2-1-2020

# LTE network slicing and resource trading schemes for machine-to-machine communications

Sylvia Nader Gendy

The American University in Cairo

The School of Sciences and Engineering

# LTE NETWORK SLICING AND RESOURCE TRADING SCHEMES FOR MACHINE-TO-MACHINE COMMUNICATIONS

A Thesis Submitted to

The Department of Electronics and Communications Engineering

in partial fulfillment of the requirements for
the degree of Master of Science

Sylvia Nader Farah Gendy
SID 800120119

under the supervision of Dr. Yasser Gadallah
August 2019

Approval sheet goes here

# DEDICATION

*To the soul of Nader Gendy, an exceptionally loving father and gifted engineer*

# ABSTRACT

The Internet of Things (IoT) is envisioned as the future of human-free communications. IoT relies on Machine-to-Machine (M2M) communications rather than conventional Human-to-Human (H2H) communications. It is expected that billions of Machine Type Communication Devices (MTCDs) will be connected to the Internet in the near future. Consequently, the mobile data traffic is poised to increase dramatically. Long Term Evolution (LTE) and its subsequent technology LTE-Advanced (LTE-A) are the candidate carriers of M2M communications for the IoT purposes. Despite the significant increase of traffic due to IoT, the Mobile Network Operators (MNOs) revenues are not increasing at the same pace. Hence, many MNOs have resorted to sharing their radio resources and parts of their infrastructures, in what is known as Network Virtualization (NV).

In the thesis, we focus on "slicing" in which an operator known as Mobile Virtual Network Operator (MVNO), does not own a spectrum license or mobile infrastructure, and relies on a larger MNO to serve its users. The large licensed MNO divides its spectrum pool into *slices*. Each MVNO reserves one or more slice(s). There are 2 forms of slice scheduling: Resource-based in which the slices are assigned a portion of radio resources or Data rate-based in which the slices are assigned a certain bandwidth.

In the first part of this thesis we present different approaches for adapting resource-based NV, Data rate-based NV to Machine Type Communication (MTC). This will be done in such a way that resources are allocated to each slice depending on the delay budget of the MTCDs deployed in the slice and their payloads. The adapted NV schemes are then simulated and compared to the Static Reservation (SR) of radio resources. They have all shown an improved performance over SR from deadline missing perspective.

In the second part of the thesis, we introduce a novel resource trading scheme that allows sharing operators to trade their radio resources based on the varying needs of their clients with time. The Genetic Algorithm (GA) is used to optimize the resource trading among the virtual operators. The proposed trading scheme is simulated and compared to the adapted schemes from the first part of the thesis. The novel trading scheme has shown to achieve significantly better performance compared to the adapted schemes.

# TABLE OF CONTENTS

# LIST OF ACRONYMS

| | |
|---|---|
| 3G | Third generation of mobile communications |
| 3GPP | 3rd Generation Partnership Project |
| 4G | Fourth generation of mobile communications |
| 5G | Fifth generation of mobile communications |
| BDSM | Bargaining Based Dynamic Spectrum Management |
| BLER | Block Error Rate |
| CN | Core Network |
| CQI | Channel Quality Indicator |
| D2D | Device to Device |
| DMR | Deadline Missing Ratio |
| DMR | Deadline Missing Ratio |
| EDF | Earliest Deadline First |
| eNB | eNodeB |
| EPC | Evolved Packet Core |
| E-UTRAN | Evolved UMTS Terrestrial Radio Access Network |
| FCC | Federal Communications Commission |
| FDPS | Frequency Domain Packet Scheduling |
| GA | Genetic Algorithm |
| GBR | Guaranteed Bit Rate |
| GWCN | Gateway Core Network |
| H2H | Human To Human |
| HARQ | Hybrid Automatic Repeat Request |
| IMT-2020 | International Mobile Telecommunications-2020 |
| IoT | Internet of Things |
| IP | Internet Protocol |
| ITU | International Telecommunications Union |
| LB | Load Balancing |
| LTE | Long Term Evolution |
| LTE-A | LTE-Advanced |
| M2M | Machine To Machine |
| MCS | Modulation and Coding Scheme |
| MIMO | Multiple-Input Multiple-Output |

| | |
|---|---|
| MNO | Mobile Network Operator |
| MOCN | Multi-Operator Core Network |
| MTCD | Machine Type Communication Device |
| MTCG | MTC Gateway |
| MVNO | Mobile Virtual Network Operator |
| NV | Network Virtualization |
| NVS | Network Virtualization Substrate |
| OAM | Operations, Administration and Management |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| PRB | Physical Resource Block |
| PRR | Partial Resource Reservation |
| QAM | Quadrature amplitude modulation |
| QoS | Quality of Service |
| QPSK | Quadrature Phase Shift Keying |
| RAN | Radio Access Network |
| RE | Resource Element |
| RENEV | Resources nEgotiation for NEtwork Virtualization Analysis |
| SC-FDMA | Single Carrier-FDMA |
| SC-PTM | Single-cell Point-to-Multipoint |
| SINR | Signal-to-Inference-and-Noise Ratio |
| SLA | Service Level Agreement |
| SNR | Signal-to-Noise Ratio |
| SR | Static Reservation |
| STI | Slicing Time Interval |
| STI | Slicing Time Interval |
| TDPS | Time Domain Packet Scheduling |
| TODA | Truthful Online Double Auction |
| TRUST | TRuthful doUble Spectrum aucTions |
| TTI | Transmission Time Interval |
| UE | User Equipment |
| UMTS | Universal Mobile Telecommunication System |
| V2X | Vehicle-to-Everything |
| VNF | Virtual Network Functions |

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1: Introduction

The expansion of mobile coverage worldwide and the constantly growing popularity of the mobile applications have led to a dramatic increase in the data traffic of smart phones and other handheld devices. Cisco Global Mobile Data Traffic 2017-2022 Forecast Report estimates that 12.3 billion mobile devices will be connected by 2022. The report also estimates that the data traffic will witness a 7-fold increase from 2017 to 2022 [1].

## 1.1 Network Slicing and Resource Trading

The Mobile Network Operators (MNOs) are facing various technical and financial challenges to expand their infrastructures to adapt to the increase in data traffic. Their revenues are not increasing at the same rate. So, they resorted to sharing their resources. The most basic form of sharing is passive sharing, in which the MNOs share physical resources such as the datacenters or power generators. The other more complex form is active sharing where they share radio communication elements such as radio frequencies or base stations. The solutions offered to allow the operators to share the same infrastructure and manage their resources efficiently are known as Network Virtualization (NV).

The 3rd Generation Partnership Project (3GPP) has specified two active sharing architectures: Gateway Core Network (GWCN) and Multi-Operator Core Network (MOCN). In GWCN, the MNOs share both the Radio Access Network (RAN) and the Core Network (CN). Whereas, in MOCN, they only share the RAN [2].

The 3GPP has also defined 5 deployments scenarios [3]. In the first scenario, the operators share the same RAN elements without sharing their radio resources i.e. multiple MNOs share the same eNodeB (eNB) but each MNO owns its own frequency license. In the second scenario, each MNO has its own frequency license and covers a separate geographic area but collectively the operators cover the whole country. In the third scenario, an MNO allows another one to use its coverage in a certain geographic area; outside of this MNO's area, they operate independently. In the fourth scenario, MNOs share the same CN but own

Individual RANs. In the last scenario, MNOs share their spectra. This scenario can be implemented in two ways: multiple large MNOs, each having its license, regroup their radio resources and reallocate them based on their needs or small MNOs rely on large MNOs that own frequency licenses to serve their user. The small MNOs are usually specialized in a certain service such as video streaming, and they are known as Mobile Virtual Network Operators (MVNOs).

The last scenario of NV implementation (spectrum sharing) is the scope of the thesis. We focus on network slicing; it is a form of network virtualization that allows multiple virtual networks to share the same physical network. An MNO may choose to divide its network into multiple slices each serving a different application and has different needs.

The other form of spectrum sharing is radio resource trading. Many studies in the literature have been conducted to offer solutions for the sharing MNOs to trade or bid on radio resources. Most commonly, the licensed MNO acts as the "seller" and the MVNOs bidding on the radio resources act as the "buyers", the selling MNO offers its unused resources for trade. The buying MNOs bid on the resources that they need to serve their end users. They make the trading decisions based on their valuations of the radio resources which translate to the price that they are willing to pay/receive. The objective of any trade is to maximize the utility of both the seller and the buyer

## 1.2 Machine-to-Machine Communications

Many states are planning to build new cities with smart transportation, waste management, health care, water supply and power grids. The cities management will rely on data collected by sensors and other devices [4]. Billions of sensors and other small deployable devices are envisioned to be connected to the Internet with no human intervention in what is known as the Internet of Things (IoT). IoT will offer a variety of services to individuals, organizations and communities. It relies on Machine-to-Machine (M2M) communication rather than Human-to-Human (H2H) communication. The devices used in M2M communication are known as Machine-Type Communications Devices (MTCDs). MTCDs types include surveillance cameras used in security systems, alarm devices used in critical emergencies such as fires or floods and monitoring sensors used to track different environmental conditions such as temperature, pressure or water level. The Cisco forecast 2017-2022 report predicts that, out of 12.3 billion devices, 3.9 billion MTCDs will be connected in 2022 [1]. Long Term Evolution (LTE) is the candidate platform for IoT deployment due to its ubiquity and long reach. Many studies have been conducted to adapt the current LTE and its subsequent technology LTE-Advanced (LTE-A) to M2M traffic needs. The MTCDs deployment challenges include the battery life, processing power and wireless ranges of the devices. The other obstacle is standardization; European Telecommunication Standards Institute (ETSI) and other stakeholders have been working on a unified M2M architecture that consists of two domains namely Network/application domain and device/gateway domain [4].

We believe that NV can help MNOs reduce the capital and operational expenses of the IoT deployments. Moreover, sharing the radio spectrum allows the MNOs to overcome the inherited spectrum scarcity issues. Most of the NV schemes have been designed to maximize the overall throughput of the system without affecting the individual performances of the sharing operators. These requirements are suitable for H2H traffic. Additional parameters need to be considered to adapt the NV schemes to M2M communications.

## 1.3 Problem Statement

In the thesis, we assume that a large MNO slices its pool of radio resources and allocates the slices to multiple MVNOs. To allocate the radio resources to the slices, we use some of the known NV schemes such as Network Virtualization Substrate (NVS) [5] and Partial Resource Reservation (PRR) [6] . However, as the NV schemes were designed for H2H traffic, NVS and PRR must be adapted to the needs of mission-critical M2M deployments. Critical M2M traffic is mostly event-driven and events such as those encountered in emergency situations have very strict delay budgets. Therefore, the slice scheduling is designed to minimize the probability of deadline missing.

