanlarına (UE) gelen görevler, uygulamalara özel farklılıklar göstermektedir. Yürütülen uygulamalarda birbirine bağımlı olan görevlerin tamamının, MEC sunucularına aktarılması mümkün değildir. Bu durumda, uygulama gereksinimleri dikkate alınarak görevler için yerel hesaplama veya kısmen aktarım kararı verilmelidir. Örneğin, yüz tanıma uygulamalarında görevlerin kısmen aktarımı tercih edilmektedir. Bu uygulamalar genellikle; görüntü yakalama, yüz tespiti, görüntü işleme, öznitelik çıkarma ve sınıflandırma olmak üzere beş ana görevden oluşmaktadır [37]. Bu uygulamada, görüntü yakalama (kullanıcı arayüzü desteği, kameranın çalıştırılması, yakalanan görüntünün kaydedilmesi) görevinin UE tarafından icra edilmesi gerekirken sinyal işleme algoritmaları gibi karmaşık hesaplamalar içeren diğer görevler MEC sunucularına aktarılmaktadır [1].



### 3. ULTRA GÜVENİLİR DÜŞÜK GECİKMELİ HABERLEŞME

#### 3.1. Sistem Modeli

Geliştirilen MEC sistemi UE'lerin bulunduğu  $U$  ve sunucuların bulunduğu  $S$  kümelerinden oluşmaktadır. Sistem dahilinde, her UE'nin kendi görevlerini hesaplama kabiliyeti mevcuttur. Fakat sınırlı hesaplama kabiliyetleri nedeniyle hesap yoğunluğu yüksek uygulamaların icra edilebilmesi için UE'ler görevlerini hesaplanmak üzere kablosuz olarak MEC sunucularına aktarabilmektedir.

MEC sunucuları çoklu işlem birimine sahip işlemcilerle donatılmıştır. Bu sayede, farklı UE'ler tarafından aktarılan görevler paralel olarak hesaplanabilmektedir. Buna ek olarak, hesaplama ve haberleşme işlemleri için zaman akışı bölümlenerek  $t \in N$  olarak indekslenmiştir ve her bir zaman bölümünün uzunluğu  $\tau$  olarak verilmiştir.

İncelenen MEC ağında konumları itibariyle UE'ler rastgele dağıtılmıştır ve sunucuların konumları sabitlenmiştir. UE'ler görevlerini daha yakın konumda olan, daha güçlü hesaplama sağlayabilen veya mevcut eşleştirilmiş sunucudan daha az iş yüküne sahip başka bir sunucuya aktarabilmektedir. Bu bağlamda, UE-sunucu eşleştirme ve görev aktarımı işlemleri için iki zaman ölçeği önerilmiştir. Ardışık her  $T_0$  uzunluğundaki zaman dilimleri gruplanarak  $n \in Z^+$  olarak indekslenmiştir. Bu zaman dilimleri uzun/yavaş zaman ölçeğini temsil etmektedir. Zaman dilimleri dahilindeki  $t \in N$  indeksli her zaman bölümü de kısa/hızlı zaman ölçeğini temsil etmektedir.

UE-sunucu eşleştirmesi, her zaman diliminin başlangıcında gerçekleştirilmektedir. Eşleştirme sonucu  $\eta_{ij}(n)$  indikatörüyle ifade edilmektedir. Buna göre,  $n$  indeksli zaman diliminde UE  $i$  ve sunucu  $j$  eşleştirmesi,  $\eta_{ij}(n) = 1$  olarak gösterilmektedir. Aksi durumda,  $\eta_{ij}(n) = 0$  olmaktadır. UE-sunucu eşleştirmesi, (1)'de verilen kurala göre yürütülmektedir:

$$\begin{aligned} \eta_{ij}(n) &\in \{0, 1\}, \quad \forall i \in U, j \in S, \\ \sum_{j \in S} \eta_{ij}(n) &= 1, \quad \forall i \in U \end{aligned} \quad (1)$$

Bu kurala göre her UE'nin, belirli bir zaman diliminde yalnızca bir sunucuyla eşleşmesine izin verilmektedir. UE-sunucu eşleştirmesinin ardından bir sonraki zaman dilimine kadar

olan ardışık zaman bölümlerinde her bir UE, ilişkili MEC sunucuya görevlerinin bir kısmını dinamik olarak aktarıp kalan görevleri de yerel olarak hesaplamaktadır.

### 3.1.1. UE Tarafındaki Trafik Modeli

UE'lerde yürütülen bir uygulamada bütün görevler rastgele bir biçimde gelmektedir. Görevlerin yerel olarak (UE üzerinde) veya uzaktan (sunucu üzerinde) hesaplanabildiği, farklı görevlerin birbirinden bağımsız olduğu ve bu sayede paralel olarak hesaplanabildiği varsayılmıştır. Her bir zaman bölümünde ortalama görev gelme oranı;  $\lambda_i = E[Av_i]/\tau$  olarak verilmiştir. Gelen görevler bağımsız ve özdeş olasılık dağılımına sahiptir. UE'lerin her biri, yerel hesaplama ve sunucuya aktarım için gelen görevlerini iki parçaya ayırmaktadır. UE tarafında görevlerin ayrılması işlemi aşağıda verilmiştir:

$$\begin{aligned} Av_i(t) &= Av_i^{Loc}(t) + Av_i^{Off}(t), \\ Av_i^{Loc}(t), Av_i^{Off}(t) &\in \{0, Av_{unit}, 2Av_{unit}, \dots\} \end{aligned} \quad (2)$$

Her UE, yerel hesaplama ve sunucuya aktarım için ayrılmış görevlerini iki ayrı sıra belleğinde saklamaktadır. UE  $i$ 'nin yerel hesaplanan ve MEC sunucusuna aktarılan görev sıra belleklerinin  $t$  indeksli zaman bölümündeki uzunlukları ve AWGN kanal üzerinden görev aktarım hızı sırasıyla aşağıda verildiği gibi ifade edilmektedir:

$$Q_i^{Loc}(t+1) = \max \left\{ Q_i^{Loc}(t) + Av_i^{Loc}(t) - \frac{\tau f_i(t)}{L_i}, 0 \right\} \quad (3)$$

$$Q_i^{Off}(t+1) = \max \left\{ Q_i^{Off}(t) + Av_i^{Off}(t) - \sum_{j \in S} \tau R_{ij}(t), 0 \right\} \quad (4)$$

$$R_{ij}(t) = W \log_2 \left( 1 + \frac{\eta_{ij}(n) P_i(t) h_{ij}(t)}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j}(n) P_{i'}(t) h_{i'j}(t)} \right) \quad (5)$$

Burada verilen  $f_i(t)$ , görevler yerel olarak hesaplanırken UE  $i$  tarafından tahsis edilen bir saniyedeki veri işleme hızıdır.  $L_i$  ise bir bit hesaplama için gereken çevrim sayısını göstermektedir.  $Q_i^{Loc}(t)$ , önceki zaman bölümleri boyunca ( $t$  indeksli zaman bölümüne kadar) biriken ve UE  $i$  üzerinde işlenecek veri miktarını göstermektedir.  $Av_i^{Loc}(t)$  ise,  $t$  indeksli zaman bölümünde UE  $i$ 'nin yerel hesaplanacak görevlerinin bulunduğu sıra belleğine eklenen veri miktarını temsil etmektedir. Yerel hesaplamalar için, UE'lerin

dinamik gerilim ve veri işlem hızı ölçekleme (DVFS) kabiliyetine sahip olduğu ve bu sayede işlemci hızının ayarlanabildiği varsayılarak, her bir UE'nin yerel hesaplamalardaki güç tüketim değeri  $\kappa[f_i(t)]^3$  olarak verilmiştir. Burada güç tüketim parametresini temsil eden  $\kappa$  değeri; UE'lerin donanım mimarisi ile ilişkili etkin anahtarlama kapasitesidir [20, 33].

$R_{ij}(t)$ , UE  $i$  görevlerini MEC sunucu  $j$ 'ye aktarırken  $t$  zaman bölümündeki veri iletim hızı olarak tanımlanmıştır ve verilerin iletiminde, normal dağılım esaslı kanal modeli kullanılmıştır.  $P_i(t)$  ve  $N_0$ , sırasıyla UE  $i$ 'nin iletim gücü ve AWGN kanalın spektral yoğunluğu olarak verilmiştir. Buna ek olarak, her bir MEC sunucuya ayrılan bant genişliği  $W$ ; UE  $i$  ve MEC sunucu  $j$  arasındaki kanal kazancı  $h_{ij}$  olarak gösterilmiştir.  $Q_i^{Off}(t)$ , önceki zaman bölümleri boyunca ( $t$  indeksli zaman bölümüne kadar) biriken ve UE  $i$  tarafından MEC sunucuya aktarılacak veri miktarını göstermektedir.  $Av_i^{Off}(t)$  ise,  $t$  indeksli zaman bölümünde UE  $i$  tarafından MEC sunucuya aktarılacak görevlerin bulunduğu sıra belleğine eklenen veri miktarını temsil etmektedir.

Yerel hesaplama ve görev aktarımı işlemleri kapsamında yapılan kaynak tahsisinde her bir UE için aşağıda verilen kısıtlar uygulanarak toplam güç tüketimi sınırlandırılmıştır:

$$\begin{aligned} \kappa[f_i(t)]^3 + P_i(t) &\leq P_i^{max}, \\ f_i(t) &\geq 0, \quad P_i(t) \geq 0 \end{aligned} \quad (6)$$

### 3.1.2. Sunucu Tarafındaki Trafik Modeli

Her bir MEC sunucuda, eşleştirilen farklı UE'lerden aktarılan görevleri saklamak üzere kullanılan birbirinden ayrı görev sıra belleklerinin mevcut olduğu varsayılmıştır. MEC sunucu  $j$ 'ye  $t$  indeksli zaman bölümünde UE  $i$  tarafından aktarılan görev sıralarının uzunluğu ve bu uzunlukların zamanla değişimi (7) olarak verilmiştir:

$$\begin{aligned} M_{ji}(t+1) = \max \left\{ M_{ji}(t) + \min \{ Q_i^{Off}(t) + Av_i^{Off}(t), \tau R_{ij}(t) \} \right. \\ \left. - \frac{\tau f_{ji}(t)}{L_i}, 0 \right\} \leq \max \left\{ M_{ji}(t) + \tau R_{ij}(t) - \frac{\tau f_{ji}(t)}{L_i}, 0 \right\} \end{aligned} \quad (7)$$

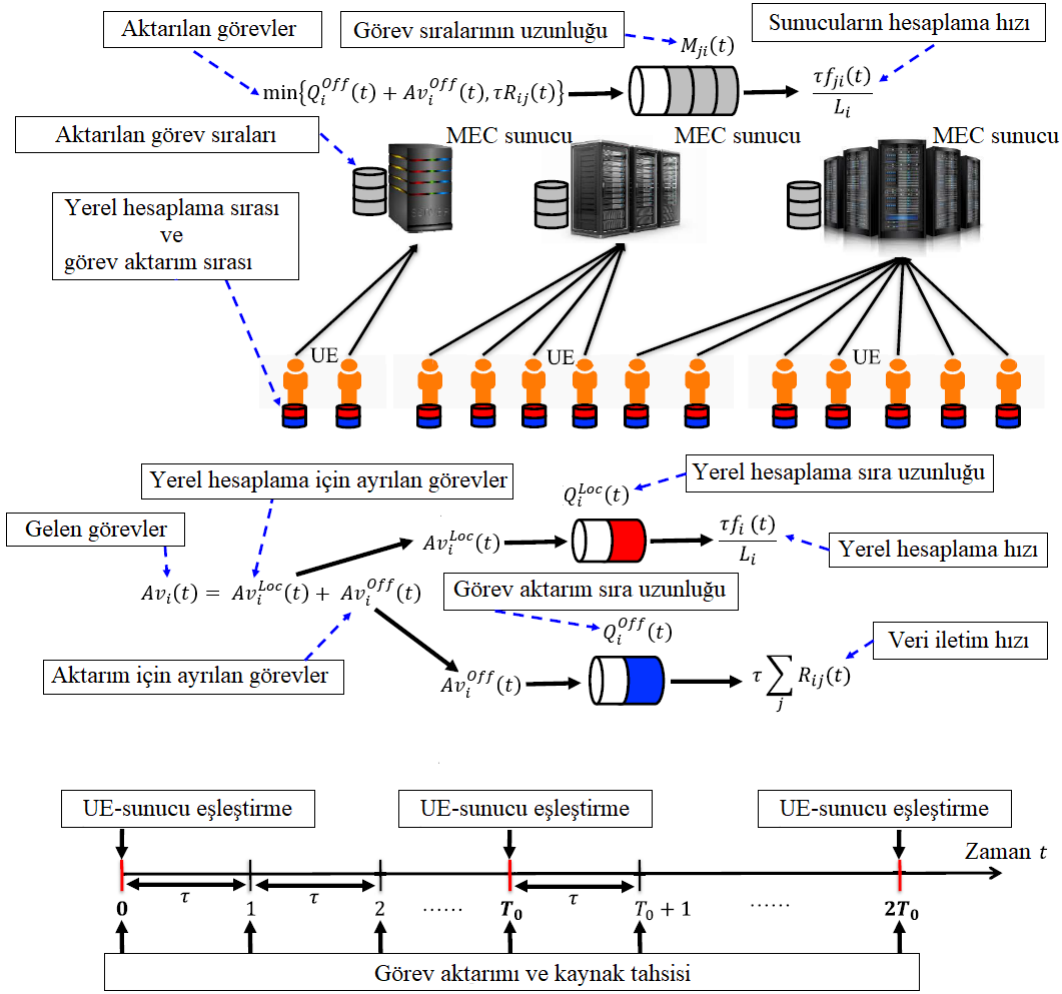
Burada verilen  $f_{ji}(t)$ , UE  $i$ 'nin aktardığı görevleri hesaplamak üzere MEC sunucu  $j$  tarafından  $t$  indeksli zaman bölümünde tahsis edilen bir saniyedeki veri işleme hızıdır.

$M_{ji}(t)$ , önceki zaman bölümleri boyunca ( $t$  indeksli zaman bölümüne kadar) biriken ve MEC sunucu üzerinde işlenecek veri miktarını göstermektedir. MEC sunucuya aktarılan görevler (sunucunun görev sıra belleğine eklenen veri miktarı), UE tarafında görev aktarım sıra belleğindeki veri miktarının veya zaman bölümü boyunca aktarılan veri miktarının minimum değeri olarak tanımlanmıştır. Buna göre MEC sunucuya gelen görevler,  $\min\{Q_i^{Off}(t) + Av_i^{Off}(t), \tau R_{ij}(t)\}$  olarak verilmiştir.

