# Latency-Aware Service Function Chain Placement in 5G Mobile Networks

## 1. Introduction

This paper focuses on optimizing Service Function Chain (SFC) placement in 5G networks to minimize latency. It uses Integer Linear Programming (ILP) to determine the best placement for Virtualized Service Functions (VSFs) in a network while ensuring low end-to-end (E2E) latency.

### Key Objectives:

1. Minimize latency by placing VNFs in the fastest possible location.  
2. Ensure optimal resource usage without overloading any network node.  
3. Guarantee QoS for real-time applications like AR/VR, gaming, and remote surgery.

## 2. Network Model

The network consists of three types of nodes where VNFs can be placed:

• DU (Distributed Unit): Closest to users with the lowest latency but limited processing power.

• CU (Centralized Unit): Intermediate location balancing speed and computing resources.

• 5GC (5G Core): Far from users with the highest computing power but also the highest latency.

Each User Equipment (UE) sends an SFC request, which consists of multiple VNFs. Where these VNFs are placed affects both latency and network congestion.

## 3. ILP-Based Optimization Model

ILP is used to find the best placement of VNFs that minimizes total end-to-end latency.

### Decision Variables:

Define if VNF of user is placed at node, otherwise 0.

### Objective Function (Minimize Latency):

Minimize: ∑(Latency \* Placement)

### Constraints:

1. Each UE must be assigned to exactly one DU.  
2. Each VNF must be placed exactly once.  
3. Network Resource Constraints (CPU, Memory).

## 4. Numerical Example: Solving SFC Placement Using ILP

Problem Setup:  
Two users (UE1, UE2) request SFCs:  
• UE1 needs Firewall (FW) and Load Balancer (LB).  
• UE2 needs Intrusion Detection System (IDS) and Video Optimizer (VO).

Latency values for each node:  
• DU latency = 1 ms  
• CU latency = 3 ms  
• 5GC latency = 8 ms

### ILP Solution:

• Firewall (FW) is placed at DU → Latency = 1 ms.  
• Load Balancer (LB) is placed at CU → Latency = 3 ms.  
• Intrusion Detection System (IDS) is placed at CU → Latency = 3 ms.  
• Video Optimizer (VO) is placed at 5GC → Latency = 8 ms.  
Total Latency Calculation: 1 + 3 + 3 + 8 = 15 ms

## 5. Key Takeaways from Paper 1

• ILP finds the optimal SFC placement for minimizing latency.  
• Placing VNFs at DUs provides the lowest latency, but resource constraints limit edge placements.  
• ILP is computationally expensive and does not handle mobility well.

## 6. Why MILP is Needed (Covered in Paper 2)

ILP struggles with mobility. Paper 2 introduces MILP (Mixed-Integer Linear Programming), which:  
• Optimizes both latency and mobility by allowing fractional VNF placement.  
• Minimizes handover disruptions by reducing unnecessary migrations.

# Paper 2: Latency and Mobility-Aware Service Function Chain Placement in 5G Networks

## 1. Introduction

Paper 2 extends Paper 1 by considering mobility in addition to latency. Instead of ILP, it uses MILP, allowing fractional VNF placement to reduce migration costs.

## 3. MILP-Based Optimization Model

### Decision Variables:

Define as the fraction of VNF for user placed at node (between 0 and 1).

### Objective Function:

Minimize: (Latency + Migration Cost)

## 4. Numerical Example: Solving SFC Placement Using MILP

### MILP Optimization Strategy:

• Minimize latency while ensuring minimal migrations.  
• Allow VNFs to be split across nodes to balance latency and cost.

## 5. Heuristic Algorithm (Fast Approximation of MILP)

MILP is computationally expensive, so a heuristic approach finds near-optimal solutions faster.  
• Sort users by latency sensitivity.  
• Assign VNFs to the lowest-latency node.  
• Minimize migrations by keeping VNFs stable.

## 7. Final Comparison of Paper 1 (ILP) vs. Paper 2 (MILP)

• ILP focuses only on latency, while MILP optimizes both latency and mobility.  
• MILP allows VNFs to be split across nodes, while ILP requires a single placement.  
• Heuristic solutions provide a fast, near-optimal alternative to MILP.