Moreover, the MTCDs are involved in different applications each having its own characteristics and needs. Hence, we allow the sharing MVNOs to trade Physical Resource Blocks (PRBs) from their allocated slices. Most of the trading schemes in the literature focus on the resources prices to make the trading decisions. We introduce a novel trading scheme that allows the slices to estimate their valuations for the traded goods as a function of their qualities or more specifically their abilities to minimize the probability of deadline missing. The adapted NV schemes and the novel trading schemes are evaluated based on their performances in terms of their abilities to minimize the percentage of packets missing the deadline.

## 1.4 Thesis Contribution and Structure

The first part of the thesis work is to adapt NVS and PRR to the delay sensitive nature of M2M traffic. The slice scheduling in NVS has 2 models: Resource-based and Data rate-based scheduling. So, we present different approaches to adapt Resource-based NVS, Data rate-based NVS and PRR to massive mission-critical MTCDs deployment.

The second part is to introduce a novel resource trading that allows the operators to trade the resources based on their channel conditions. The trading is initiated at constant time intervals know as Slicing Time Intervals (STIs). There are 2 implementations for the trading schemes: static and dynamic. In the static implementation, the STI is pre-defined and does not change for the duration of the communication. In the dynamic implementation, the STI changes dynamically depending on the traffic.

The main contributions of the thesis can therefore be summarized as follows:

1. Adapt some of the well-known network slicing schemes to M2M communication.
2. Formulate the resource trading optimization problem.
3. Introduce a novel network resources trading scheme under static and dynamic operating conditions.

The comparison between resource-based NVS, Data rate-based NVS and PRR has been published in [7]. The trading scheme has been accepted for publication later this year [8].

The rest of this thesis is organized as follows. In chapter 2, we present the necessary background information about LTE fundamentals, M2M communication and network slicing. We also review the previous work from the literature about network virtualization and radio resources trading. In chapter 3, Resource-based NVS, Data rate-based NVS and PRR are explained. Multiple techniques will be introduced to adapt these schemes to mission critical M2M deployments. In addition, the novel trading algorithm is introduced as well as the dynamic STI calculation algorithm.

In chapter 4, the adapted Resource-based, Data rate-based NVS and PRR schemes simulations are evaluated. The experimental results are analyzed in depth from delay efficiency perspective. Then, the simulation of the trading scheme is presented, and its performance is compared to the other NV schemes. In chapter 5, the thesis work is concluded, and potential future work is discussed.

# Chapter 2: Background and Related Work

## 2.1 Introduction

In this chapter, we present the LTE fundamentals that support the work in this thesis such as radio resource grid and allocation. Then we go through some background information about M2M communication and the main characteristics of machine type communications, traffic and architecture. We also discuss the network virtualization and slicing focusing on its objectives and requirements. We then review the studies published in the literature regarding network virtualization and radio resources trading. We aim to highlight the contribution of the thesis as compared to the previous studies.

## 2.2 Background

### 2.2.1 Long Term Evolution (LTE)

In the early 2000s, the third generation of mobile network (3G) was introduced. The most known 3G system is the Universal Mobile Telecommunication System (UMTS) developed by the 3GPP. However, the mobile phones industry had been thriving; the smart phones conquered the market. The mobile traffic became data centric and the deployed 3G networks became congested. 3GPP introduced LTE in Release 8 in December 2008. It was developed as a long-term evolution of the UMTS. By the end of 2009, LTE was launched in Europe [9].

In LTE, the Radio Access Network (RAN) is known as evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and the Core Network (CN) is known as Evolved Packet Core (EPC). The RAN has one component, which works as the base station, it is known as eNodeB (eNB) [9]. The EPC is Internet Protocol (IP) based; a User Equipment (UE) is assigned an IP when it is first turned ON and it keeps it until it gets switched off. This is the major difference between 3G and 4G. In 3G, the UE is assigned an IP when it initiates a connection request, and only lasts for the duration of the connection. LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink direction and Single Carrier-FDMA (SC-FDMA) in the uplink direction.

Starting Release 9, non-contiguous resource allocation is supported in SC-FDMA [10]. In Release 10, 3GPP has released the specifications for an enhanced LTE, known as LTE-Advanced (LTE-A). LTE-A added the support of Machine Type Communication (MTC). It also added the support of Carrier Aggregation (CA); in which carriers can be regrouped to achieve higher data rates [10].

Finally, 5G is now being launched in many countries. The main features of 5G include: peak date rate of 20 Gbits/s, user data rate can reach 100 Mbits/s, mobility of 500 Km/h, latency of 1 ms, and connection density up to $10^6$ $\text{connections/km}^2$ [11].

As explained earlier, LTE uses OFDMA in the downlink and SC-FDMA in the uplink direction. As shown in Figure 2.1, an LTE Frame is equivalent to $10ms$; it is divided into 10 Subframes (1 ms each). The subframe is also known as Transmission Time Interval (TTI). Each subframe is divided into 2 time slots (0.5 $ms$ each). Each time slot can carry up to 7 symbols or Resource Elements (REs). In the frequency domain, the grid is divided into Physical Resource Blocks (PRBs). Each PRB is equivalent to 180 KHz in the frequency domain and 0.5 ms in the time domain. Each PRB is divided into 12 subcarriers (15 KHz each). So, each PRB carries 12 subcarriers ×7 symbols= 84 REs [12]. The Number of bits per RE depends on the modulation; for QPSK 2 bits per RE, for 16QAM 4 bits per RE and for 64QAM 6 bits per RE. As the number of PRBs assigned to a user increases, the data rate increases. It is the scheduling process that determines the number of PRBs to be assigned to each user.

LTE supports multiple channel bandwidths: 1.4, 3, 5, 10, 15, and 20 MHz; the number of PRBs per channel is listed in Table 2.1. At each TTI, the scheduler allocates the available PRBs to active users. The used scheduler can be channel-independent or channel-dependent. Channel-independent scheduling does not monitor the channel quality whereas channel-dependent scheduling depends on the Channel Quality Indicator (CQI) reports received from the active users in the network to make the resource allocation decision [13]. The LTE Release 10 defined 15 CQI levels; the CQI is computed as a function of the Signal-to-Inference-and-Noise Ratio (SINR). The CQI is then mapped to the modulation and coding scheme (MCS). Table 2.2 lists the CQI mapping for 10% Block Error Rate (BLER) [14].

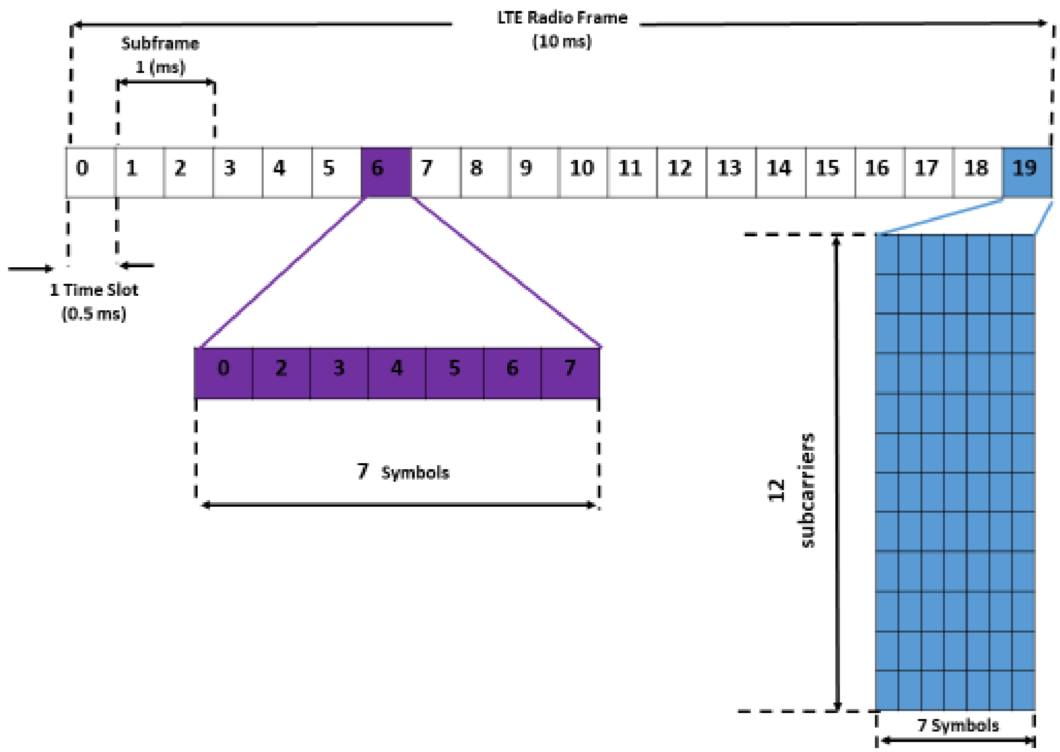


**Figure 2.1 LTE Frame Structure**

**Table 2.1 LTE Channel Bandwidths**

| Channel Bandwidth | Number of PRBs |
|---|---|
| 1.4 | 6 |
| 3 | 15 |
| 5 | 25 |
| 10 | 50 |
| 15 | 75 |
| 20 | 100 |

The scheduling decision does not only rely on the CQI; it takes into consideration other parameters including [15]:

- The data buffered in each device and its queuing time; this is estimated by the Buffer Status Report (BSR).
- The message retransmission handled by the Hybrid Automatic Repeat Request (HARQ) protocol; as it gives priority to retransmitted data.

- The Quality of Service (QoS) restrictions for the data sent by the active users. For example: Guaranteed Bit Rate (GBR) for video data.
- PRBs assigned to one user need to be contiguous; however, starting R9, non-contiguous allocation is supported in the uplink direction.
- Transmission Power Limitation.