Sistem dahilinde MEC sunucularının, UE'lere daha hızlı hesaplama kabiliyeti sağlamak üzere kullanıldığı dikkate alınarak sunuculardaki her işlem birimi, bir zaman bölümünde en fazla bir UE tarafından aktarılan görevleri hesaplamaya ayrılmaktadır [13, 33]. MEC sunucularının hesaplama kaynakları, (8)'de verilen kurala göre planlanmaktadır:

$$\begin{aligned} \sum_{i \in U} \mathbf{1}_{\{f_{ji}(t) > 0\}} &\leq Cn_j, \forall j \in S \\ f_{ji}(t) &\in \{0, f_j^{max}\}, \forall i \in U, j \in S \end{aligned} \quad (8)$$

İndikatör fonksiyonu  $\mathbf{1}_{\{f_{ji}(t) > 0\}}$ , MEC sunucu  $j$  üzerinde UE'ler için tahsis edilen sıfırdan büyük veri işleme hızlarının alt kümesini kapsamaktadır. İndikatör fonksiyonu ile bu alt kümede bulunan elemanlar bir; alt küme dışında kalan elemanlar ise sıfır olarak gösterilmektedir. Bu sayede, MEC sunucuları ile eşleştirilen UE'lere tahsis edilen faal işlem birimi sayısının toplamı, sunucudaki mevcut toplam işlem birimi sayısını gösteren  $Cn_j$  değerinden küçük veya eşit olmaktadır. Ayrıca, MEC sunucularında hesaplama kaynaklarının planlanmasında mevcut zaman dilimindeki UE-sunucu eşleştirmesine bakılmaksızın MEC sunucularındaki işlemcilerde tahsis edilen veri işleme hızları, sunucuya aktarılan görev sıralarının ağırlığına göre tüm UE'ler için hesaba katılmaktadır. Mevcut zaman diliminde MEC sunucusu  $j$  üzerinde, mevcut zaman diliminde eşleştirilen UE'ler tarafından aktarılan görevler arasında ağırlığı en fazla olanlara işlem birimlerinde hesaplama için bir saniyedeki maksimum veri işleme hızı olarak verilen  $f_j^{max}$  tahsis edilmektedir. Bununla birlikte, mevcut zaman diliminde eşleştirilmeyen veya aktarılan görev ağırlıkları daha az olan UE'ler için, veri işleme hızları sıfır olarak atanmaktadır. Böylelikle, önceki zaman dilimlerinde MEC sunucularına aktarılan fakat henüz işlenmeyen veriler de hesaba katılmaktadır. Bu çalışma kapsamında incelenen MEC sistem mimarisi ve iki ölçekli zaman akışı, sırasıyla Şekil 3.1'de verilmiştir.



Şekil 3.1 MEC sistem mimarisi ve iki ölçekli zaman akışı [14]

### 3.2. Gereksinimlerin ve Problemin Tanımı

UE'lerde yerel olarak hesaplanan ve hesaplanmak üzere sunuculara aktarılan görevlerde ortaya çıkan uçtan uca gecikmeler farklı bileşenlerden oluşmaktadır. UE üzerinde görevler yerel olarak hesaplanırken uçtan uca gecikme, görev sıralama ve hesaplama gecikmelerinden oluşmaktadır. Görevler MEC sunucularına aktarıldığında ise uçtan uca gecikme, UE üzerinde aktarım için görev sıralama, kablosuz iletim, sunucular üzerinde görev sıralama ve hesaplama gecikmelerini içermektedir. Görev sıralamanın ortalama gecikmesi, görev sırasının ortalama uzunluğu ile orantılıdır [34]. Fakat görev sıralarının istatistiklerinde kuyruk dağılımları hesaba katılmadan, yalnızca ortalama görev sıra uzunluklarına odaklanmak düşük gecikme ve güvenilirlik gereksinimlerini karşılamakta yetersiz kalmaktadır [8]. Bu kapsamda, her bir UE için yerel hesaplama ve görev aktarımı sıralarının uzunluklarına (9) ve (10) olasılıksal kısıtlamaları tanımlanmıştır:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T P(Q_i^{Loc}(t) > u_i^{Loc}) \leq \varepsilon_i^{Loc} \quad (9)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T P(Q_i^{Off}(t) > u_i^{Off}) \leq \varepsilon_i^{Off} \quad (10)$$

Burada  $u_i^{Loc}$  ve  $u_i^{Off}$ , sırasıyla yerel hesaplanan ve sunucuya aktarılan görevlerin sıra uzunluk sınırları;  $\varepsilon_i^{Loc} \ll 1$  ve  $\varepsilon_i^{Off} \ll 1$  ise, göre sıralarında uzunluk sınır ihlali olasılıkları olarak verilmiştir. Görev sıralarında uzunluk sınırının ihlali, güvenilirlik gereksiniminin karşılanması dairesinde önem arz etmektedir. Örneğin, sonlu boyuta sahip bir sıra belleği kapasitesinin üzerinde görevle yüklendiğinde, bellekte yeterli alan kalmadığından gelen yeni görevler yok sayılacağından, sistemin güvenilirlik performansı ciddi ölçüde azalacaktır. Yerel hesaplanan ve sunucuya aktarılan görev sıralarının uzunluk sınırlarını aşması olaylarını ve görev sıralarının kuyruk dağılımlarını karakterize etmek üzere ekstrem değer kuramının ilkelerinden yararlanılmıştır. Bu bağlamda, *Pickands-Balkema-De Haan kuramı* [15] aşağıda tanıtılmıştır:

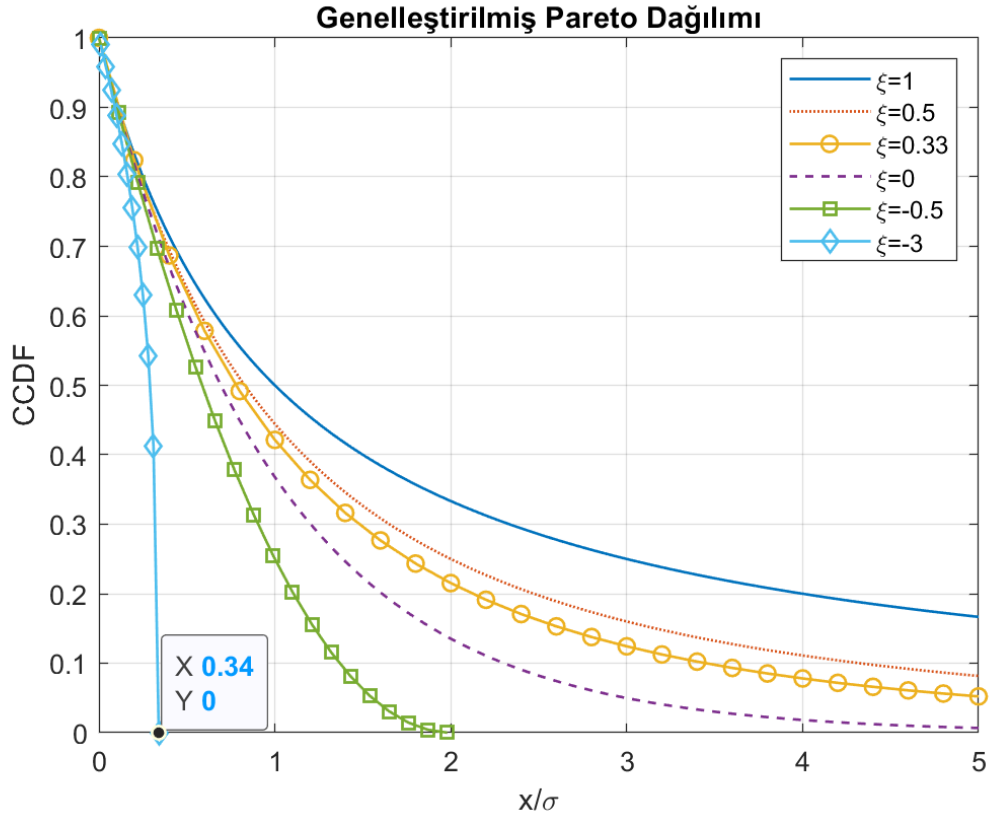
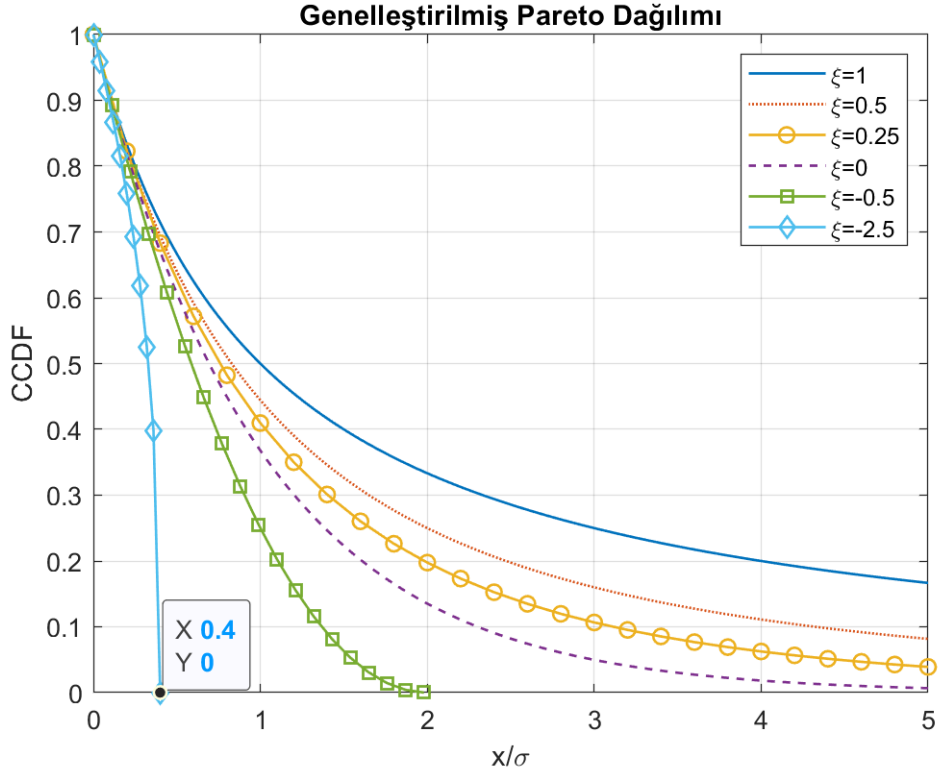
Birikimli dağılım fonksiyonu (CDF)  $F_Q(q)$  olan bir  $Q$  rastgele değişkeni ve  $u$  eşik değeri için, bu eşik değeri  $u \rightarrow \sup\{q: F_Q(q)\} < 1$  olarak seçildiğinde ve görev sırasının fazlalık değerleri  $X|Q > u = Q - u > 0$  rastgele değişken olarak verildiğinde, bu değişkenin tamamlayıcı birikimli dağılım fonksiyonu (CCDF)  $F'_{X|Q>u}(x|Q > u)$ , genelleştirilmiş Pareto dağılımı (GPD) ile hesaplanmaktadır. GPD fonksiyonu  $G(x; \sigma, \xi)$ :

$$G(x; \sigma, \xi) = \begin{cases} \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}, & x \geq 0 \text{ ve } \xi > 0 \\ e^{-x/\sigma}, & x \geq 0 \text{ ve } \xi = 0 \\ \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}, & 0 \leq x \leq -\sigma/\xi \text{ ve } \xi < 0 \end{cases} \quad (11)$$

Yukarıda verilen GPD fonksiyonu; ölçek  $\sigma > 0$  ve şekil  $\xi \in \mathcal{R}$  parametreleriyle ifade edilmektedir. Şekil 3.2’de farklı şekil parametreleri için GPD fonksiyonları gösterilmiştir. GPD fonksiyonlarında ortalama değer ve varyans sırasıyla (12) ve (13) olarak verilmiştir:

$$E[X_i^{Loc}(t)|Q_i^{Loc}(t) > u_i^{Loc}] \approx \sigma_i^{Loc}/(1 - \xi_i^{Loc}) \quad (12)$$

$$Var(X_i^{Loc}(t)|Q_i^{Loc}(t) > u_i^{Loc}) \approx (\sigma_i^{Loc})^2/[(1 - \xi_i^{Loc})^2(1 - 2\xi_i^{Loc})] \quad (13)$$



Şekil 3.2 Farklı şekil parametreleri için GPD fonksiyonları

Görev sıralarının kuyruk dağılımlarını karakterize etmek üzere kullanılacak GPD modelinde gerçek değerlere en yakın dağılımın elde edilebilmesi adına şekil ve ölçek parametreleri bulunmalıdır. Bu bağlamda, GPD fonksiyonları için parametrelerin kestirimi aşağıda verilmiştir [35]:

$$\begin{aligned}\xi_i &= \frac{E[(X_i)^2] - 2E[X_i]^2}{2E[(X_i)^2] - 2E[X_i]^2} \in \mathcal{R}, \\ \sigma_i &= \frac{E[(X_i)^2]E[X_i]}{2E[(X_i)^2] - 2E[X_i]^2} > 0\end{aligned}\quad (14)$$

$$\begin{aligned}c_i^M(t) &= \frac{\sum_{\tau=1}^t (Q_i^{Loc}(\tau) - u_i^{Loc}) \times \mathbf{1}_{\{Q_i^{Loc}(\tau) - u_i^{Loc} > 0\}}}{\sum_{\tau=1}^t \mathbf{1}_{\{Q_i^{Loc}(\tau) - u_i^{Loc} > 0\}}}, \\ c_i^V(t) &= \frac{\sum_{\tau=1}^t (Q_i^{Loc}(\tau) - u_i^{Loc})^2 \times \mathbf{1}_{\{Q_i^{Loc}(\tau) - u_i^{Loc} > 0\}}}{\sum_{\tau=1}^t \mathbf{1}_{\{Q_i^{Loc}(\tau) - u_i^{Loc} > 0\}}}\end{aligned}\quad (15)$$

$$\begin{aligned}\hat{\xi}_i(t) &= \frac{c_i^V(t) - 2[c_i^M(t)]^2}{2c_i^V(t) - 2[c_i^M(t)]^2}, \\ \hat{\sigma}_i(t) &= \frac{c_i^V(t)c_i^M(t)}{2c_i^V(t) - 2[c_i^M(t)]^2}\end{aligned}\quad (16)$$

Yerel hesaplanan ve sunucuya aktarılan görev sıralarının fazlalık değerleri GPD fonksiyonu ile karakterize edilmektedir. Fazlalık değerleri gösteren kuyruk dağılımları, (12) ve (13) dahilinde verilen ortalama değer ve varyans üzerinden sınırlandırılmıştır. Bu bağlamda, GPD şekil ve ölçek parametrelerine eşik değerler tanımlanarak her bir UE'nin yerel hesaplama ve görev aktarımı sıraları için fazlalık değerlerinin ortalamaları ve ikinci momentlerine (17), (18), (19) ve (20)'de verilen kısıtlamalar uygulanmıştır:

$$\bar{X}_i^{Loc} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[X_i^{Loc}(t) | Q_i^{Loc}(t) > u_i^{Loc}] \leq \frac{\sigma_i^{Loc,th}}{1 - \xi_i^{Loc,th}} \quad (17)$$

$$\begin{aligned}\bar{K}_i^{Loc} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ (X_i^{Loc}(t))^2 \mid Q_i^{Loc}(t) > u_i^{Loc} \right] \\ &\leq \frac{2(\sigma_i^{Loc,th})^2}{(1 - \xi_i^{Loc,th})(1 - 2\xi_i^{Loc,th})}\end{aligned}\quad (18)$$

$$\bar{X}_i^{Off} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[X_i^{Off}(t) | Q_i^{Off}(t) > u_i^{Off}] \leq \frac{\sigma_i^{Off,th}}{1 - \xi_i^{Off,th}} \quad (19)$$



$$\begin{aligned}\bar{K}_i^{Off} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \left( X_i^{Off}(t) \right)^2 \mid Q_i^{Off}(t) > u_i^{Off} \right] \\ &\leq \frac{2(\sigma_i^{Off,th})^2}{(1 - \xi_i^{Off,th})(1 - 2\xi_i^{Off,th})}\end{aligned}\quad (20)$$

MEC sunucu tarafında da benzer şekilde görev sıralamanın ortalama gecikmesi, görev sıralarının ortalama uzunluğu ve iletim hızı ile orantılıdır. Bu kapsamda, sunuculara aktarılan görev sıralarına (21)'de verilen olasılıksal kısıtlama tanımlanmıştır:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T P \left( \frac{M_{ji}(t)}{\tilde{R}_{ij}(t-1)} > u_{ji} \right) \leq \varepsilon_{ji} \quad (21)$$

Burada  $\tilde{R}_{ij}(t-1) = \frac{1}{t} \sum_{\tau=0}^{t-1} \tilde{R}_{ij}(\tau)$ , iletim hızının ilerleyen zaman ortalamasını ifade etmektedir. UE tarafındaki görev sıralarının fazlalık değerlerinin ifadesine benzer şekilde, sunuculara aktarılan görev sıralarının uzunluk sınırı  $\tilde{R}_{ij}(t-1)d_{ji}$  olarak verilmiştir. Bu sayede, sunuculardaki görev sıralarının fazlalık değerlerinin kuyruk dağılımına GPD fonksiyonu ile yaklaşım sağlanmaktadır ve böylelikle sunucu tarafındaki görev sıralama dinamikleri de hesaba katılarak sistemin gecikme performansının geliştirilmesi hedeflenmektedir. (17)-(20)'de tanımlandığı üzere, GPD şekil ve ölçek parametrelerine eşik değerler atanarak MEC sunucularına aktarılan görev sıralarının fazlalık değerlerinin ortalamalarına ve ikinci momentlerine sırasıyla (22) ve (23) kısıtlamaları uygulanmıştır:

$$\bar{X}_{ji} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ X_{ji}(t) \mid M_{ji}(t) > \tilde{R}_{ij}(t-1)u_{ji} \right] \leq \frac{\sigma_{ji}^{th}}{1 - \xi_{ji}^{th}} \quad (22)$$

$$\begin{aligned}\bar{K}_{ji} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \left( X_{ji}(t) \right)^2 \mid M_{ji}(t) > \tilde{R}_{ij}(t-1)u_{ji} \right] \\ &\leq \frac{2(\sigma_i^{th})^2}{(1 - \xi_{ji}^{th})(1 - 2\xi_{ji}^{th})}\end{aligned}\quad (23)$$

UE tarafında, yerel hesaplama gecikmesi ve görev aktarımı sırasında meydana gelen iletim gecikmesi (3) ve (4)'te verildiği üzere sırasıyla  $f_i(t)/L_i$  ve  $\sum_{j \in S} R_{ij}(t)$  ile ters orantılıdır. UE'lerde daha yüksek CPU frekansı ve iletim gücü tahsis edilmesi sayesinde daha düşük hesaplama ve iletim gecikmeleri sağlanabilmektedir. Hatta yerel hesaplanan

ve sunuculara aktarılan görev sıraları, gecikmeler ile dolaylı olarak ilişkili olduğundan bu sayede görev sıralarının uzunlukları da azaltılmaktadır. Ancak frekansın ve iletim gücünün artırılması sonucunda enerji yetersizliği ortaya çıkmaktadır.

Sunucularda bulunan gelişmiş hesaplama kabiliyetlerine sahip her işlem birimi, bir zaman bölümünde en fazla bir UE'nin aktardığı görevleri hesaplamak üzere atanmaktadır. Bu nedenle, sunucu tarafındaki hesaplama gecikmesi ihmal edilmiştir. Bunun yanı sıra, sunucular eşleştirilen UE'ler için hesaplama kaynaklarını planlamaktadır.

### **3.3. Görev Aktarımı ve Kaynak Tahsisi**

Her bir zaman diliminin başlangıcında (ardışık  $T_0$  zamanlarında); kablosuz kanal kazancı, UE'lerin ve sunucuların hesaplama kabiliyetleri ve tarihsel iş yükleri dikkate alınarak UE-sunucu eşleştirmesi gerçekleştirilmektedir. Bu bağlamda, çok miktarda UE'nin bir sunucu ile eşleştirilmesi dairesinde çoktan bire eşleştirme algoritması kullanılmıştır. Eşleştirmenin ardından her zaman bölümü dahilinde Lyapunov optimizasyon problemi çözülerek görev aktarımı ve kaynak tahsisi işlemleri yürütülmektedir. Bu kapsamda, her bir UE'ye anlık gelen görevler, yerel hesaplanmak ve sunucuya aktarılmak üzere iki parçaya ayrılırken CPU frekansı ve iletim gücü tahsis edilmektedir. MEC sunucularında ise, aktarılan görevleri hesaplamak üzere işlem birimleri planlanmaktadır. Görevlerin ayrılması, aktarımı ve kaynakların tahsisi; (1), (2), (6) ve (8) olarak verilen URLLC kısıtlamaları dikkate alınarak gerçekleştirilmektedir.

#### **3.3.1. Lyapunov Optimizasyon Problemi**

UE'lerde ve sunucularda mevcut tüm görev sıraları ve URLLC kısıtlamaları, karesel Lyapunov fonksiyonu biçiminde tanımlanmıştır. Bu sayede, ardışık zaman bölümlerinde görev sıralarının uzunlukları arasındaki sürüklenme ile birlikte yerel hesaplama ve iletim güçlerinin ağırlıklı toplamları minimize edilerek ağ dahilinde görev sıralarının ve tüketilen gücün kararlılığı sağlanmaktadır. Bu kapsamda (17)-(20), (22) ve (23) olarak verilen URLLC kısıtlamaları, (24)-(29)'da sanal görev sıraları biçiminde tanımlanmıştır:

$$Q_i^{Loc,(X)}(t+1) = \max \left\{ Q_i^{Loc,(X)}(t) + \left( X_i^{Loc}(t+1) - \frac{\sigma_i^{Loc,th}}{1 - \xi_i^{Loc,th}} \right) \right. \\ \left. \times \mathbf{1}_{\{Q_i^{Loc}(t+1) > u_i^{Loc}\}}, 0 \right\} \quad (24)$$

$$Q_i^{Loc,(K)}(t+1) = \max \left\{ Q_i^{Loc,(K)}(t) + \left( K_i^{Loc}(t+1) - \frac{2(\sigma_i^{Loc,th})^2}{(1 - \xi_i^{Loc,th})(1 - 2\xi_i^{Loc,th})} \right) \right. \\ \left. \times \mathbf{1}_{\{Q_i^{Loc}(t+1) > u_i^{Loc}\}}, 0 \right\} \quad (25)$$

$$Q_i^{Off,(X)}(t+1) = \max \left\{ Q_i^{Off,(X)}(t) + \left( X_i^{Off}(t+1) - \frac{\sigma_i^{Off,th}}{1 - \xi_i^{Off,th}} \right) \right. \\ \left. \times \mathbf{1}_{\{Q_i^{Off}(t+1) > u_i^{Off}\}}, 0 \right\} \quad (26)$$

$$Q_i^{Off,(K)}(t+1) = \max \left\{ Q_i^{Off,(K)}(t) + \left( K_i^{Off}(t+1) - \frac{2(\sigma_i^{Off,th})^2}{(1 - \xi_i^{Off,th})(1 - 2\xi_i^{Off,th})} \right) \right. \\ \left. \times \mathbf{1}_{\{Q_i^{Off}(t+1) > u_i^{Off}\}}, 0 \right\} \quad (27)$$

$$Q_{ji}^{(X)}(t+1) = \max \left\{ Q_{ji}^{(X)}(t) + \left( X_{ji}(t+1) - \frac{\sigma_{ji}^{th}}{1 - \xi_{ji}^{th}} \right) \right. \\ \left. \times \mathbf{1}_{\{M_{ji}(t+1) > \bar{R}_{ij}(t)u_{ji}\}}, 0 \right\} \quad (28)$$

$$Q_{ji}^{(K)}(t+1) = \max \left\{ Q_{ji}^{(K)}(t) + \left( K_{ji}(t+1) - \frac{2(\sigma_{ji}^{th})^2}{(1 - \xi_{ji}^{th})(1 - 2\xi_{ji}^{th})} \right) \right. \\ \left. \times \mathbf{1}_{\{M_{ji}(t+1) > \bar{R}_{ij}(t)u_{ji}\}}, 0 \right\} \quad (29)$$

Buna ek olarak, tüm muhtemel sonuçları içeren  $\Omega$  kümesi içindeki bir  $A$  olayının beklenen değeri;  $E[\mathbf{1}_{\{A\}}] = 1 \times P(A) + 0 \times P(\Omega \setminus A) = P(A)$  olarak ifade edilmektedir. Buradan hareketle (9), (10) ve (20) olarak verilen URLLC kısıtlamalarının eşdeğer ifadeleri (30), (31) ve (32); bunlara karşılık gelen sanal görev sıraları (33), (34) ve (35) olarak verilmiştir:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \mathbf{1}_{\{Q_i^{Loc}(t) > u_i^{Loc}\}} \right] \leq \varepsilon_i^{Loc} \quad (30)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \mathbf{1}_{\{Q_i^{Off}(t) > u_i^{Off}\}} \right] \leq \varepsilon_i^{Off} \quad (31)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \mathbf{1}_{\{M_{ji}(t) > \tilde{R}_{ij}(t-1)u_{ji}\}} \right] \leq \varepsilon_{ji} \quad (32)$$

$$Q_i^{Loc,(Q)}(t+1) = \max \left\{ Q_i^{Loc,(Q)}(t) + \mathbf{1}_{\{Q_i^{Loc}(t+1) > u_i^{Loc}\}} - \varepsilon_i^{Loc}, 0 \right\} \quad (33)$$

$$Q_i^{Off,(Q)}(t+1) = \max \left\{ Q_i^{Off,(Q)}(t) + \mathbf{1}_{\{Q_i^{Off}(t+1) > u_i^{Off}\}} - \varepsilon_i^{Off}, 0 \right\} \quad (34)$$

$$Q_{ji}^{(M)}(t+1) = \max \left\{ Q_{ji}^{(M)}(t) + \mathbf{1}_{\{M_{ji}(t+1) > \tilde{R}_{ij}(t)u_{ji}\}} - \varepsilon_{ji}, 0 \right\} \quad (35)$$

URLLC kısıtlamalarını içeren sanal görev sıraları ve UE'lerin yerel hesaplama ve görev aktarımı kapsamında güç tüketimi, (36)'da verilen Lyapunov sürüklenme artı ceza yöntemi ile sınırlandırılmıştır. Bu problemin optimal çözümü, üst sınırın minimizasyonu yoluyla elde edilmektedir. Optimizasyon parametresi ( $V \geq 0$ ) için seçilen değere göre görev sıralarının ardışık zaman bölümlerinde değişimi ve UE'lerin güç tüketim miktarı arasında denge kurulmaktadır.

$$E \left[ L(\mathbf{Q}(t+1)) - L(\mathbf{Q}(t)) + \sum_{i \in U} V(\kappa[f_i(t)]^3 + P_i(t)) | \mathbf{Q}(t) \right], \quad (36)$$

$$\mathbf{Q} = \left( Q_i^{Loc,(X)}, Q_i^{Loc,(K)}, Q_i^{Off,(X)}, Q_i^{Off,(K)}, Q_{ji}^{(X)}, Q_{ji}^{(K)}, \right. \\ \left. Q_i^{Loc,(Q)}, Q_i^{Off,(Q)}, Q_{ji}^{(M)} \right)$$

Burada verilen  $L(\mathbf{Q}(t))$ ;  $\mathbf{Q}$  görev sıra vektörünün  $t$  indeksli zaman bölümündeki karesel Lyapunov fonksiyonudur ve (37)'de verildiği gibi ifade edilmektedir. Lyapunov sürüklenme artı ceza ifadesinde (24)-(29) ve (33)-(35) olarak verilen sanal görev sıraları,  $\mathbf{Q}$  görev sıra vektöründe birleştirilmiştir. Bu görev sıralarının uzunluklarının ardışık zaman bölümlerinde sürüklenmesi (artışı) optimizasyon probleminde sınırlandırılarak görev sıralarında kararlılık sağlanmaktadır ve bu sayede (17)-(20), (22), (23) ve (30)-(32) olarak verilen URLLC kısıtlamaları karşılanmaktadır.

$$\begin{aligned}
& \frac{1}{2} \sum_{i \in U} \left( [Q_i^{Loc,(X)}(t)]^2 + [Q_i^{Loc,(K)}(t)]^2 + [Q_i^{Off,(X)}(t)]^2 + [Q_i^{Off,(K)}(t)]^2 \right. \\
& \quad \left. + [Q_i^{Loc,(Q)}(t)]^2 + [Q_i^{Off,(Q)}(t)]^2 \right) \\
& \quad + \frac{1}{2} \sum_{i \in U} \sum_{j \in S} \left( [Q_{ji}^{(X)}(t)]^2 + [Q_{ji}^{(K)}(t)]^2 + [Q_{ji}^{(M)}(t)]^2 \right)
\end{aligned} \tag{37}$$

Lyapunov sürüklenme artı ceza ifadesinin üst sınırının minimizasyonu, üç ayrı optimizasyon problemi çözülerek gerçekleştirilmektedir. Birinci problem, UE'lerin MEC sunucularıyla eşleştirilmesi ile birlikte UE'lerin hesaplama ve haberleşme kaynaklarının tahsis edilmesini içermektedir. İkinci problem, UE'lere  $t$  indeksli zaman bölümünde gelen görevlerin, mevcut görev sıralarının ağırlıklarına göre yerel hesaplama veya MEC sunucularına aktarım için ayrılmasına odaklanmaktadır. Üçüncü problem ise, UE'ler tarafından aktarılan görevlerin sıra ağırlıklarına göre MEC sunucularında kaynakların tahsis edilmesini kapsamaktadır.