**Table 2.2 4 -Bit CQI Mapping for 10% Block Error Rate (BLER)**

| CQI | Range of SINR (dB) | Modulation | Code Rate (×1024) | Spectral Efficiency |
|---|---|---|---|---|
| 0 | SINR< -6.936 | Out of Range | | |
| 1 | -6.936 ≤ SINR< -5.146) | QPSK | 78 | 0.1523 |
| 2 | -5.147 ≤ SINR <-3.18 | QPSK | 120 | 0.2344 |
| 3 | -3.18≤ SINR <-1.253 | QPSK | 193 | 0.3770 |
| 4 | -1.253≤ SINR < 0.761 | QPSK | 308 | 0.6016 |
| 5 | 0.761≤ SINR <2.699 | QPSK | 449 | 0.8770 |
| 6 | 2.699≤ SINR < 4.694 | QPSK | 602 | 1.1758 |
| 7 | 4.694≤ SINR < 6.525 | 16 QAM | 378 | 1.4766 |
| 8 | 6.525≤ SINR <8.573 | 16 QAM | 490 | 1.9141 |
| 9 | 8.573≤ SINR <10.366 | 16 QAM | 616 | 2.4063 |
| 10 | 10.366≤ SINR <12.289 | 64 QAM | 466 | 2.7305 |
| 11 | 12.289≤ SINR <14.173 | 64 QAM | 567 | 3.3223 |
| 12 | 14.173≤ SINR <15.889 | 64 QAM | 666 | 3.9023 |
| 13 | 15.889≤ SINR <17.814 | 64 QAM | 772 | 4.5234 |
| 14 | 17.814≤ SINR <19.829 | 64 QAM | 873 | 5.1152 |
| 15 | SINR ≥ 19.829 | 64 QAM | 948 | 5.5547 |

In LTE, the packet scheduling runs in both the Time Domain and frequency; they are known as Time Domain Packet Scheduling (TDPS) and Frequency Domain Packet

Scheduling (FDPS). TDPS determines the users to be served; whereas FDPS allocates the PRBs to the selected users [15].

### 2.2.2 Machine-to-Machine Communication

M2M deployment is considered one of the pillars of the constantly-growing IoT paradigm. Billions of MTCDs are envisioned to be connected to the Internet to offer all kinds of services. Many studies have attempted to adapt the current architecture and communication schemes to the needs of M2M communications. LTE and its subsequent technology LTE-A are the candidate technologies for M2M deployment. The networks are IP-based; and provide large capacity for users. Moreover, they allow flexible resource management and scalability [16].

The two important elements of M2M communication are the MTCDs and MTC Gateway (MTCG). The role of the MTCD is to collect data; for example, the sensors used to monitor environmental conditions such as temperature, water level, and humidity. The role of the MTCG is to manage the communication of the MTCD with the rest of the network in terms of power consumption and other communication parameters.

There are three supported transmission modes [17]: direct transmission, multi-hop transmission and peer-to-peer transmission. In direct transmission, the MTCD communicates directly with the eNB. This may cause network congestion in case a large number of transmission requests were sent to the eNB simultaneously. In multi-hop transmission, communication goes through the MTCG first. It handles the MTCD communications better in terms of power consumption but introduces additional complexity on the LTE network. The solution is to allow a single MTCG to manage multiple MTCDs in the same geographic area. The peer-to-peer transmission allows the MTCDs to communicate together or communicate with the MTC server. This architecture can help reduces the power consumption significantly as it facilities the spreading of messages.

The maim characteristic of MTC include [16] [17]:

- MTC traffic is mostly in the uplink direction (between the MTCD and the eNB).
- QoS customization: there are many applications for M2M deployments; and each application has different needs.
- Most M2M traffic is event-driven and bursty.

- Most of the time, M2M traffic has low rate and small packets. However, in cases of an incident, the payload increases dramatically within a short period of time.
- Delay-sensitivity: most MTCDs are deployed in mission critical applications with strict delay budgets.
- Power consumption limitations: the transmission power needs to be optimized for low-power MTCDs.
- Traffic prioritization: critical traffic needs to be prioritized over delay-tolerant traffic.
- Security: MTC connection especially in peer-peer communications need to be secured and encrypted.

### 2.2.3 Network Virtualization and Slicing

With the on-going growth of the IoT, it is expected that billions of MTCDs will be connected. The MNOs are resorting to sharing their resources in what is known as Network Virtualization (NV). NV provides solutions for the sharing MNOs to pool their resources and assign them fairly without affecting their individual performances and the QoS they offer to their individual users.

One of the approaches for network virtualization is slicing; the MNO that owns the spectrum license, "slices" its spectrum pool into portions and allocate them to the sharing virtual operators fairly according to an agreed upon Service Level Agreement (SLA).

Any slicing scheme needs to satisfy the three basic objectives for the sharing: isolation, customization, and resource utilization. The isolation ensures that the traffic variations in one MNO should not affect other MNOs. Customization allows the flexible handling of the MNOs' different needs. The utilization is maximized by reallocating unused resources in one MNO to the other MNOs without affecting the service level agreements (SLAs) offered to their customers [18].

3GPP has also defined the fundamental requirements of RAN sharing [2]. The requirements are shown in Table 2.3. They can be summarized in four categories: resources allocations, capacity negotiations, Operations, Administration and Management (OAM) and handovers [18].

**Table 2.3 RAN Sharing Requirements**

| Resources Allocations | Capacity Negotiation | OAM | Handover |
|---|---|---|---|
| • Each virtual operator is assigned a certion portion of resources.<br>• The portion may be static or dynamic<br>• Each virtual has the right to different Service Level Agreement (SLA) depending on its need<br>• Admission control needs to be applied to make sure that the virtual operators' SLAs are met. | • The virtaul operator may request additional resources based on their need; for example they may request additional capacity during an event or holiday season | • OAM is controlled by the hosting RAN.<br>• the hosting RAN sends periodic OAM updates to the virtual operators. | • The hosting RAN is allowed to handover devices to neighboring less congested cells to avoid exceeding the agreed upon virtual operators portion.<br>• The traffic is transferred from congested base stations to low-volume base stations, this is know as Load balancing (LB). |

## 2.3 Related Work

Motivated by the massive increase in data traffic, many studies have suggested different techniques for MNOs to use their spectrum efficiently. The possible spectrum efficiency approaches are network slicing or radio resource trading.

### 2.3.1 Network Virtualization and Slicing

In [5] the study discusses the Network Virtualization Substrate (NVS) scheme that uses a two-level hierarchical scheduler to allocate the wireless resources to multiple MVNOs. The first level handles the selection of the slice to be served. The second level deals with allocating the resources to the data bearers within the selected slice.

The Partial Resource Reservation (PRR) scheme is based on the NVS scheme. It introduces an additional shared slice. All the MVNOs can be allocated resources from the shared portion [6].

The Resource nEgotiation for NEtwork Virtualization (RENEV) scheme presented in [19], [20] allows the base stations to make the resource allocation decisions together by sharing a common baseband module among them. In NVS and PRR the resources are allocated among multiple MNOs within the same base station whereas RENEV allows the resource allocation among multiple MNOs within different base stations.

The scheme presented in [21] shows another approach to scheduling, different from the hierarchical scheduling used in NVS, in which each eNB has a hypervisor that allocates the resources. The scheme adds safety margins when estimating Quality of service (QoS) of the traffic and sets a maximum bit rate on non QoS traffic to guarantee that the SLA requirements are met.

The study in [22] proposes the CellSlice scheme in which the resource allocation decision is made by the gateway allowing less changes in the current base stations' schedulers compared to NVS; the main drawback in CellSlice is the gateway response delay.

The study in [23] presents a slicing technique for M2M communication in which a portion of the resources is reserved for M2M communication. This portion is computed at the end of each slicing time interval (STI). To adapt to the dynamic nature of M2M traffic, if the traffic conditions change at a high rate, the STI is small whereas, at times where the traffic conditions are stable, a larger STI may be used to reduce the computational load on eNBs.

As 5G is being launched in many countries in 2019, one of its cutting edges that differentiates 5G from the previous generations is end-to-end slicing, i.e. allowing the existence of end-to-end virtual networks (also known as logical networks) in both the RAN and CN. Recent papers address the issues caused by the end-to-end slicing in 5G. The study in [24] propose a solution for isolating different Virtual Network Functions (VNFs) in the same slice. Their motivation is that if all functions were on the same server, in case of failure, the whole slice goes down. The study in [25] suggests applying the concept of Earliest Deadline First (EDF) for slice scheduling. The scheduler prioritizes the slices whose deadlines are imminent for radio resources allocation. The study in [26] suggests using a "Network Slice Broker" which acts as a mediator between the resource providers and the virtual operators. The broker receives anonymous offers from the virtual operators and uses the Blockchain technology [27] to allocate the best resources in terms of pricing and quality to the virtual operators.

### 2.3.2 Radio Resource Trading

The second approach for spectrum efficiency is the radio resource trading. In the trading schemes, the large MNO acts as seller and the MVNOs act as buyers. The goods they are bidding on are the radio resources [28].The authors in [20] categorize the spectrum trading models into bidding, pricing and bargaining games. The pricing is usually set using an optimization or game theory problem formulation.

The work in [29] summarizes the process of bidding on radio resources by defining the true valuation of the resources, the seller and the buyer utilities. The seller and each one of the bidders estimate the value of the good of interest at a certain price known as "True Valuation". Similar to any auction, the bidder that offers the highest price wins the auction. After the auction, the utility of the seller is the difference between the payment he receives for the good he sells and his true valuation of the good. The utility of the buyer is the difference between his true valuation of the good and the price he pays for it.

This auction has been used for decades by The Federal Communications Commission (FCC) and similar authorities all over the world to distribute spectrum fairly among mobile operators.

This model for bidding is referred to as "single-sided" auction. Such auctions can last for months or years to be sealed. That's why many studies have suggested using "double-sided" auction. Unlike conventional auctions, double-sided auctions allow both the sellers and the buyers to announce their prices to a mediator. The mediator chooses a clearing price. All the sellers who asked for a price lower than the clearing price win and the all bidders that offered to pay more than the clearing price win. Examples of double-sided auctions include TRuthful doUble Spectrum aucTions (TRUST) [29], VERITAS [30] auctions and Truthful Online Double Auction (TODA) [31] .

The authors in [32] also propose an auction-based approach in which two spectrum pools are used: a price-based pool for delay-sensitive M2M traffic and a Waiting-time based pool for delay-tolerant M2M and H2H traffic.

As for the pricing-based trading, the study in [33] proposes a dynamic pricing scheme based on the Stackelberg game [34]. It allows the licensed MNO to maximize its revenue and the sharing MVNO to minimize the price it pays for the radio resources.

The study in [35] introduces an “application-aware pricing-scheme” that allows the coexistence of conventional H2H and M2M traffic on the same infrastructure. This pricing model is designed to adapt to the needs of M2M traffic when deployed in femtocells.