$$\begin{aligned}
& \text{minimize}_{\boldsymbol{\eta}(n), \mathbf{f}(t), \mathbf{P}(t)} \sum_{j \in S} \sum_{i \in U} (\varphi_{ji}(t) - \varphi_i^{Off}(t)) \tau W \\
\mathbf{P1:} & \quad \times \log_2 \left( 1 + \frac{\eta_{ij}(n) P_i(t) h_{ij}(t)}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j}(n) P_{i'}(t) h_{i'j}(t)} \right) \\
& \quad + \sum_{i \in U} \left[ V(\kappa [f_i(t)]^3 + P_i(t)) - \frac{\varphi_i^{Loc}(t) \tau f_i(t)}{L_i} \right]
\end{aligned}$$

*Kısıtlamalar:*

$$\begin{aligned}
& \eta_{ij}(n) \in \{0, 1\}, \quad \forall i \in U, j \in S, \\
& \sum_{j \in S} \eta_{ij}(n) = 1, \quad \forall i \in U \\
& \kappa [f_i(t)]^3 + P_i(t) \leq P_i^{max}, \\
& f_i(t) \geq 0, \quad P_i(t) \geq 0
\end{aligned}$$

**P1**'de verilen  $\varphi_i^L$ ,  $\varphi_i^O$  ve  $\varphi_{ji}$ ; sırasıyla yerel hesaplama, görev aktarım ve sunucudaki görev sıralarının  $t$  indeksli zaman bölümündeki ağırlıklarını temsil etmektedir. Bununla birlikte  $\boldsymbol{\eta}(n) = (\eta_{ij}(n): i \in U, j \in S)$ , ağ genelinde UE-sunucu eşleştirme indikatörünü,  $\mathbf{P}(t) = (P_i(t): i \in U)$ , ağ genelinde UE'lerin tahsis ettiği iletim gücü vektörünü,  $\mathbf{f}(t) = (f_i(t), \mathbf{f}_j(t): i \in U, j \in S)$ , ağ genelinde hesaplama kaynakları tahsis vektörünü ve  $\mathbf{f}_j(t) = (f_{ji}(t): i \in U, j \in S)$ , MEC sunucu  $j$ 'nin hesaplama kaynağı tahsis vektörünü göstermektedir. **P1**'de verilen optimizasyon probleminde, UE'lerin tahsis ettiği iletim

gücü ile yerel veri işleme hızı bağlantılıdır. Ayrıca iletim gücü, UE'ler ile eşleştirilen MEC sunucuları arasındaki kanal kazancına ve sunucudaki görev sıralarının ağırlığına ( $\varphi_{ji}$ ) bağlıdır. Kanal kazancı, UE'ler ve MEC sunucuları arasındaki mesafeler ile;  $\varphi_{ji}$  ise MEC sunucularının hesaplama kabiliyetleri ve sunucular ile eşleştirilen UE'lerin miktarı ile ilişkilidir. Bu nedenle, UE'ler ve MEC sunucuları eşleştirilirken, UE'lerin coğrafi konfigürasyonu ve MEC sunucularının hesaplama kabiliyetleri dikkate alınmaktadır. Buna ek olarak, UE-sunucu eşleştirmesi ve kaynak tahsisi, iki farklı zaman ölçeğinde gerçekleştirilmektedir. **P1** optimizasyon problemi iki adımda çözülmüştür, bu bağlamda ilk olarak UE-sunucu eşleştirmesi yürütülmektedir. Eşleştirme sonucuna göre, UE'lerin yerel hesaplama için veri işleme hızları ve sunuculara aktarım için iletim güçleri tahsis edilmektedir. Yerel hesaplama, görev aktarım ve sunucudaki görev sıralarının  $t$  indeksli zaman bölümündeki ağırlıkları, sırasıyla (38), (39) ve (40) olarak verilmiştir.

$$\begin{aligned} \varphi_i^{Loc}(t) = & \left( Q_i^{Loc,(X)}(t) + 2Q_i^{Loc,(K)}Q_i^{Loc}(t) + 2[Q_i^{Loc}(t)]^3 + Q_i^{Loc}(t) \right) \\ & \times \mathbf{1}_{\{Q_i^{Loc}(t)+Av_i(t) > u_i^{Loc}\}} + Q_i^{Loc,(Q)}(t) + Q_i^{Loc}(t) \end{aligned} \quad (38)$$

$$\begin{aligned} \varphi_i^{Off}(t) = & \left( Q_i^{Off,(X)}(t) + 2Q_i^{Off,(K)}Q_i^{Off}(t) + 2[Q_i^{Off}(t)]^3 + Q_i^{Off}(t) \right) \\ & \times \mathbf{1}_{\{Q_i^{Off}(t)+Av_i(t) > u_i^{Off}\}} + Q_i^{Off,(Q)}(t) + Q_i^{Off}(t) \end{aligned} \quad (39)$$

$$\begin{aligned} \varphi_{ji}(t) = & \left( Q_{ji}^{(X)}(t) + 2Q_{ji}^{(K)}M_{ji}(t) + 2[M_{ji}(t)]^3 + M_{ji}(t) \right) \\ & \times \mathbf{1}_{\{M_{ji}(t)+\tau R_{ij}^{max}(t) > \bar{R}_{ij}(t-1)u_{ji}\}} + Q_{ji}^{(M)}(t) + M_{ji}(t) \end{aligned} \quad (40)$$

### 3.3.2. UE-Sunucu Eşleştirmesi, Görev Ayırma ve Kaynak Tahsisi

UE'lerin MEC sunucuları ile eşleştirilmesi kapsamında **P1**'de veri iletim kısmına bakılmaktadır. Bu nedenle, eşleştirme sırasında her UE için  $P_i(t) = P_i^{max}$ ,  $f_i(t) = 0$  olarak seçilmektedir. Ayrıca, kanal kazancı  $h_{ij}(t)$  ve görev sıralarının ağırlıkları  $\varphi_{ji}(t)$  ve  $\varphi_i^{Off}(t)$  ardışık zaman bölümlerinde değişirken UE'ler ardışık  $T_0$  uzunluğundaki zaman dilimlerinde yeniden eşleştirilmektedir. Kanal kazancının ve görev sıralarının ağırlıklarının etkilerini hesaba katmak üzere  $h_{ij}(t) = E[h_{ij}]$  kanalın ortalama kazancı ile birlikte, (41) ve (42) olarak verilen görev sıralarının ağırlıklarının kestirimi için deneysel ortalamaları alınmaktadır.

$$\tilde{\varphi}_i^{Off}(n) = \frac{1}{(n-1)T_0} \sum_{t=0}^{(n-1)T_0-1} \varphi_i^{Off}(t), \quad \forall i \in U \quad (41)$$

$$\tilde{\varphi}_j(n) = \frac{1}{(n-1)T_0} \sum_{k=1}^{n-1} \sum_{t=(k-1)T_0}^{kT_0-1} \sum_{i \in U} \frac{\eta_{ij}(k) \varphi_{ji}(t)}{\sum_{i \in U} \eta_{ij}(k)}, \quad \forall j \in S \quad (42)$$

Her bir UE'nin görev aktarım sırasının ağırlığının kestirim sonucu (41)'de verilmiştir. Ardışık zaman dilimlerinde, bir MEC sunucu ile eşleştirilen tüm UE'lerin görev aktarım sıralarına ait ağırlıkların kestirim sonucu (42)'de verilmiştir. Buradan hareketle, **P1** optimizasyon problemi, UE-sunucu eşleştirmesi kapsamında **P1-1** problemi olarak yeniden düzenlenmiştir.

$$\begin{aligned} \mathbf{P1-1:} \quad & \text{maximize}_{\eta(n)} \sum_{j \in S} \sum_{i \in U} \left( \tilde{\varphi}_i^{Off}(n) - \tilde{\varphi}_j(n) \right) \tau W \\ & \times \log_2 \left( 1 + \frac{\eta_{ij}(n) P_i^{max} E[h_{ij}]}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j}(n) P_{i'}^{max} E[h_{i'j}]} \right) \end{aligned}$$

*Kısıtlamalar:*

$$\begin{aligned} \eta_{ij}(n) & \in \{0, 1\}, \quad \forall i \in U, j \in S, \\ \sum_{j \in S} \eta_{ij}(n) & = 1, \quad \forall i \in U \end{aligned}$$

**P1-1**'de verilen optimizasyon probleminin çözümü için eşleştirme kuramından [16] yararlanılmıştır. UE'ler ve MEC sunucularının eşleştirmesinde çoktan bire eşleştirme modeli kullanılmıştır [41]. Buna göre her bir UE, bir MEC sunucu ile eşleştirilirken bir MEC sunucu birden fazla UE ile eşleştirilmektedir. Uygulanan çoktan bire eşleştirme yönteminde UE'lerin kümesi  $U$ ; MEC sunucuların kümesi  $S$  olarak gösterilmiştir ve  $U \times S$  alanından  $U \times S$ 'nin tüm alt kümelerini kapsayan kümeye tanımlı  $\eta$  fonksiyonu, çoktan bire eşleştirmenin sonucunu vermektedir. Çoktan bire eşleştirme yönteminin kuralları sırasıyla (43)'te verilmiştir.

$$\begin{aligned} |\eta(i)| & = 1, \quad \forall i \in U \\ |\eta(j)| & \leq U, \quad \forall j \in S \\ j = \eta(i) & \leftrightarrow i \in \eta(j) \end{aligned} \quad (43)$$

UE'lerin MEC sunucuları üzerinde eşleştirme tercihleri, **P1-1**'de verilen ağırlıklı veri iletim hızına dayalı olarak azalan sırayla yapılmaktadır. Bir MEC sunucu ile eşleştirilen birden fazla UE'nin kanaldaki girişimi, veri iletim hızlarını düşürmektedir. Bu nedenle, bir UE'nin eşleştirme tercihi, diğer UE'lerin eşleştirme durumlarına bağlıdır. Aynı kümedeki UE'lerin mevcut eşleştirme durumlarına göre tercihleri dinamik olarak değişmektedir. Buradan hareketle, verilen bir  $j = \eta(i)$  eşleştirmesinin etkinliğini hesaplamak üzere (44) ve (45)'te sırasıyla UE  $i$  ve MEC sunucu  $j$  için fayda fonksiyonları tanımlanmıştır.

$$\Gamma_i(\eta) = (\tilde{\varphi}_i^{off} - \tilde{\varphi}_j) \times \log_2 \left( 1 + \frac{P_i^{max} E[h_{ij}]}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j} P_{i'}^{max} E[h_{i'j}]} \right) \quad (44)$$

$$\Gamma_j(\eta) = (\tilde{\varphi}_i^{off} - \tilde{\varphi}_j) \times \log_2 \left( 1 + \frac{\eta_{ij} P_i^{max} E[h_{ij}]}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j}(n) P_{i'}^{max} E[h_{i'j}]} \right) \quad (45)$$

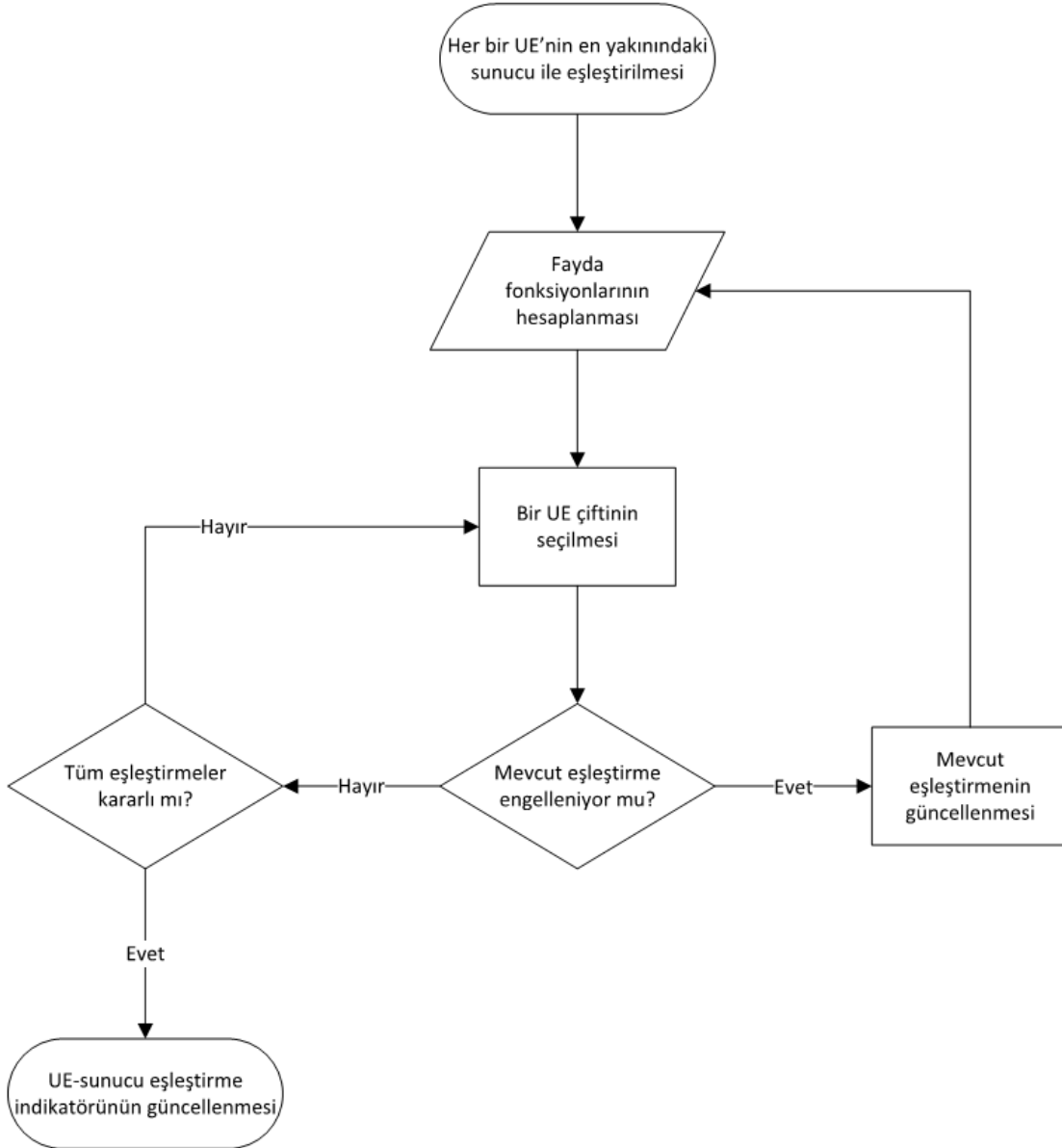
İncelenen çoktan bire eşleştirme yönteminde, fayda fonksiyonunun çıktısına göre UE-sunucu eşleştirmeleri takas edilebilmektedir. Bu bağlamda, verilen bir  $(i, i')$  UE çifti ve  $(j, j')$  MEC sunucu çifti için mevcut eşleştirmeler,  $j = \eta(i)$  ve  $j' = \eta(i')$  ise eşleştirme takası  $\eta_{ij}^{i'j'}$ , mevcut diğer UE-sunucu eşleştirmelerini etkilemeden yalnızca hedef UE ve MEC sunucu çiftini takas etmektedir. Eşleştirme takası (46)'da ifade edilmiştir.

$$\eta_{ij}^{i'j'} = \{\eta \setminus \{(i, j), (i', j')\}\} \cup \{(i', j), (i, j')\} \quad (46)$$

Mevcut UE-sunucu eşleştirmeleri, fayda fonksiyonu çıktısı daha yüksek olan UE ve MEC sunucu çiftleri tarafından engellenebilir. Bu durumda, eşleştirmelerde kararlılığın sağlanabilmesi için mevcut eşleştirmeyi engelleyen UE çiftleri takas edilmektedir. Bir UE çifti, eşleştiği MEC sunucularını takas ettiğinde hiçbir UE'nin ve MEC sunucunun fayda fonksiyonu çıktısı bir önceki eşleştirmenin çıktısından daha düşük değilse ve en az bir fayda fonksiyonunun çıktısı eskiye göre daha yüksek ise bu UE çifti, eşleştikleri MEC sunucularını takas etmektedir. **P1-1**'in çözümü kapsamında kullanılan çoktan bire eşleştirme yöntemi adımları, Şekil 3.3'te akış şeması biçiminde gösterilen Algoritma-1 dahilinde verilmiştir. Buna göre ilk olarak, her bir UE en yakınında bulunan MEC sunucu ile eşleştirilerek tüm UE'lerin ve MEC sunucuların fayda fonksiyonu hesaplanmaktadır. Daha sonra, rastgele bir UE çifti seçilerek mevcut eşleştirmeyi engelleme durumu kontrol



edilmektedir. Eğer seçilen UE çifti, mevcut eşleştirmeyi engelliyor ise UE-sunucu eşleştirmesi takas edilerek tüm UE'lerin ve MEC sunucuların fayda fonksiyonları tekrardan hesaplanmaktadır. Bu adım, tüm UE çiftleri seçilip kontrol edilene kadar tekrarlanmaktadır. Kontrol edilmeyen UE çifti kalmadığında ve mevcut eşleştirmeyi engelleyen UE çifti bulunmadığı (tüm eşleştirmelerin kararlı olduğu) durumda ağ genelinde UE-sunucu eşleştirme indikatörü  $\eta(n)$  güncellenmektedir.



Şekil 3.3 Algoritma-1 akış şeması

UE-sunucu eşleştirmesi tamamlandıktan sonra UE'lerde yerel hesaplama için veri işleme hızlarının ve sunuculara aktarım için iletim güçlerinin tahsisi kapsamında **P1**'in ikinci adımı tanımlanmıştır. Bu bağlamda UE'lerin kaynak tahsisi problemi **P1-2**'de verilmiştir. Burada, UE i'nin eşleştiği MEC sunucu,  $j^*$  olarak gösterilmiştir.

$$\begin{aligned}
& \text{minimize}_{\mathbf{f}(t), \mathbf{P}(t)} \sum_{j \in S} \sum_{i \in U} \left( \varphi_{j^*i}(t) - \varphi_i^{Off}(t) \right) \tau W \\
\mathbf{P1-2:} \quad & \times \log_2 \left( 1 + \frac{P_i(t) h_{ij^*}(t)}{N_0 W + \sum_{i' \in U \setminus i} \eta_{i'j^*}(n) P_{i'}(t) h_{i'j^*}(t)} \right) \\
& + \sum_{i \in U} \left[ V(\kappa [f_i(t)]^3 + P_i(t)) - \frac{\varphi_i^{Loc}(t) \tau f_i(t)}{L_i} \right]
\end{aligned}$$

*Kısıtlamalar:*

$$\begin{aligned}
\kappa [f_i(t)]^3 + P_i(t) &\leq P_i^{max}, \\
f_i(t) &\geq 0, \quad P_i(t) \geq 0
\end{aligned}$$

İkinci optimizasyon probleminin çözümü ile,  $t$  indeksli zaman bölümünde UE'lere gelen görevler yerel hesaplama ve görev aktarım sıralarının ağırlıklarına göre ayrılarak yerel hesaplanan veya MEC sunucusuna aktarılan görev sıra belleğine eklenmektedir.

$$\mathbf{P2:} \quad \text{minimize}_{Av_i^{Loc}(t), Av_i^{Off}(t)} \varphi_i^{Loc}(t) Av_i^{Loc}(t) + \varphi_i^{Off}(t) Av_i^{Off}(t)$$

*Kısıtlamalar:*

$$\begin{aligned}
Av_i(t) &= Av_i^{Loc}(t) + Av_i^{Off}(t), \\
Av_i^{Loc}(t), Av_i^{Off}(t) &\in \{0, Av_{unit}, 2Av_{unit}, \dots\}
\end{aligned}$$

**P2** optimizasyon probleminin çözümü (47)'de verilmiştir. Buna göre, eğer UE  $i$ 'nin yerel hesaplama görev sırasının ağırlığı görev aktarım sırasının ağırlığından küçük veya eşit ise gelen görevler yerel hesaplanan görev sıra belleğine eklenmektedir, yerel hesaplama görev sırasının ağırlığı görev aktarım sırasının ağırlığından büyük ise gelen görevler MEC sunucusuna aktarılan görev sıra belleğine eklenmektedir.

$$\begin{aligned}
(Av_i^{Loc*}(t), Av_i^{Off*}(t)) &= (Av_i(t), 0), \quad \text{eğer } \varphi_i^{Loc}(t) \leq \varphi_i^{Off}(t) \\
(Av_i^{Loc*}(t), Av_i^{Off*}(t)) &= (0, Av_i(t)), \quad \text{eğer } \varphi_i^{Loc}(t) > \varphi_i^{Off}(t)
\end{aligned} \tag{47}$$

### 3.3.3. Sunucu Tarafında Kaynak Planlama

Üçüncü optimizasyon probleminin çözümü ile MEC sunucuları tarafındaki işlem birimleri, UE'ler tarafından aktarılan görevleri hesaplamak üzere tahsis edilmektedir. Bu problemde, mevcut zaman dilimindeki eşleştirmeye bakılmaksızın MEC sunucularında tüm UE'ler için tahsis edilen veri işleme hızları dikkate alınmıştır.

$$\mathbf{P3:} \quad \text{maximize}_{f_j(t)} \sum_{i \in U} \frac{\varphi_{ji}(t) f_{ji}(t)}{L_i}$$

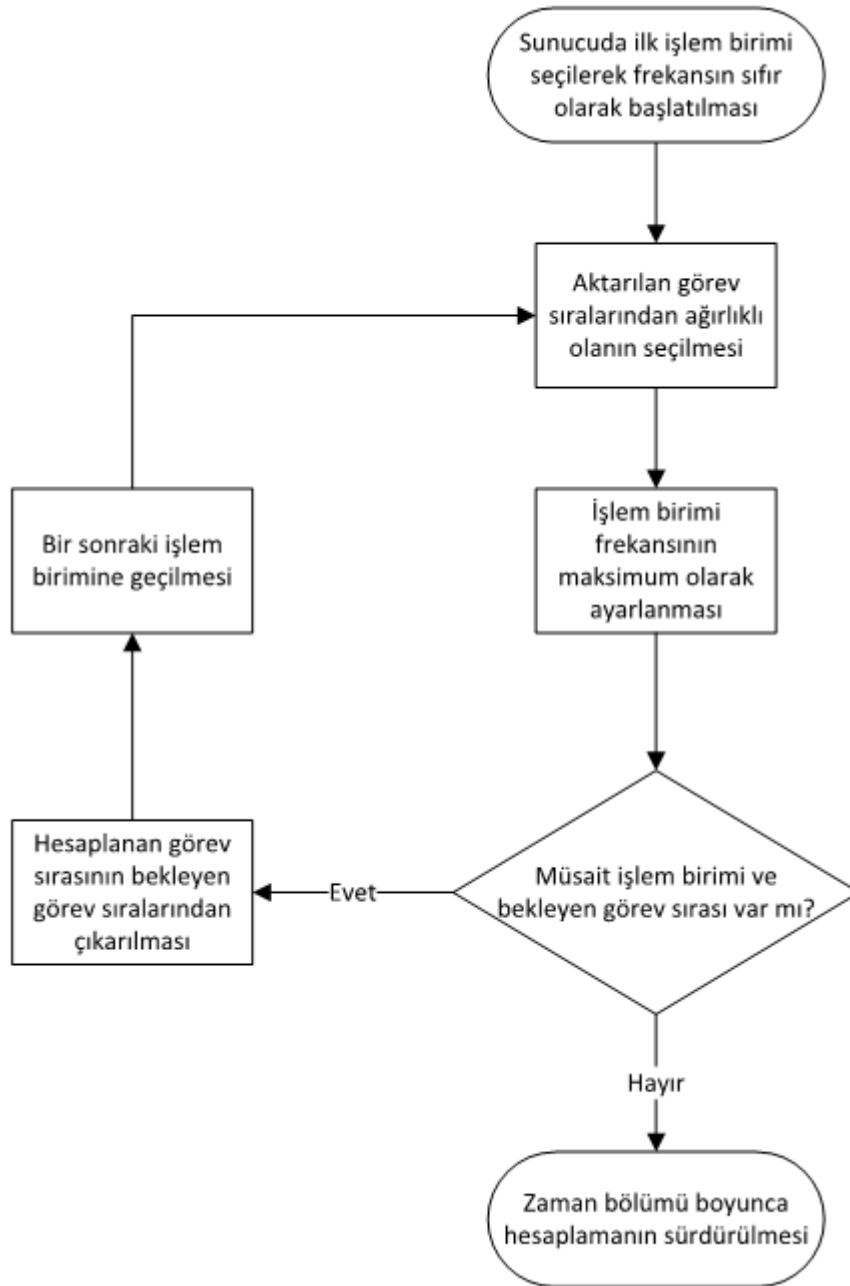
*Kısıtlamalar:*

$$\sum_{i \in U} \mathbf{1}_{\{f_{ji}(t) > 0\}} \leq Cn_j, \forall j \in S$$
$$f_{ji}(t) \in \{0, f_j^{max}\}, \forall i \in U, j \in S$$

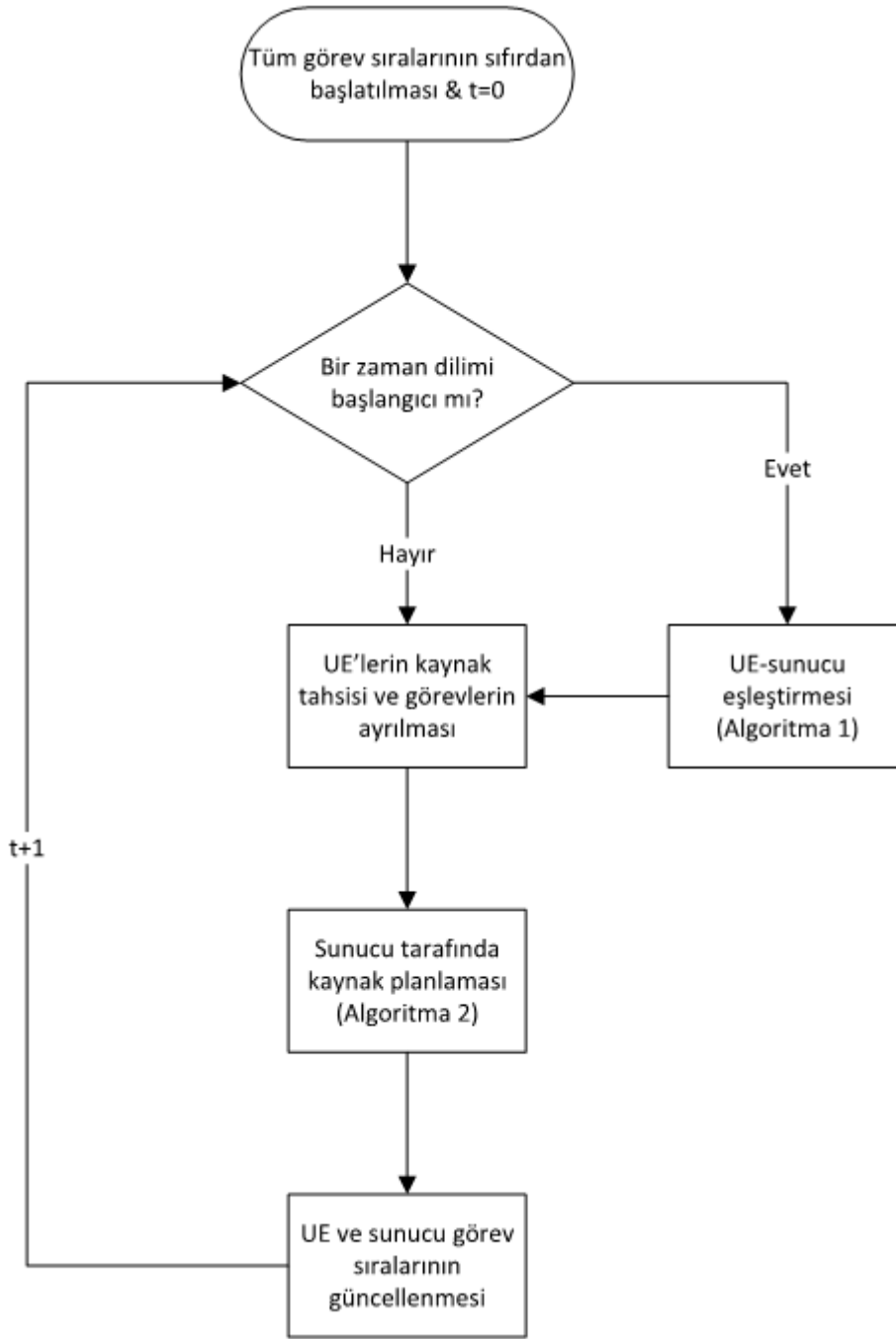
**P3**'ün çözümü kapsamında, sunucuların işlem birimlerinin tahsisi adımları Şekil 3.4'te akış şeması biçiminde gösterilen Algoritma-2 dahilinde verilmiştir. Bu algorithmada, ilk olarak kaynak tahsisi gerçekleştirilecek MEC sunucuda bir işlem birimi seçilerek veri işleme hızı (işlem birimi frekansı), sıfır olarak başlatılmaktadır. Bir sonraki adımda, MEC sunucuya UE'ler tarafından aktarılan ve hesaplanmayı bekleyen görev sıralarından ağırlığı en yüksek olan seçilerek bu görevleri hesaplamak üzere işlem birimine maksimum veri işleme hızı olan  $f_j^{max}$  atanmaktadır ve bu görev sırası, hesaplama bekleyen görev sıraları kümesinden çıkarılmaktadır. Daha sonra MEC sunucuda başka bir işlem birimi seçilerek UE'ler tarafından aktarılan ve hesaplanmayı bekleyen görev sıralarının ağırlığına göre kaynak tahsisi tekrarlanmaktadır. Bu işlem, MEC sunucuda müsait işlem birimi kalmayana veya UE'ler tarafından aktarılan görev sıralarının tamamı hesaplanana kadar devam etmektedir.