A bargaining-based dynamic spectrum management (BDSM) was introduced in [36]. It allows RANs to rent radio resources from each other based on their needs. To facilitate the transactions, the authors suggest adding an agent to each RAN to make the rent and pricing decisions.

The authors in [37] introduce a dynamic bargaining game for radio resource sharing among the primary users (the licensed MNOs) and the secondary users (MVNOs). The study also suggest that the secondary users relay the primary users’ traffic. This model allows both the primary and secondary users to offer higher data rates to their end-users.

## 2.4 Thesis Scope

Most of the network slicing schemes mentioned earlier were designed for H2H communication. They therefore focus on maximizing the throughput of the sharing operators. Due to the delay-sensitive nature of M2M communications, other parameters such as deadline missing need to be considered. Moreover, operating conditions change with time. This renders the slicing inefficient as such conditions change.

We therefore devise a slicing scheme that is based on optimization of resource division on the basis of the needs of M2M communications. Moreover, we handle the changing conditions by introducing a trading scheme that relies on the quality of the radio resources traded. The trading decisions are based on the ability of the traded resources to minimize the probability of the deadline missing for the clients of the requesting slice owner.

## 2.5 Chapter Summary

In this chapter we presented the necessary background information about LTE, focusing on the resource grid structure and the scheduling process. Then, we discussed the main elements of MTC, the possible communication modes and the characteristics of M2M traffic. We discussed the network virtualization objectives and requirements. These points will help explain the challenges that need to be addressed by the network virtualization schemes to serve M2M traffic and adapt to different MTCDs types. Then, we reviewed previous network that addresses virtualization schemes such as NVS, PRR, RENEV and

CellSlice. In addition, we discussed some of the published work about radio resources trading. Finally, we discussed the scope of the thesis highlighting the significance of the adapted NV schemes and the novel trading schemes compared to the previous work.

# Chapter 3: Slice-Based Resource Trading System Design

## 3.1 Introduction

As explained earlier in the introduction, there are many scenarios for NV deployment; in the thesis we focus on slicing in which a large MNO owning frequency license and network infrastructure allows multiple MVNOs to use its spectrum pool to serve their end-users. The MNO slices the spectrum into “slices” or portions; each MVNO is assigned one or multiple slices.

In the first part of the chapter, we discuss two of the most known network virtualization schemes (NV): Network Virtualization Substrate (NVS) and Partial Resource Reservation (PRR). NVS uses a resource-based and Data rate-based slice allocation.

We focus on the high-level allocation of PRBs to MVNOs. It is the scheduling scheme that handles the PRBs allocations to the users of each MVNO, and it is outside the scope of the thesis. For Resource-based NVS, Data rate-based PRR, we present the different approaches to adapt the schemes MTC traffic. Two major issues need to be addressed: minimizing the deadline missing ratio and customizing the SLA offered to each MVNO. To evaluate the performance of the adapted NV schemes we compare them to the static reservation of resources (SR).

In the second part of the chapter, we propose a novel trading scheme that allows the sharing operators to trade their PRBs based on their Channel Quality Indicators (CQIs). The trading decisions are optimized using Genetic Algorithm (GA). We explain the trading scheme flow in details; then we present some background information about GA. In the last section of the chapter, we explain the dynamic Slicing Time Intervals (STIs) calculations to allow the scheme to initiate the trading at changing time intervals depending on the rate of change of the traffic.

## 3.2 Adapting Network Virtualization Substrate (NVS) to M2M Communication Requirements

### 3.2.1 Network Virtualization Substrate (NVS)

The simplest form of slicing is SR in which, each MVNO is reserved a static portion, $Pr_i^{ref}$ such that $\sum_{i=1}^{N} Pr_i^{ref} = 1$.

The main advantage of SR slicing is that the fluctuations of traffic in one MVNO does not affect the other ones as their portions of resources are allocated to them permanently. The obvious drawback of SR is that unused radio resources from one slice cannot be reallocated to the other slices. SR is mentioned in this work as a baseline to which we compare the adapted schemes.

NVS, on the other hand, uses a 2-level hierarchical scheduling. The first level of allocation is selecting which slice to be served while the second level is allocating the radio resources to the data bearers within the selected slice [2] [5].

The slice selection can be either Data rate-based or resource-based. In the Data rate-based implementation ( also known as Bandwidth-based NVS), each slice, $i$, reserves a certain overall data rate $r_i^{ref}$ for all its users in Megabits per second($Mbps$). In the resource-based implementation, each slice, $i$, reserves a certain portion of the PRBs, $t_i^{ref}$. To select the slice to be served, the slice weights are calculated periodically.

If a slice, $i$, uses Data rate-based reservation, the weight function at time instant $t$, $W_{i,t}$, is given by (3.1 a) . If $i$ uses resource-based reservation, $W_{i,t}$ is given by (3.1 b) :

$$W_{i,t} = \begin{cases} \frac{r_i^{ref}}{r_{i,t}^{exp}} & (a) \\ \frac{t_i^{ref}}{t_{i,t}^{exp}} & (b) \end{cases} , \qquad (3.1)$$

where $r_{i,t}^{exp}$ and $t_{i,t}^{exp}$ are the exponential moving averages of the total data rate allocated to slice $i$ and the portion of PRBs allocated to slice $i$ at time $t$, respectively. They are computed as follows:

$$r_{i,t}^{exp} = (1-\alpha) r_{i,t-1}^{exp} + \alpha r_{i,t}^{exp}$$
$$t_{i,t}^{exp} = (1-\alpha) t_{i,t-1}^{exp} + \alpha t_{i,t}^{exp} , \qquad (3.2)$$

where $\alpha$ is a positive weighting factor between 0 and 1; larger $\alpha$ signifies larger weight for current values.

Since $r_{i,t}^{ref}$ and $t_{i,t}^{ref}$ are the agreed SLAs between the MNO and the MVNO, higher slice weight indicates that the slice is farther from its SLA. The slice with the highest weight will be selected and its MTCDs will be allocated PRBs by the eNodeB (eNB) scheduler. Any remaining PRBs will be allocated to the MTCDs of the slice having the second highest weight. This goes on until all PRBs are allocated or all slices are served.

The obvious advantage of NVS over SR is its ability to allocate unused resources by one slice to the other slices.

To adapt resource-based NVS to the needs of the M2M traffic, the portion reserved to each slice $t_i^{ref}$ depends on its delay budget. Slices that have strict delay budgets are reserved a larger portion of PRBs to minimize the probability of deadline missing. We also put in consideration the packet size of the owning MVNOs. For example, surveillance cameras have strict delay budgets and large payloads, they may be assigned larger portion than less demanding slices (such as environmental sensors).

To adapt Data rate-based NVS to M2M traffic, $r_i^{ref}$ must also be evaluated as a function of the probability of deadline missing.

Let $\lambda_m$ be the arrival rate of an MTCD, $m$, $D_m^{bound}$ its delay budget. In Poisson Processes, the events occur continuously, independently and at a constant mean rate (i.e. The interarrival time is exponentially distributed); so, the event-triggered nature of M2M traffic can be modeled as a Poisson Process [38].

The serving rate of MTCD $m$, $\mu_m$, is computed as the sum of the achievable throughput, $Thr_{m,k}$, of the $K$ PRB(s) assigned to it:

$$\mu_m = \sum_{k=1}^{K} Thr_{m,k} \quad (3.3)$$

$Thr_{m,k}$ is a function of the spectral efficiency, which depends on the Channel Quality Indicators (CQIs) reported by the MTCDs. The CQI mapping to the spectral efficiency is listed in Table 2.3.

$$Thr_{m,k} = \frac{spetrcal\ efficiency\ (number\ of\ bits\ per\ RE) \times RE}{TTI\ (ms)}, \quad (3.4)$$

where $RE$ is number of resource elements per TTI ($RE = 2 \times 12$ subcarriers $\times$ $7$ symbols $= 168$), and *TTI=1 ms*.

Assuming that 20 out of the 168 *REs* are used for control and signaling, the achievable throughput of MTCD, *m*, for assigned PRB *k* [39]

$$Thr_{m,k} = Spetrcal\ efficiency\ of\ RB_{m,k} \times 148 \times 10^{-3}\ (Mbps) \quad (3.5)$$

Each MTCD queue can be modeled as M/D/1, assuming that $\tau_m = 1/\mu_m$ is the deterministic servicing time, the probability of deadline missing for MTCD, $m$, is computed as follows [40]:

$$\Pr[D_m > D_m^{bound}] = 1 - \left(1 - \frac{\lambda_m}{\mu_m}\right) . \sum_{v=0}^{z} \frac{[-\lambda_m(D_m^{bound} - \tau_m - v.\tau_m)^v]}{v!} . e^{(\lambda_m(D_m^{bound} - \tau_m - v.\tau_m)}, \quad (3.6)$$

where $z$ is an integer such that: $z\tau_m \leq ((D_m^{bound} - \tau_m) \leq (z+1)\tau_m$

The scheme therefore aims at keeping the probability of deadline missing below a certain threshold, $P_m^{max}$.

The minimum serving rate $\mu_m^{min}$ is calculated by setting $\Pr[D_m > D_m^{bound}] = P_m^{max}$.

The reserved data rate for slice $i$ having $M_i$ MTCDs can therefore be written as:

$$r_i^{ref} = \sum_{m=1}^{M_i} \mu_m^{min} \quad (3.7)$$

### 3.2.2 Partial Resource Reservation (PRR)

PRR [6] is based on Resource-based NVS, it has introduced an additional slice known as the shared slice; it is accessible to all MVNOs.

Each slice $s$ reserves a certain portion $t_i^{ref}$ of PRBs However, the shared slice portion is computed as follows:

$$t_{shared}^{ref} = 1 - \sum_{i=1}^{N'} t_i^{ref}, \quad (3.8)$$

where $N'$ is the number of the non-shared slices.

Similar to NVS, the slice weights are calculated by (3.1) and the slice having the highest weight is selected. MTCDs within the slice are allocated the PRBs by the eNB scheduler. If there are remaining PRBs, the slice having the second highest weight will be served and so on until all slices are served.

The only difference is that, after all the non-shard slices have been allocated the PBRs from their respective portions, any unserved MTCD belonging to any slice can be allocated PRBs from the shared part.

The simplest approach to adapt PRR to M2M traffic is to prioritize delay-sensitive traffic to minimize the probability of deadline missing. The shared PRBs are allocated in such a way that the MTCDs with the tightest delay budgets are prioritized.