Algoritma-1 ile çözülen **P1** ve **P2** optimizasyon problemleri, UE-sunucu eşleştirmesi, UE tarafında gelen görevlerin yerel hesaplama veya MEC sunucuya aktarım için ayrılması, UE tarafında kaynak tahsisi işlemlerini kapsamaktadır. Algoritma-2 ile çözülen **P3** optimizasyon problemi ise MEC sunucuları tarafında hesaplama kaynaklarının tahsisi işlemini kapsamaktadır. Bu algoritmaların birleştirilerek MEC sistemi dahilinde iki zaman ölçeğinde gerçekleştirilen URLLC için dinamik görev aktarımı ve kaynak tahsisi işlemlerinin adımları, Şekil 3.5'te akış şeması biçiminde gösterilen Algoritma-3 dahilinde verilmiştir. Buna göre, ilk olarak UE'lerdeki ve MEC sunucularındaki tüm görev sıraları

sıfırdan başlatılarak UE-sunucu eşleştirmesi (Algoritma-1) icra edilmektedir. UE-sunucu eşleştirme işlemi her zaman dilimi başlangıcında tekrarlanmaktadır. Bir sonraki adımda UE tarafında gelen görevler, yerel hesaplama ve MEC sunucuya aktarım görev sıra belleklerine ayrılarak yerel hesaplamalar için veri işleme hızı, MEC sunucuya aktarımlar için iletim gücü tahsis edilmektedir. Daha sonra, MEC sunucu tarafında UE'ler tarafından aktarılan görevlerin hesaplanması kapsamında işlem birimleri tahsis edilmektedir. Zaman bölümü sonunda tüm görev sıraları güncellenerek görev ayırma ve kaynak tahsisi işlemleri bir sonraki zaman bölümünde tekrarlanmaktadır.



Şekil 3.4 Algoritma-2 akış şeması

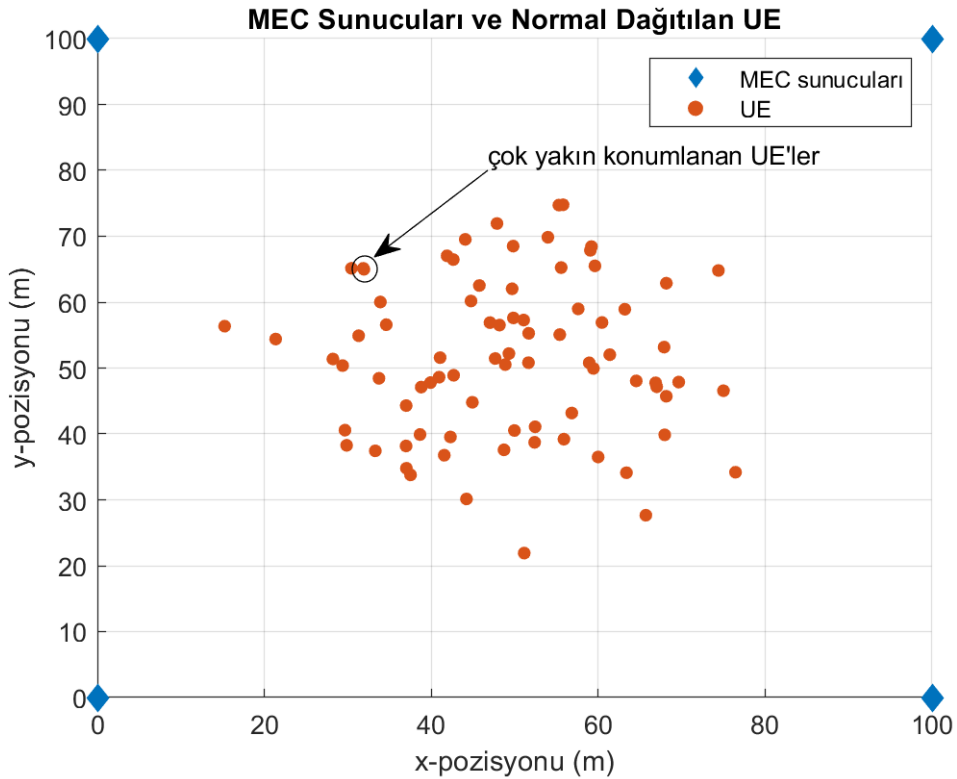


Şekil 3.5 Algoritma-3 akış şeması

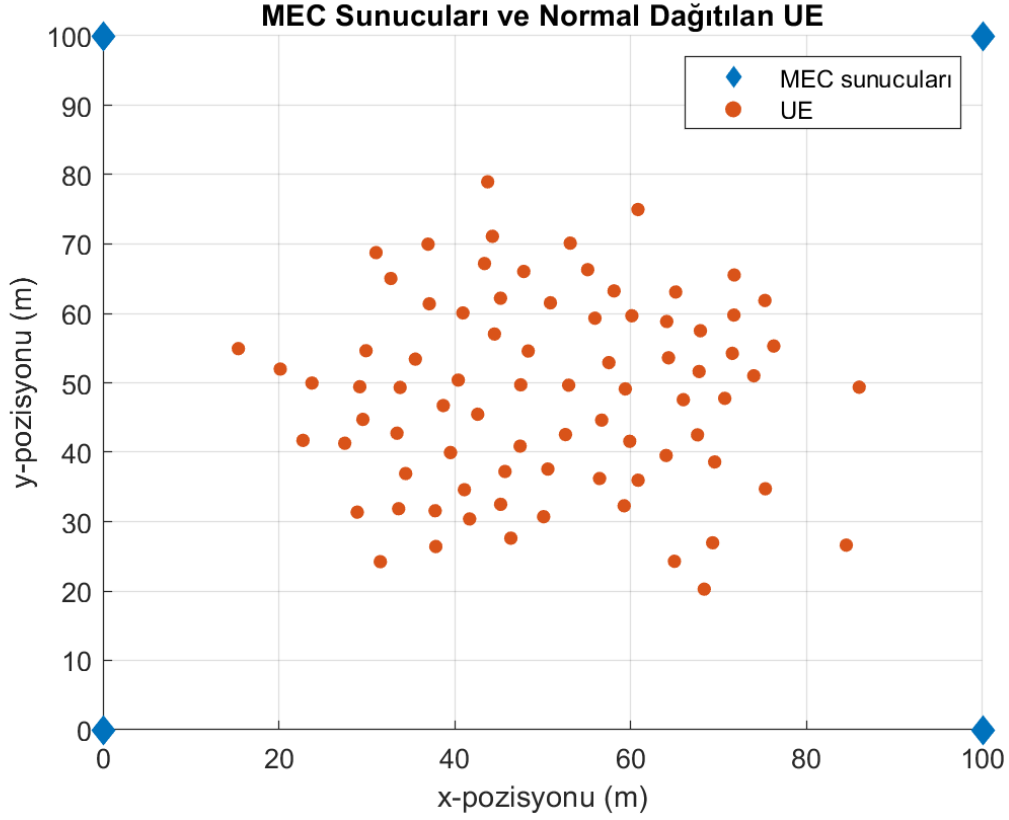
## 4. NÜMERİK SONUÇLAR

Bu bölümde, UE'lerin rastgele dağıtıldığı bir MEC ağında, işlem yoğunluğu esasına dayalı olarak görevlerin yerel hesaplama ve sunucuya aktarım için ayrılma oranlarının hesaplandığı ve farklı olasılık dağılımlarına sahip görev sıralarının fazlalık değerlerine GPD yaklaşımlarının etkinliğinin değerlendirildiği benzetim sonuçları verilmiştir.

Bu kapsamda,  $100 \times 100 \text{ m}^2$  alana sahip bir iç mekan dikkate alınarak dört adet MEC sunucu, bu alanın köşelerine yerleştirilmiştir. MEC ağı dahilinde UE'lerin konumları, alan merkezi ortalamalı normal dağılım ile belirlenmiştir. UE pozisyonlarının daha iyi gösterilebilmesi adına konumları çok yakın seçilen UE'ler birbirinden ayrılmıştır. MEC sunucularının ve UE'lerin konumlarını gösteren benzetim sonuçları sırasıyla Şekil 4.1 ve Şekil 4.2'de verilmiştir. Buna göre, tüm UE'lerin pozisyonları belirlendikten sonra her bir UE arasındaki mesafe ölçülerek coğrafi konfigürasyonu neredeyse aynı olan UE'lerin konumları yeniden hesaplanmaktadır.

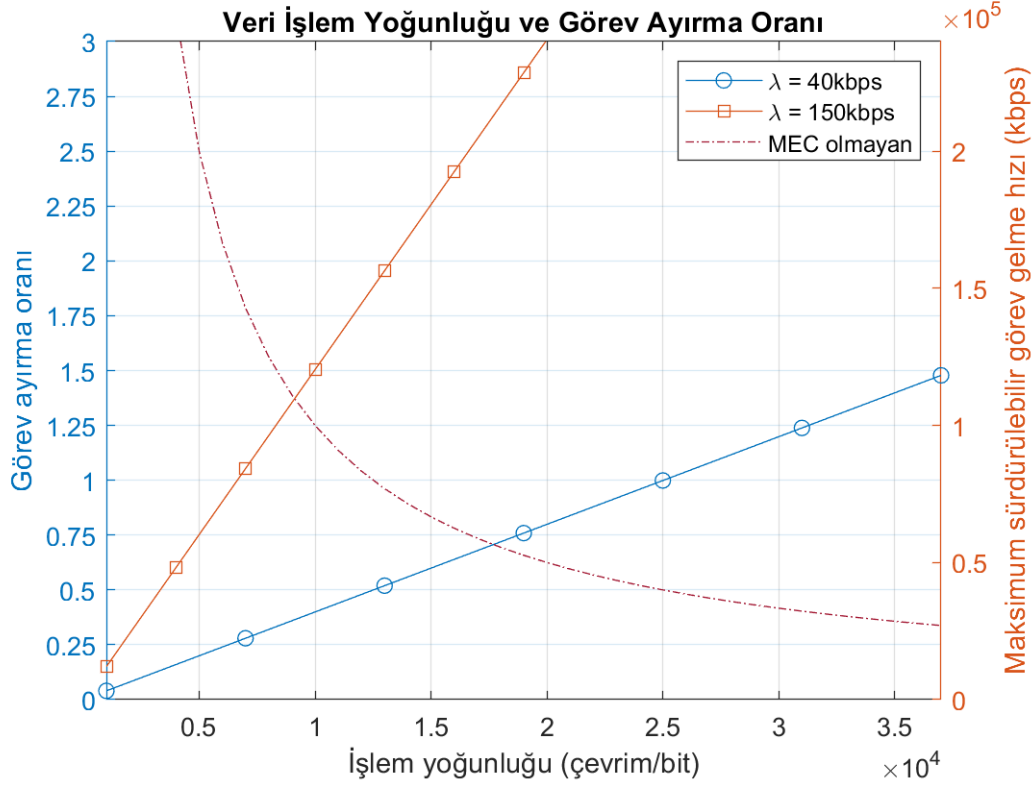


Şekil 4.1 MEC sunucularının ve UE'lerin konumları-1



Şekil 4.2 MEC sunucularının ve UE'lerin konumları-2

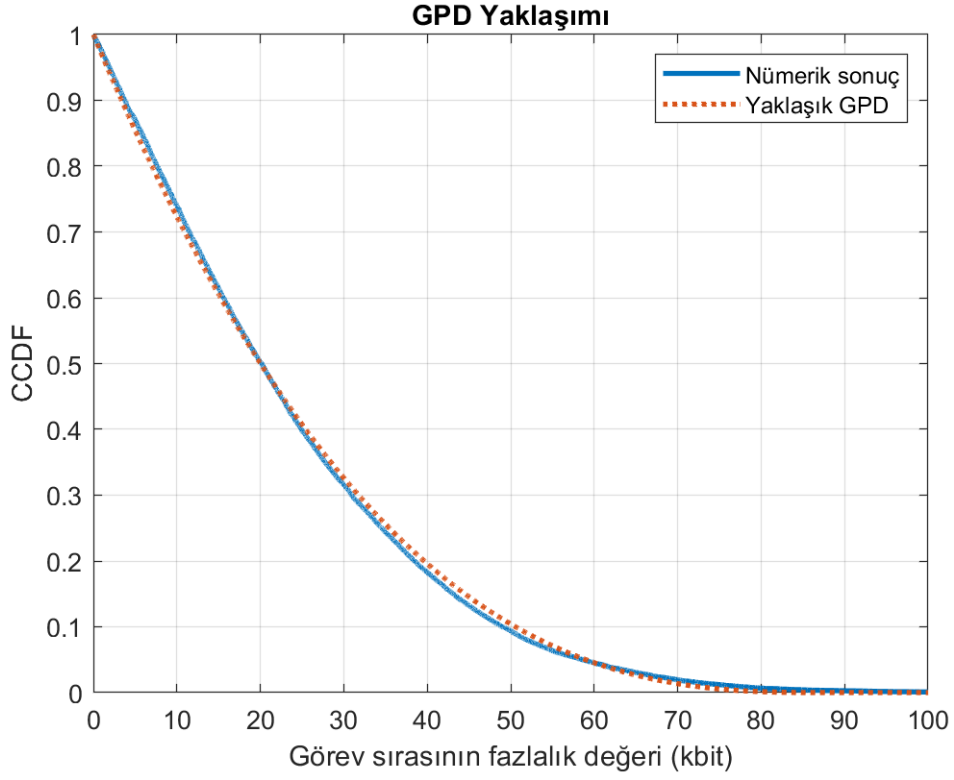
MEC sunucularının kullanılmadığı durumda, UE'lerde görev sıralarının kararlılığın korunabilmesi için yerel hesaplama hızının, görev gelme hızından daha yüksek olması gerekmektedir. UE'lerde yerel hesaplama ve görev aktarım sıralarının ilerleyen zaman ortalamalarının oranı, ayrılan görev miktarını göstermektedir. Burada verilen maksimum sürdürülebilir görev gelme hızı, görev aktarımı ve yerel hesaplama arasındaki sınırı temsil etmektedir. Bu bağlamda, görevlerin gelme hızı, maksimum izin verilen sınırı aştığında görevlerin yarısından fazlası hesaplanmak üzere sunucuya aktarılmaktadır. Bununla birlikte, daha yoğun veri işleme gerektiren veya UE'lere daha hızlı gelen görevler için daha büyük oranda sunucuya aktarım gerçekleştirilmektedir. İşlem yoğunluğuna karşı görev ayırma oranı Şekil 4.3'te verilmiştir.



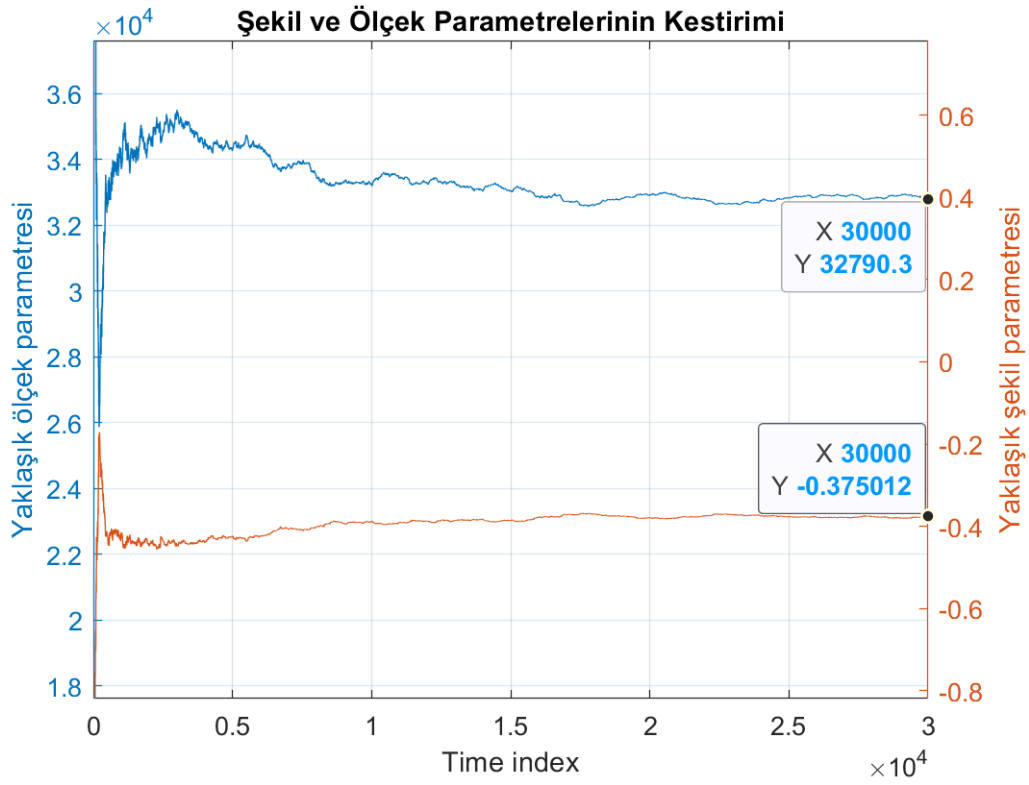
Şekil 4.3 İşlem yoğunluğu ve görev ayırma oranı

Güvenilir haberleşmenin sağlanabilmesi için görev sıralarının fazlalık değerleri hesaba katılmaktadır. MEC sunucuları tarafında da görev sıralarının uzunlukları hesaplama gecikmesi ile ilişkili olduğundan UE'lerdeki yerel hesaplama ve görev aktarım sıralarının ve MEC sunucularına aktarılan görev sıralarının fazlalık değerlerinin olasılık dağılımları yaklaşımı GPD fonksiyonları ile sağlanmaktadır. Fazlalık değerlere yaklaşım için GPD fonksiyonunda şekil ve ölçek parametrelerinin uygun olarak seçilmesi gerekmektedir. Buradan hareketle, UE'lerde ve MEC sunucularında mevcut tüm görev sıra uzunluklarına eşik değer tanımlanarak, farklı olasılık dağılımlarına sahip görev gelme olayları için fazlalık değerlerin deneysel ortalaması alınarak GPD fonksiyonlarının şekil ve ölçek parametreleri kestirilmiştir. Buna göre, normal dağılımlı fazlalık değerler için yaklaşık GPD fonksiyonu ve parametre kestirimi sonuçları sırasıyla Şekil 4.4 ve Şekil 4.5'te; Weibull dağılımlı fazlalık değerler için yaklaşık GPD fonksiyonu ve parametre kestirimi sonuçları sırasıyla Şekil 4.6 ve Şekil 4.7'de verilmiştir.

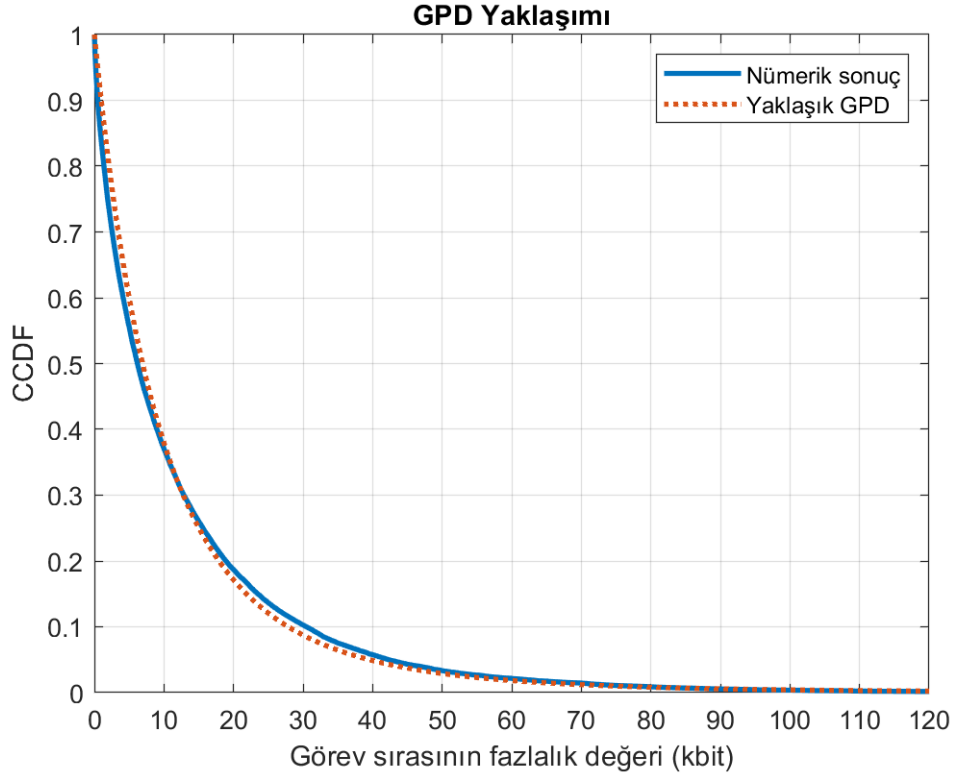




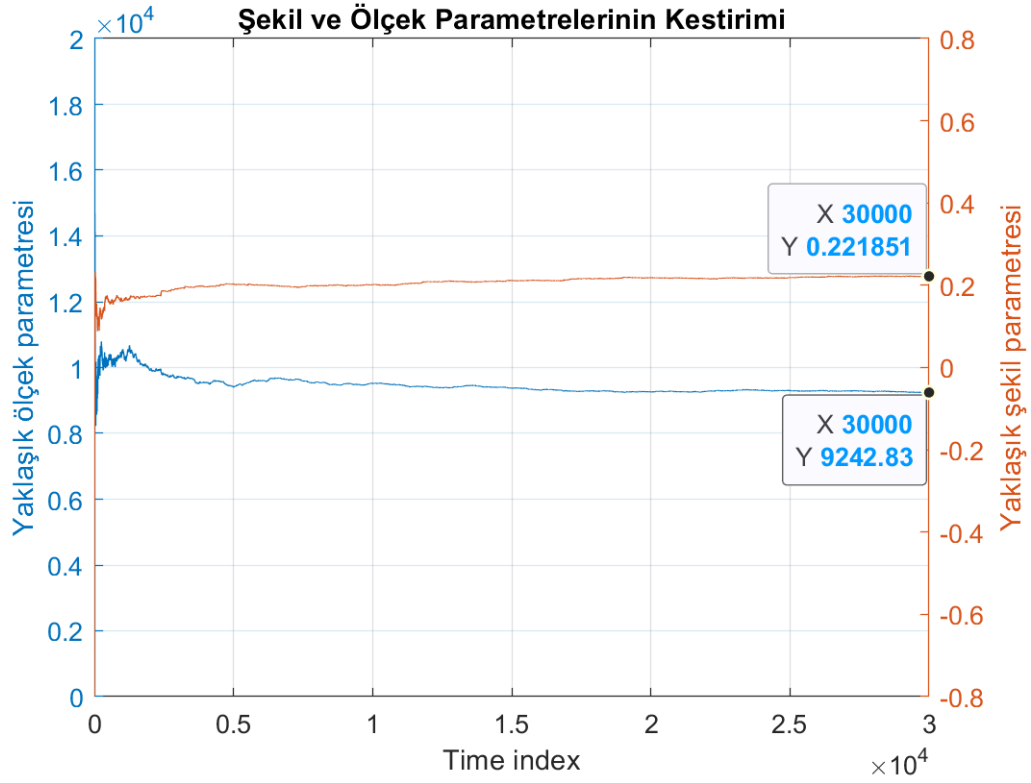
Şekil 4.4 Fazlalık değerlerin GPD ile yaklaşımı-1