## 3.3 Slice-Based Resource Trading Algorithm

In the proposed scheme, the mediator of the auction is the hosting eNB (owned by the licensed MNO); and the bidders are the sharing MVNOs. Each MVNO is assigned a slice; and similar to SR, each slice $i$ reserves a portion, $Pr_i^{ref}$, of the PRBs pool. The scheme allows the sharing MVNOs to trade PRBs from their respective portions periodically.

The trading time intervals are known as the Slicing Time Intervals (STIs). At the end of each STI, the Deadline Missing Ratio (DMR) is computed. It is the ratio of the packets that missed their transmission deadlines to the total number of packets for all MTCDs in every slice $i$.

If this ratio exceeds a certain threshold, the MVNO owning the slice sends a trading request to the mediator. If the mediator receives 2 or more trading requests within the same STI, a trade is initiated. The slices will trade a certain number of PRBs based on their CQI.

In our model, we assume that the CQIs differ from one PRB to the other due to the multipath fading. We also assume that the CQI experienced by one MTCD differs from the others as they are deployed over a large geographic area. This is the main objective of the trade; we assign the PRB to the MTCDs depending on their reported CQIs. As the CQI differs from one MTCD to the other within the same slice; the spectral efficiency and the achievable throughput will also differ.

Thus, we depend on the median of the achievable throughput experienced by all the MTCDs in the same slice to estimate the selling/buying slice valuation for a certain PRB. We aim to maximize the utility of the trade by calculating the difference between the achievable throughputs by the “buying” slice and the achievable throughput by the “selling” slice. We also relate the achievable throughput to the minimum serving rate needed by the slice to keep the probability of deadline missing below a certain threshold.

### 3.3.1 The Proposed Trading Problem Formulation

Assuming that there are *n* slices that are willing to participate in the trade, the steps taken, according to the proposed technique, are as follows:

1. Each slice, $i$, orders its PRBs in a descending order of channel quality. It participates in the trade with the last $\varkappa_i$ PRBs (i.e. the least performing PRBs).
2. Let $M$ be the set of all users in the system such that $M = \bigcup_{i=1}^{n} M_i$, where $M_i$ is the set of users in each slice.
3. Slice $i$ announces the set $K_i$ of all the PRBs it is willing to trade such that:

$$K_i = \{1, \dots , k, \dots , \varkappa_i\}$$

   The total number of PRBs to be trade is $\kappa = \sum_{i=1}^{n} \varkappa_i$ and $TR$ is the set of all traded PRBs $TR = \bigcup_{i=1}^{n} K_i$.

4. Slice $i$ also announces its valuation $R_{i,k}$ for every PRB $k$ offered for trade. The achievable throughput if PRB $k$ is assigned to MTCD $m$, $THR_{m,k}$ depends on the CQI; it is computed using Equation 3.5.

   As the CQI differs from one MTCD to another within the same slice, we estimate the achievable throughput among all MTCDs in slice *i* as:

$$R_{i,k} = \underset{m \in M_i}{\text{Median}}(THR_{m,k}) \ \forall \ k \epsilon K_i \tag{3.9}$$

5. Each slice $i$ also declares its valuation for every PRB offered by every other slice $j$ as follows

$$V_{i,k'}^{j} = \underset{m \in M_i}{\text{Median}}(THR_{m,k'}) \ \forall \ k' \in K_j \tag{3.10}$$

6. For all possible combinations of $i, j \ and \ k$, the utility of the trade is computed. It is the difference between the achievable throughput when the PRB is allocated to its new slice and the achievable throughput when the PRB was allocated to its old slice $i$.

   If slice $i$ receives PRB $k$ from slice $j$, the utility of the trade $U_{i,k}^{j}$ is calculated as follows

$$U_{i,k}^{j} = V_{i,k}^{j} - R_{i,k} \tag{3.11}$$

7. $I$ is an index matrix such that:

$$I_{i,k} = \begin{cases} 1 & \text{if PRB k belongs to i} \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

8. Each slice should receive the same number of PRBs it gave away, so that $\varkappa_i$ would remain the same after the trade.
9. Let $\lambda_m$ be the arrival rate for an MTCD $m$, $D_m^{bound}$ its delay budget and $\mu_m$ the serving rate. The Probability of Deadline Missing for MTCD $m$ is calculated using Equation 3.6. The scheme keeps the probability of the deadline missing below a certain threshold, $P_{max}$. The minimum serving rate $\mu_m^{min}$ is then calculated by setting the probability of deadline missing to $P_m^{max}$ as explained earlier. We aim to keep the serving rate $\mu_m \geq \mu_m^{min}$ for each MTCD.
10. Each slice $i$ has multiple MTCDs with different arrival rates and serving rates $\mu_m$. The minimum serving rate of slice $i$ is calculated as follows:

$$\mu_i^{min} = \underset{m \in M_i}{\text{Median}}(\mu_m^{min}) \tag{3.13}$$

The median is used rather than the average to estimate the values of the slice minimum serving rate to ensure that extreme values do not affect the slice-level estimated value.
11. Since we aim to trade resources between slices not MTCDs, $\mu_m$ is replaced by $R_{i,k}$ and $\mu_m^{min}$ by $\mu_i^{min}$.

$$R_{i,k} \geq \mu_i^{min} \tag{3.14}$$

12. As explained earlier in (3.11), the trade utility is the difference between the achievable throughput by PRB $k$ before and after the trade. So, the objective function can be expressed as

$$\max \sum_j \sum_k I_{j,k} U_{i,k}^j , \tag{3.15}$$

$$\text{where} \quad j = 1,2, \dots n ;$$

$$i = 1, \dots , j-1, j+1, \dots n;$$

$$and \;\; k \in TR$$

$$\text{Subject to } \textstyle\sum_{k \in K} I_{j,k} = \kappa_j \tag{3.15a}$$

$$\textstyle\sum_{j=1}^{n} I_{j,k} = 1 \tag{3.15b}$$

$$R_{j,k} \geq \mu_j^{min} \tag{3.15c}$$

Condition (3.15a) ensures that each slice receives the same number of PRBs it gave away. Condition (3.15b) ensures that no PRB is allocated to 2 slices at the same time. Condition

(3.15c) ensures that the minimum serving rate is achieved. We use the GA technique to solve the objective function in (3.15).

### 3.3.2 Trading Algorithm Description

The Genetic Algorithm (GA) is a well-known heuristic optimization algorithm, it is inspired by Darwin's Evolution theory. GA was chosen over other known heuristic optimization methods such as Ant Colony, Particle Swarm and Trust-region as it suits our use case better. Ant Colony is best used to determine the best route from one start point to a destination (i.e. it needs a closed loop problem). Particle Swarm optimization is mostly used for continuous values; whereas, in our case we use GA to estimate the index matrix *I*. Trust-region optimization needs a near-optimal start point; which is not applicable in the trading case.

The values resulting from the iterations are known as "population", and the alterations applied on the individuals to reach the optimal solution are known as "genetic operations". An initial value is selected randomly, it is referred to as the initial population. A series of genetic operations are applied on the initial population to produce the following generations. They are namely Selection, Crossover and Mutation.

The first step is selection, the "fitness" of the population is evaluated; and then the individuals that show better fitness are selected to produce the new generation. This selection is based on the objective function (also known as fitness function in GA).

Some of the selected parents are then recombined to produce children; this is the crossover operation. The selected parents may also be mutated to form the new generation. A simple example of mutation is random bit flipping [41].This goes on until the termination condition is met or the maximum number of generations is reached [41]. Figure 3.1 illustrates the flow of the GA optimization.

In the proposed scheme, the population is the $n \times \kappa$ index matrix $I$. GA chooses a random initial value for $I$; then calculated the utility of the trade in case a PRB previously owned by a certain slice *i* is given to a certain slice *j*. The GA aims to maximize the utility of the trade as explained in equation 3.15. The selected values for *I* must satisfy the 3 constraints a, b and c. ALGORITM 1 explains briefly the flow of the proposed trading scheme.

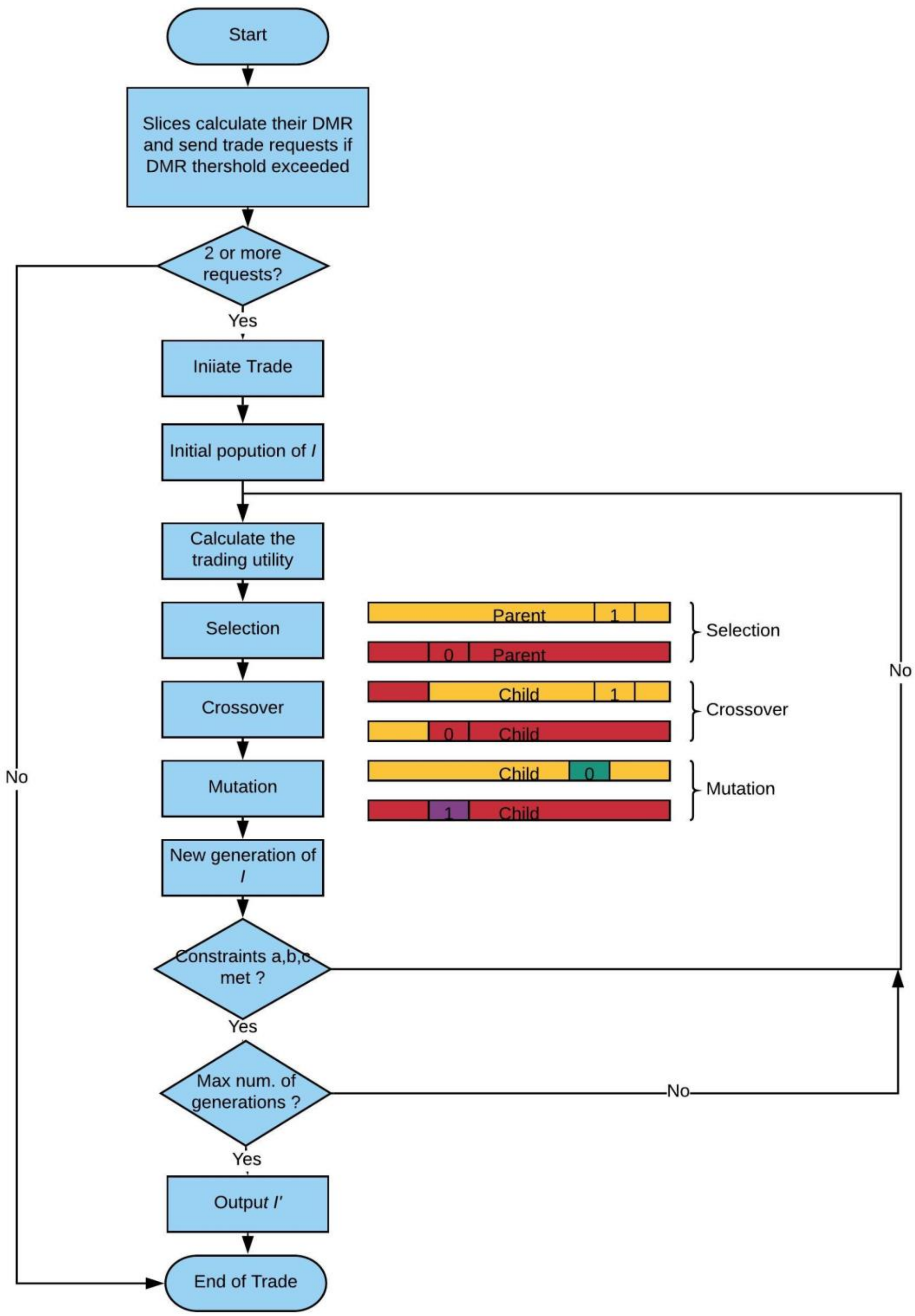


**Figure 3.1 Flowchart of the Genetic Algorithm**

**ALGORITHM 1: The Resource Trading Scheme**

| Step | Description |
|---|---|
| 1. | **procedure** TRADING_ALOGORITHM (n) |
| 2. | **for** i ←1:n **do** |
| 3. | $K_i$ ← the set of the indices of the PRBs to trade by $i$ |
| 4. | $\mu_i^{min}$ ← minimum serving rate of slice $i$ |
| 5. | $R_{i,k}$ ← the valuation of the PRBs to be traded |
| 6. | **end for** |
| 7. | $\kappa$ ← the total number of traded PRBs |
| 8. | $TR$ ← the set of all traded PRBs |
| 9. | //Setting the index matrix |
| 10. | **for** $i$ ← $1{:}n$ **do** |
| 11. | **for** $k$ ← $1{:}\kappa$ **do** |
| 12. | **If** $k$ belongs to I **then** |
| 13. | $I(i,k) = 1$ |
| 14. | **else** |
| 15. | $I(i,k) = 0$ |
| 16. | **end if** |
| 17. | **end for** |
| 18. | **end for** |
| 19. | // Estimating the valuation of PRBs shared by other slices |
| 20. | **for** $i$ ← $1{:}n$ **do** |
| 21. | **for** $j$ ← $1{:}n$ **do** |

```
                for k ← 1: κ do
                    if i ≠ j   then
                        V_{i,k}^j ← the valuation of the PRBs shared by other slices
                    end if
                end for
            end for
        end for
        // Starting GA
        I′=GA (I, V,R)
        // Assigning the PRBs to slices after trade
        for i ← 1: n do
            K_i′=[ ]
            for k ← 1: κ do
                If I′(i, k) = 1 then
                    K_i′=[K_i′  TR(k)]
                end if
            end for
        end for
end procedure
```

### 3.3.3 Dynamic STI Calculation

Due to the dynamic nature of M2M traffic, the Deadline Missing Ratio (DMR) evaluations and trading requests time intervals need to adapt continuously. In times where the traffic is stable, we may use large STIs to reduce the computational load of the eNB. In times where the traffic is changing rapidly, slice need to evaluate their DMR results and trade resources more often. The rate of change of traffic is estimated using the autocorrelation.

Let $M$ be the total number of MTCDs in the system; and let $X_m(t)$ be the random traffic generation process by each MTCD, $m$.

So, the total traffic generated in the system can be represented by the random process $X_{tot}(t)$ such that:

$$X_{tot}(t) = \sum_{m=1}^{M} X_m(t)$$

The autocorrelation of $X_{tot}$ , $A(\tau)$, is computed as follows:

$$A(\tau) = \frac{1}{ht_0} \sum_{j=i-h+1}^{i} X_{tot}(jt_0)\, X_{tot}(jt_0 - \tau)$$

Where $t_0$ is the TTI, $i$ is the TTI index and $h$ is the previous sample size used for the calculations.

We estimate $\tau$ that keeps $A(\tau)$ less than a certain threshold $A_{th}$

Below is the process for dynamic STI calculation:

A. We set a default sample size of the total generated data $X_{tot}(t)$,$h_{dft}$
B. We set a standard deviation threshold $Std_{th}$, upper and lower limit for autocorrelation $A_{th,upper}$ and $A_{th,lower}$
C. We calculate the STI as shown in ALGORITM 2.

**ALGORITHM 2: Dynamic STI Calculation**

| Step | Description |
|---|---|
| 1 | **procedure** CALCULATE_STI (samples) |
| 2 | Std ← Standard deviation of the $h_{dft}$samples |
| 3 | **if** Std < $\mathrm{Std_{th}}$ |
| 4 | $A_{th} \leftarrow A_{th,lower}$ |
| 5 | **else** |
| 6 | $A_{th} \leftarrow A_{th,upper}$ |
| 7 | **end if** |

```
        A(0) ← autocorrelation of samples at τ = 0
        for τ ← 1: h_dft do
            A(τ) ← autocorrelation of samples at τ
                If A(τ)/A(0) <  A_th then
                    break
                end if
        end for
        STI ← τ
        return STI
end procedure
```

## 3.4 Chapter Summary

In this chapter, we explained Resource-based NVS, Data rate-based NVS and PRR. We showed that these schemes can be adapted to MTC needs. Two requirements have been satisfied: delay-sensitivity and customization for different applications. We then presented a novel resource trading scheme. It allows the resource-sharing MVNOs to trade a certain number of PRBs from their reserved slices. The valuation of the PRBs is based on their ability to minimize the probability of deadline missing. The trading scheme is therefore adapted to M2M traffic.

# Chapter 4: Experimental Validation of the Proposed Techniques

## 4.1 Introduction

In this chapter, we present the experimental setup and simulation results of the performance evaluation of the proposed techniques. First, we present the comparison between the adapted Resource-based NVS, Data rate-based NVS, PRR and SR. Then, we compare the novel trading scheme to the adapted NV schemes. We use both static and dynamic STIs to compare the different slicing schemes. As we focus on massive mission-critical MTCDs deployment, we evaluate the performance of the schemes from deadline missing perspective.

## 4.2 Experimental Evaluation of the Adapted Network Virtualization Schemes

### 4.2.1 Simultation Setup

We consider 4 MVNOs sharing a single eNB. Each MVNO is allocated a single slice and it serves a different type of MTCDs. We run a series of simulations on MATLAB with 2 bandwidth configurations (10 and 20 MHz) and 4 numbers of MTCDs (100, 200, 300 and 400). The simulation parameters are shown in Table 4.1. The slice weights are calculated at each TTI ( i.e. STI=1 TTI).

Slice 1 servers alarm MTCDs whose role is to report major incidents (such as fires or earthquakes). They are characterized by small data packets and the most rigid delay budgets. Slice 2 serves motion-activated surveillance cameras (i.e. security cameras that only capture videos when there is some movement around them). These are characterized by larger packet sizes and relatively small delay budgets. Slice 3 serves monitoring sensors (e.g. temperature or humidity sensors) and they are characterized by smaller packets compared to video packets and less strict delay budgets than alarms systems and surveillance cameras. Slice 4 serves multiple MTCDs types such that 30% are alarm devices, 30% are surveillance cameras and 40% are monitoring sensors [16] [23].

The parameters of each device in slice 4, such as the delay budget or packet size, depend on its type. The detailed configuration of each MTCD type is shown in Table 4.2.

**Table 4.1 Simulation 1 Setup**

| Parameter | Value |
|---|---|
| Number of Runs | 5 |
| Confidence Level | 95% |
| SNR Range | 0-20 dB (uniformly distributed) |
| Number of eNBs | 1 |
| Number of MTCDs | 100, 200, 300, 400 |
| Number of Subframes | 1,000 |
| Number of Slices | 4 |
| Channel Bandwidth (MHz) | 10 MHz ( 50 PRBs) and 20 MHz ( 100 PRBs) |
| Probability of deadline missing threshold ($\boldsymbol{P_m^{max}}$) | 0.1 |
| α | 0.5 |
| STI | 1 TTI |

**Table 4.2 Slices Configuration in Simulation 1**

| Parameter | Slice 1 | Slice 2 | Slice 3 | Slice 4 |
|---|---|---|---|---|
| Traffic Type | Alarm | Surveillance Cameras | Monitoring Sensors | Mixed |
| Arrival Distribution | Poisson | Poisson | Poisson | Poisson |
| Delay Budget (ms) | 5-10 | 20 | 20-50 | Varied |
| Arrival Rate (Packets/s) | 15 | 20 | 5 | Varied |
| Packet Size (Bytes) | 256 | 1024 | 512 | Varied |
| Percentage of MTCDs | 20% | 20% | 30% | 30% |

For SR and the resource-based NVS implementations, surveillance cameras were assigned 40% of the PRBs as they transmit the largest packets. The other 3 slices were assigned 20% of the resources each. Even though alarm systems have the smallest packet sizes, they were also assigned 20% of PRBs to ensure that their strict delay budget need is always satisfied. For Data rate-based NVS, each of the 4 slices was reserved a minimum data-rate $r_s^{ref}$ which is the sum of the minimum serving rates of all MTCDs in the slice, $\mu_m^{min}$ (as explained earlier in section 3.2).

In PRR, alarm, monitoring sensors and mixed traffic slices were allocated 20% of the PRBs each (similar to SR and Resource-based NVS). Surveillance cameras were allocated 30% of PRBs (instead of 40%) as the remaining 10% were assigned to the shared slice. The SLA allocated to each slice is shown in Table 4.3.

**Table 4.3 Slices Service Level Agreements in Simulation 1**

| NV Scheme | Slice 1 Alarm | Slice 2 Cameras | Slice 3 Sensors | Slice 4 Mixed | Slice 5 Shared (PRR only) |
|---|---|---|---|---|---|
| SR | $t_1^{ref} = 0.2$ | $t_2^{ref} = 0.4$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | NA |
| Resource-based NVS | $t_1^{ref} = 0.2$ | $t_2^{ref} = 0.4$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | NA |
| Data rate-based NVS | $r_1^{ref}$ $Mbps$ | $r_2^{ref}$ $Mbps$ | $r_3^{ref}$ $Mbps$ | $r_4^{ref}$ $Mbps$ | NA |
| PRR | $t_1^{ref} = 0.2$ | $\boldsymbol{t_2^{ref} = 0.3}$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | $\boldsymbol{t_5^{ref} = 0.1}$ |

### 4.2.2 Experimental Results

We ran a series of experiments on MATLAB; the results were computed as the average of five independent runs with 95% confidence interval analysis. As we focus on MTC deployment; the most import metric is deadline missing ratio (i.e. the ratio of the packets missing deadlines to the total number of packets).