Şekil 4.5 GPD şekil ve ölçek parametrelerinin kestirimi-1

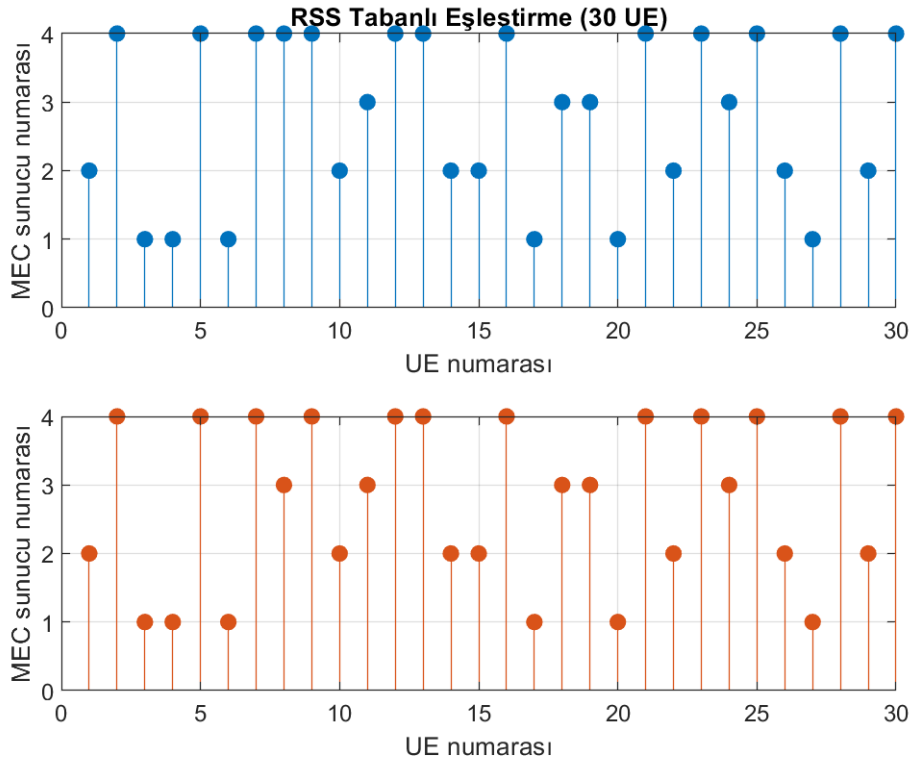


Şekil 4.6 Fazlalık değerlerin GPD ile yaklaşımı-2

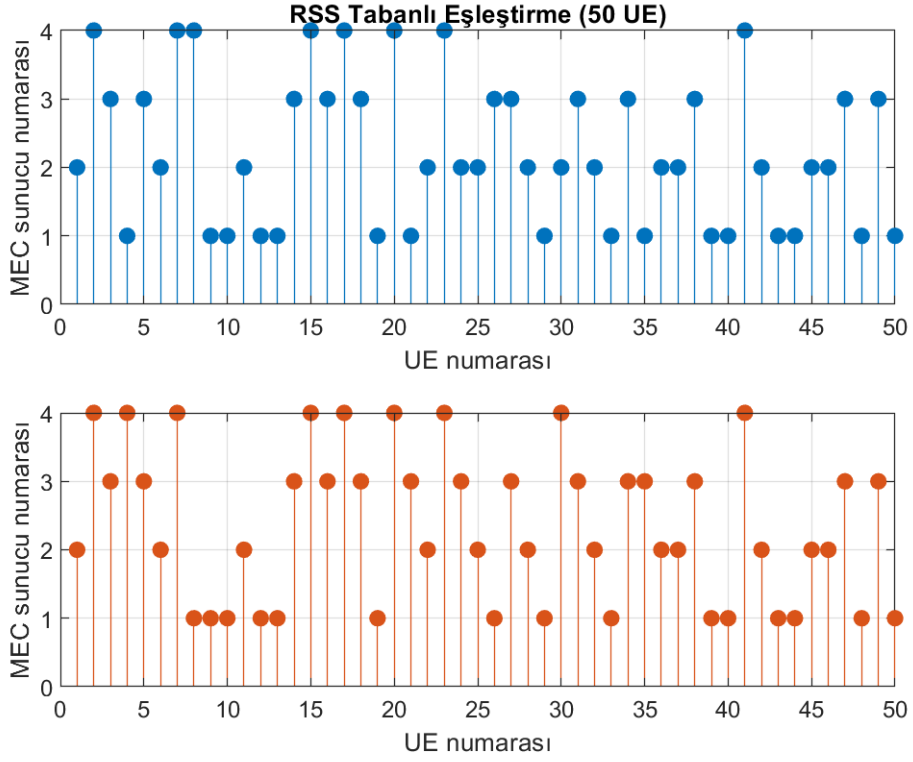


Şekil 4.7 GPD şekil ve ölçek parametrelerinin kestirimi-2

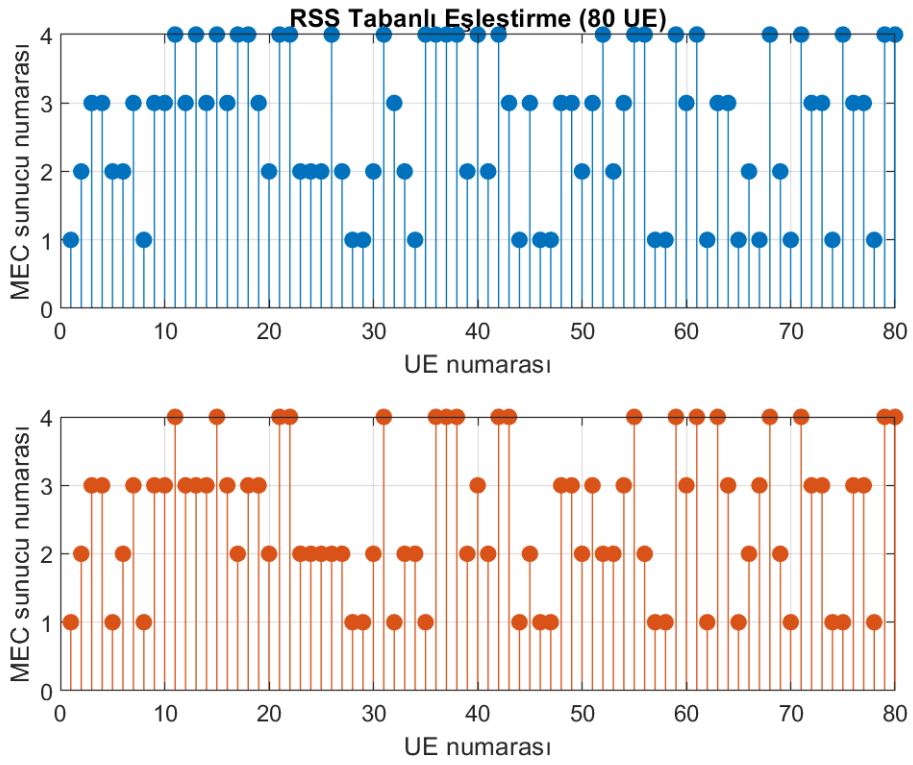
UE'ler ve MEC sunucuları, yalnızca alınan sinyal gücüne dayalı olarak eşleştirildiğinde sunucuların hesaplama kabiliyetleri ve mevcut iş yükleri hesaba katılmadığından düşük gecikme gereksinimini karşılamakta yetersiz kalmaktadır. Farklı miktarda işlem birimine sahip MEC sunucuların bulunduğu bir ortamda UE'lerin, alınan sinyal gücünün azalması nedeniyle azalan veri iletim hızına karşın daha güçlü hesaplama kabiliyetine sahip MEC sunucularıyla eşleştirilmesi sayesinde sunucuların iş yükleri dengelenmektedir. Fakat, işlem birimi miktarının her MEC sunucuda aynı olması durumunda, alınan sinyal gücünün azalması nedeniyle görev aktarımında ilave gecikmeler ortaya çıkmaktadır. Farklı miktarlardaki UE'ler, dört farklı MEC sunucu ile alınan sinyal gücüne (RSS) dayalı olarak eşleştirilmiştir. Ağ dahilinde normal dağıtılan UE'ler; ilk konumlarından ve UE'ler arasındaki minimum mesafelerin hesaplanarak güncellenen konumlarından MEC sunucuları ile eşleştirilmiştir. UE-sunucu eşleştirme sonuçları otuz, elli ve seksen UE için sırasıyla Şekil 4.8, Şekil 4.9 ve Şekil 4.10'da verilmiştir. Normal dağıtılan UE'lerin MEC sunucuları ile eşleştirme sonuçları otuz UE için ilk konumlarından (13, 4, 7, 6); yeniden konumlarından (12, 5, 7, 6), elli UE için ilk konumlarından (8, 12, 14, 16); yeni konumlarından (9, 14, 12, 15), seksen UE için ilk konumlarından (25, 23, 16, 16); yeni konumlarından (18, 23, 21, 18) olarak alınmıştır.



Şekil 4.8 Alınan sinyal gücüne dayalı eşleştirme-1



Şekil 4.9 Alınan sinyal gücüne dayalı eşleştirme-2



Şekil 4.10 Alınan sinyal gücüne dayalı eşleştirme-3

## 5. SONUÇ

Bu tez kapsamında, düşük gecikme ve yüksek güvenilirlik gereksinimlerini karşılamak üzere geliştirilen bir haberleşme-hesaplama sistem tasarımı referans alınmıştır. Referans alınan çalışmada, görev sıralarının uzunluklarına uygulanan olasılıksal kısıtlamaların formülasyonu, bu kısıtlamaların optimizasyon problemlerine dahil edilmesi ve çözüm algoritmaları incelenmiştir. UE'lerin ve MEC sunucularının eşleştirilmesi, gelen görevlerin yerel hesaplanmak ve MEC sunucularına aktarılmak üzere ayrılması ve MEC sunucularında hesaplama kaynaklarının tahsis edilmesi kapsamında önerilen algoritmalar incelenmiştir. Benzetim sonuçlarında, görev sıralarının fazlalık değerlerinin farklı olasılık dağılımlarına GPD fonksiyonu ile yaklaşımın etkinliği doğrulanmıştır. Alınan sinyal gücüne (RSS) dayanan eşleştirme yönteminin güvenilir ve düşük gecikmeli haberleşmede yetersiz olduğu gösterilerek, bu bağlamda incelenen çoktan bire eşleştirme yöntemine duyulan gereksinim doğrulanmıştır. Buna ek olarak, MEC sunucularının bulunmadığı ağlarda UE'lerin sınırlı hesaplama kabiliyetleri nedeniyle birim görevlerdeki işlem yoğunluğunun artmasına karşılık gelen sürdürülebilir görev gelme hızının doğal bir sınırı olduğu, buna karşın MEC sistemlerinde gelen görevlerin, işlem yoğunluğuna uygun olarak sunuculara aktarılmak üzere ayrılmasıyla daha yoğun veri işleme gerektiren görevlerin, sıralarda kararlılığın korunarak hesaplanabileceği gösterilmiştir.

Gelecek çalışmalarda, UE'lerin MEC sistemi dahilinde hareketli olması ve önerilen algoritmaların farklı miktarda UE'lerin bulunduğu sistem konfigürasyonlarında sönümlü kanal modelleri kullanılarak çalıştırılması, ağ dahilinde mevcut UE sayısının ve işlem yoğunluğunun artmasına karşı uçtan uca gecikmenin hesaplanması hedeflenmektedir.

## 6. KAYNAKLAR

- [1] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, A Survey on Mobile Edge Computing: The Communication Perspective, in *IEEE Communications Surveys & Tutorials*, 19(4):2322-2358, **2017**.
- [2] Y. Zhang, Mobile Edge Computing. In: Mobile Edge Computing, Simula SpringerBriefs on Computing, Vol 9, Springer, **2022**.
- [3] P. Mach and Z. Becvar, Mobile Edge Computing: A Survey on Architecture and Computation Offloading, in *IEEE Communications Surveys & Tutorials*, 19(3):1628-1656, **2017**.
- [4] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, Mobile Edge Computing A key technology towards 5G, ETSI, **2015**.
- [5] H. Sun, Y. Yu, K. Sha and B. Lou, mVideo: Edge Computing Based Mobile Video Processing Systems, in *IEEE Access*, 8:11615-11623, **2020**.
- [6] Y. Liu, M. Peng, G. Shou, Y. Chen and S. Chen, Toward Edge Intelligence: Multiaccess Edge Computing for 5G and Internet of Things, in *IEEE Internet of Things Journal*, 7(8):6722-6747, **2020**.
- [7] B. Soret, P. Mogensen, K. I. Pedersen and M. C. Aguayo-Torres, Fundamental tradeoffs among reliability, latency and throughput in cellular networks, *2014 IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, pages 1391-1396, **2014**.
- [8] M. Bennis, M. Debbah and H. V. Poor, Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale, in *Proceedings of the IEEE*, 106(10):1834-1853, **2018**.
- [9] Chih-Ping Li, Jing Jiang, W. Chen, Tingfang Ji and J. Smee, 5G ultra-reliable and low-latency systems design, *2017 European Conference on Networks and Communications (EuCNC)*, Oulu, Finland, pages 1-5, **2017**.
- [10] J. Kwak, Y. Kim, J. Lee and S. Chong, DREAM: Dynamic Resource and Task Allocation for Energy Minimization in Mobile Cloud Systems, in *IEEE Journal on Selected Areas in Communications*, 33(12):2510-2523, **2015**.
- [11] Y. Kim, J. Kwak and S. Chong, Dual-Side Optimization for Cost-Delay Tradeoff in Mobile Edge Computing, in *IEEE Transactions on Vehicular Technology*, 67(2):1765-1781, **2018**.
- [12] S. Mao, S. Leng, K. Yang, Q. Zhao and M. Liu, Energy Efficiency and Delay Tradeoff in Multi-User Wireless Powered Mobile-Edge Computing Systems, *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, pages 1-6, **2017**.
- [13] Z. Jiang and S. Mao, Energy Delay Tradeoff in Cloud Offloading for Multi-Core Mobile Devices, in *IEEE Access*, 3:2306-2316, **2015**.

- [14] C. -F. Liu, M. Bennis, M. Debbah and H. V. Poor, Dynamic Task Offloading and Resource Allocation for Ultra-Reliable Low-Latency Edge Computing, in *IEEE Transactions on Communications*, 67(6):4132-4150, **2019**.
- [15] S. Coles, An Introduction to Statistical Modeling of Extreme Values. London, U.K.: Springer, **2001**.
- [16] A. E. Roth and M. A. O. Sotomayor, Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis. New York, NY, USA: Cambridge Univ. Press, **1992**.
- [17] M. J. Neely, Stochastic Network Optimization With Application to Communication and Queueing Systems. San Rafael, CA, USA: Morgan and Claypool, **2010**.
- [18] J. G. Andrews *et al.*, What Will 5G Be?, in *IEEE Journal on Selected Areas in Communications*, 32(6):1065-1082, **2014**.
- [19] ETSI, Mobile-edge computing—Introductory technical white paper, White Paper, Sophia Antipolis, France, **2014**.
- [20] T.D. Burd and R.W. Brodersen, Processor design for portable systems. in *J. VLSI Signal Process. Syst.*, 13(2–3):203-221, **1996**.
- [21] M. Bennis, M. Debbah and H. V. Poor, Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale, in *Proceedings of the IEEE*, 106(10):1834-1853, **2018**.
- [22] G. Durisi, T. Koch and P. Popovski, Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets, in *Proceedings of the IEEE*, 104(9):1711-1726, **2016**.
- [23] M. Chen, W. Saad and C. Yin, Virtual Reality Over Wireless Networks: Quality-of-Service Model and Learning-Based Resource Management, in *IEEE Transactions on Communications*, 66(11):5621-5635, **2018**.
- [24] A. Al-Shuwaili and O. Simeone, Energy-Efficient Resource Allocation for Mobile Edge Computing-Based Augmented Reality Applications, in *IEEE Wireless Communications Letters*, 6(3):398-401, **2017**.
- [25] S. -W. Ko, K. Han and K. Huang, Wireless Networks for Mobile Edge Computing: Spatial Modeling and Latency Analysis, in *IEEE Transactions on Wireless Communications*, 17(8):5225-5240, **2018**.
- [26] C. You, K. Huang, H. Chae and B. -H. Kim, Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading, in *IEEE Transactions on Wireless Communications*, 16(3):1397-1411, **2017**.
- [27] N. Cheng *et al.*, Air-Ground Integrated Mobile Edge Networks: Architecture, Challenges, and Opportunities, in *IEEE Communications Magazine*, 56(8):26-32, **2018**.
- [28] F. Zhou, R. Q. Hu, Z. Li and Y. Wang, Mobile Edge Computing in Unmanned Aerial Vehicle Networks, in *IEEE Wireless Communications*, 27(1):140-146, **2020**.

- [29] S. Cao, X. Tao, Y. Hou and Q. Cui, An energy-optimal offloading algorithm of mobile computing based on HetNets, *2015 International Conference on Connected Vehicles and Expo (ICCVE)*, Shenzhen, China, pages 254-258, **2015**.
- [30] Y. Zhao, S. Zhou, T. Zhao and Z. Niu, Energy-efficient task offloading for multiuser mobile cloud computing, *2015 IEEE/CIC International Conference on Communications in China (ICCC)*, Shenzhen, China, pages 1-5, **2015**.
- [31] C. You and K. Huang, Multiuser Resource Allocation for Mobile-Edge Computation Offloading, *2016 IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, USA, pages 1-6, **2016**.
- [32] Maofei Deng, Hui Tian and Bo Fan, Fine-granularity based application offloading policy in cloud-enhanced small cell networks, *2016 IEEE International Conference on Communications Workshops (ICC)*, Kuala Lumpur, pages 638-643, **2016**.
- [33] Y. Mao, J. Zhang, S. H. Song and K. B. Letaief, Power-Delay Tradeoff in Multi-User Mobile-Edge Computing Systems, *2016 IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, USA, pages 1-6, **2016**.
- [34] J. D. C. Little, A proof for the queuing formula:  $L = \lambda W$ , <http://fisherp.scripts.mit.edu/wordpress/wp-content/uploads/2015/11/ContentServer.pdf> (Erişim tarihi: **4 Nisan 2023**).
- [35] C. -F. Liu and M. Bennis, Ultra-Reliable and Low-Latency Vehicular Transmission: An Extreme Value Theory Approach, in *IEEE Communications Letters*, 22(6):1292-1295, **2018**.
- [36] F. Raissi, S. Yangui and F. Camps, Autonomous Cars, 5G Mobile Networks and Smart Cities: Beyond the Hype, *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Napoli, Italy, pages 180-185, **2019**.
- [37] Z. Q. Jaber and M. I. Younis, Design and Implementation of Real Time Face Recognition System (RTFRS), in *International Journal of Computer Applications*, 94(12):15-22, **2014**.
- [38] M. Simsek, A. Aijaz, M. Dohler, J. Sachs and G. Fettweis, 5G-Enabled Tactile Internet, in *IEEE Journal on Selected Areas in Communications*, 34(3):460-473, **2016**.
- [39] A. Varghese and D. Tandur, Wireless requirements and challenges in Industry 4.0, *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, India, pages 634-638, **2014**.
- [40] J. Arata et al., A remote surgery experiment between Japan and Thailand over Internet using a low latency CODEC system, in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, Rome, Italy, pages 953-959, **2007**.
- [41] E. Bodine-Baron, C. Lee, A. Chong, B. Hassibi, A. Wierman, Peer Effects and Stability in Matching Markets. in G. Persiano,(Eds) *Algorithmic Game Theory*. SAGT 2011. Lecture Notes in Computer Science, vol 6982. Springer, Berlin, Heidelberg, **2011**.