The below figures show the deadline missing results for Resource-based NVS, Data rate-based NVS (also known as Bandwidth-based NVS), PRR and SR. As the number of PRBs available to MTCDs increase from 50 (Figure 4.1) to 100 (Figure 4.2), the deadline missing ratio decreases across all NV schemes. Similarly, as the number of MTCDs increases, the number of PRBs that can be allocated to each MTCD decreases so the deadline missing ratio increases across all NV schemes.

Both resource-base and Data rate-based NVS implementations have improved the deadline missing ratio compared to SR. NVS allows the unused PRBs from one slice to be used by the other slices in need. Whereas in SR, each slice can only be allocated PRBs from its reserved portion. Therefore, more packets missed their deadlines as they were not assigned the PRBs on time. For example in the extreme case of 400 MTCDs with 50 PRBs, the deadline missing ratio in SR reached 24%; whereas, in Resource-based NVS, it did not exceed 16%. PRR performed slightly worse than both NVS models (17%).

In the 10 MHz Bandwidth, PRR had a higher deadline missing ratio than NVS for all MTCDs numbers as a portion of PRBs were reserved to the shared slice. However, in the 20 MHz simulation, PRR had the lowest deadline missing ratio. PRR gave higher priority to MTCDs with smaller delay budgets to be assigned the shared PRBs. So, it improved the overall performance of the system; this improvement is more pronounced in higher bandwidth systems.

Since one of the major objectives of NV is customization to satisfy the different needs of the slices, the deadline missing ratio was evaluated for each slice. Figure 4.3 shows the deadline missing ratio for each slice when 400 MTCDs are served using 50 PRBs. Figure 4.4 also illustrates the deadline missing ratio for each slice when serving 400 MTCDs with 100 PRBs. The surveillance cameras have the largest packet size and relatively rigid delay budgets so they have the highest deadline missing ratio in all NV implementations.

In the Data rate-based and resource-based NVS implementations, the agreed upon data rate $r_s^{ref}$ and PRBs portion $t_s^{ref}$ were both related to the delay budget of each slice. Both approaches succeeded to improve the performance of NVS over SR (for slice 1, 3 and 4).
However, SR performed better in the surveillance cameras slice in the extreme case of 400 MTCDs served with only 50 PRBs. This is due to fact that SR permanently allocates 40% of PRBs to the cameras. Even if the traffic increases in the other slices, the slice reservation is not affected. This quality was more pronounced when the payload is high in low-bandwidth implementation. Both models of NVS and PRR performed better than SR when 100 PRBs were used to serve the surveillance cameras.
In PRR simulation, the surveillance cameras were assigned 30% of resources instead of 40%. Therefore, the deadline missing percentage for this slice was higher in PRR than NVS.
It is worth mentioning that in the case of alarm slice, PRR, Resource-based NVS and Data rate-based NVS performed much better than SR in both 10 and 20 MHz bandwidth cases. They all achieved 0% deadline missing ratio (in the 20 MHz-bandwidth case). In the 10 MHz case, Resource-based NVS achieved 2% whereas SR achieved 12%.

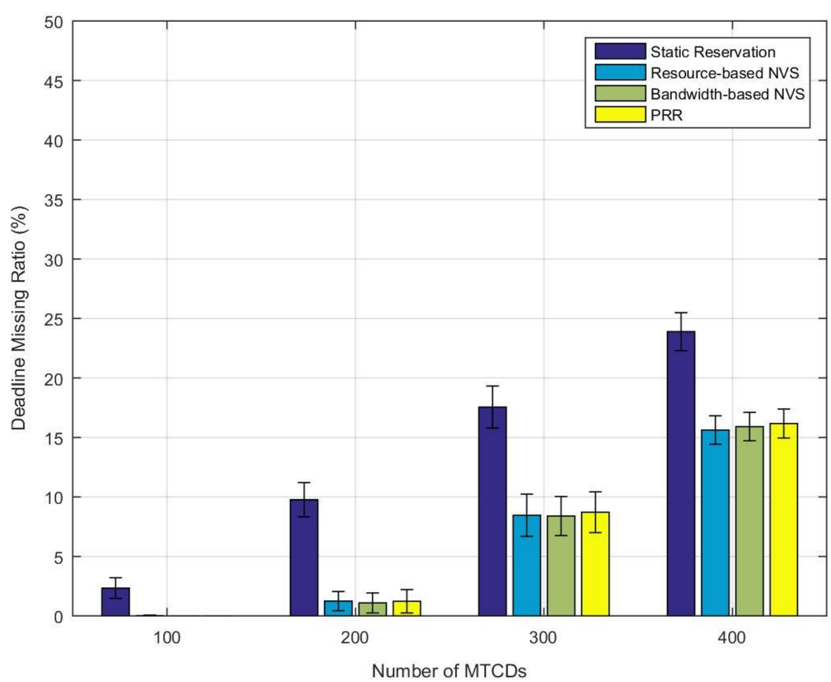


**Figure 4.1 Deadline missing ratio, bandwidth = 10MHz**

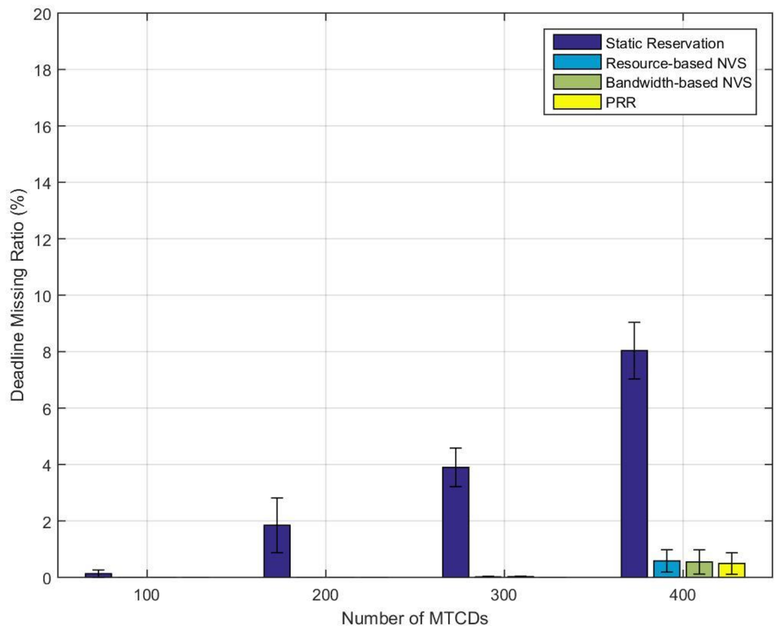


**Figure 4.2 Deadline missing ratio, bandwidth = 20MHz**

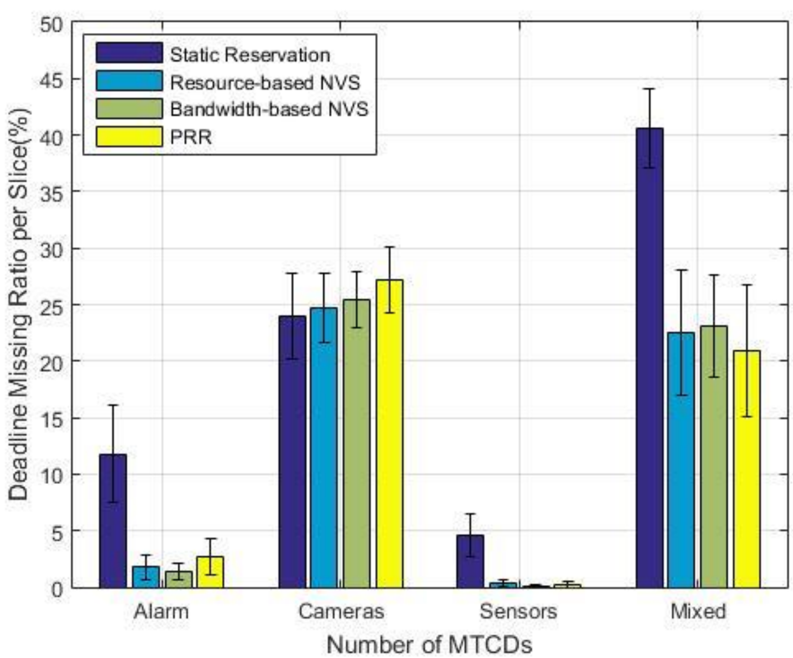


**Figure 4.3 Deadline missing ratio per slice, bandwidth =10 MHz**

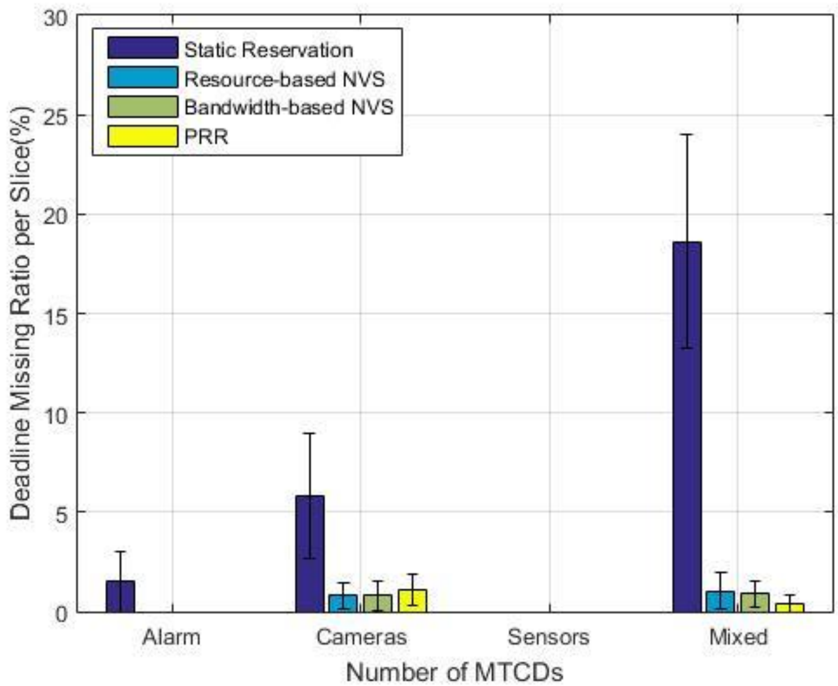


**Figure 4.4 Deadline missing ratio per slice, bandwidth = 20 MHz**

Another metric was also considered; which is the aggregate system throughput. Figure 4.5 and 4.6 illustrate the overall throughput (in Mbps) in the case of 10 and 20 MHz respectively. Naturally as the number of PRBs available in the system increase, the throughput increases. The throughput results are consistent with the deadline missing results. Both Models of NVS, PRR achieve higher throughput than SR. for example when 100 PRBs were used, SR achieved a throughput of 21 Mbps; whereas, PRR and both NVS models achieved approximately 24 Mbps.

The last parameter to examine is utilization; Figure 4.7 and 4.8 illustrate the PRBs utilization in the 10 MHz and 20 MHz bandwidth cases respectively. Naturally, the utilization is higher in the 10 MHz compared to the 20 MHz case; it reached almost 100% when serving 400 MTCDs. Since SR does not allow the unused PRBS from one slice to be reallocated to the others; the PRBs were not best utilized.

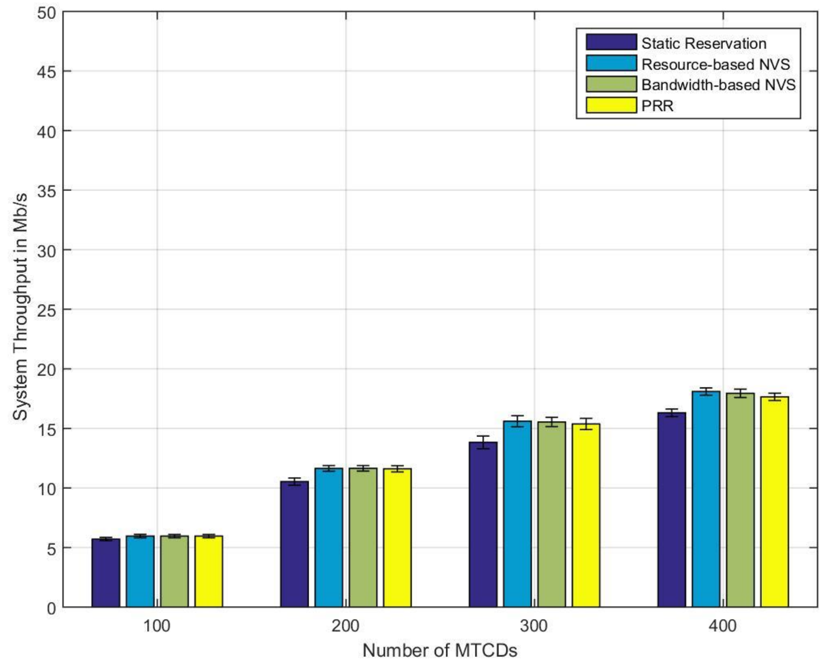


**Figure 4.5 Overall system throughput, bandwidth=10 MHz**

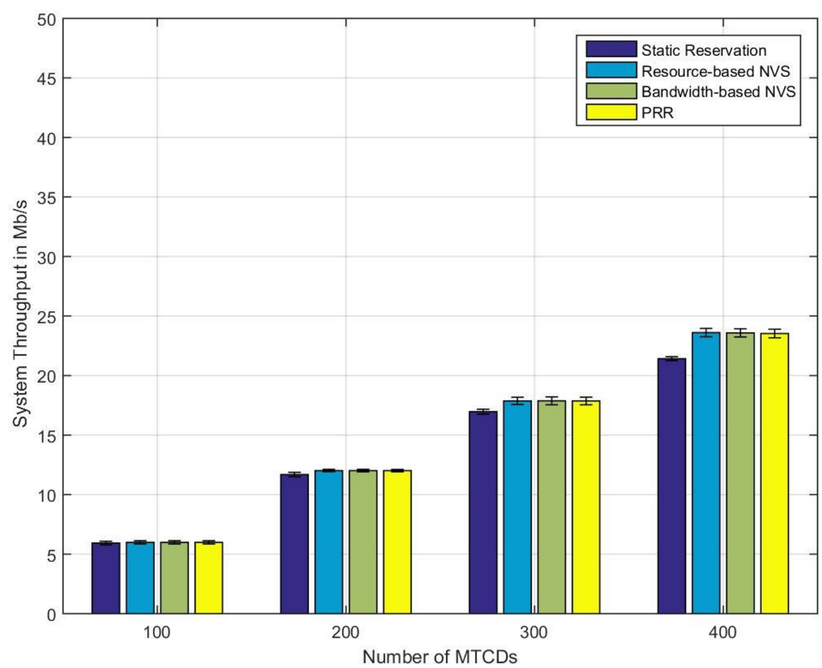


**Figure 4.6 Overall system throughput, bandwidth=20 MHz**

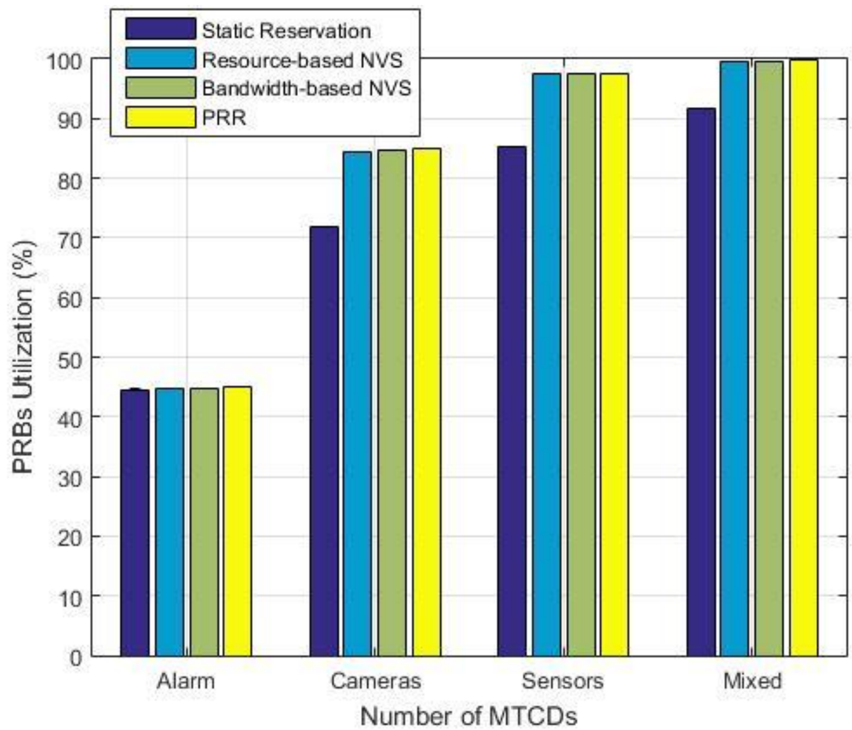


**Figure 4.7 PRBs utilization, bandwidth=10 MHz**

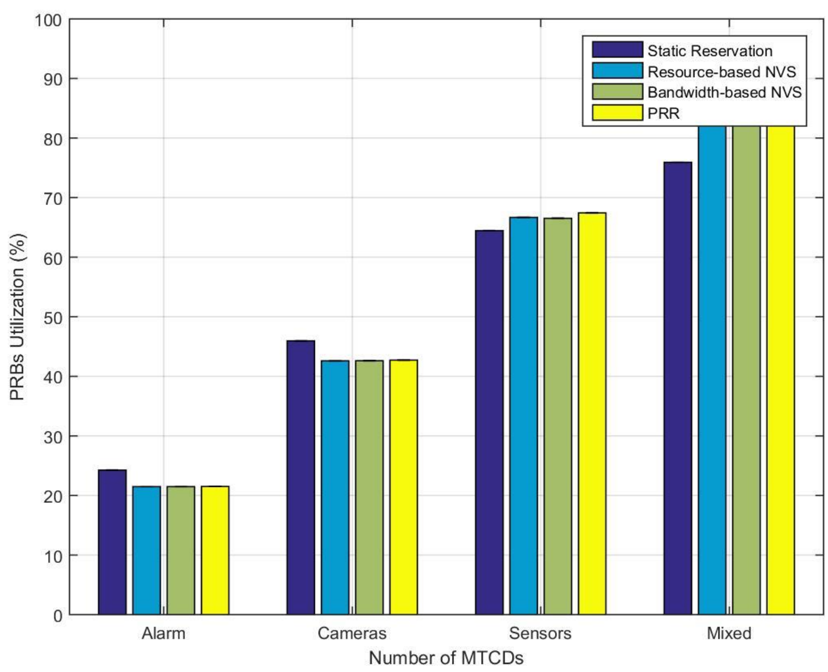


**Figure 4.8 PRBs utilization, bandwidth=20 MHz**

## 4.3 Experimental Evaluation of the Proposed Resource Trading Scheme

### 4.3.1 Simultation Setup

In this section, we compare the performance of the novel trading schemes to the adapted Resource-based NVS, Data rate-based NVS (a.k.a. Bandwidth-based NVS) and PRR. We run a series of simulations using MALAB with 4 slices. Each slice belongs to a different MVNO and serves a different type of MTCDs.

The simulation setup is identical to the first round of experiments. Except that the slice weights are calculated at larger STIs. In the static implementation, we used 10,20,30,40 and 50 TTIs; in the dynamic case, the STI was computed as function of the traffic auto-correlation.

We test the trading scheme in massive MTCDs deployment so larger numbers of MTCDs were used; ranging from 400 to 700 MTCDs. We assume that the system bandwidth is 10 MHz; the detailed simulation setup is shown in Table 4.4. The MTCDs configuration is identical to the previous experiments (Table 4.2).

The SLA reserved to each slice is listed in Table 4.5. In the trading and the resource-based NVS schemes, the alarm, surveillance cameras, sensors and mixed slices reserve 20%, 40%, 20% and 20% of the PRBS respectively. PRR uses the same distribution except that the surveillance cameras are assigned 30% instead of 40% and the remaining 10% are assigned to the shared slice. In the Data rate-based NVS, each slice is assigned a certain reference bandwidth, $r^{ref}$. Each slice is allowed to trade up to 50% of its reserved PRBs.

**Table 4.4 Trading Scheme Simulation Setup**

| Parameter | Value |
|---|---|
| Number of Runs | 5 |
| Confidence Level | 95% |
| SNR Range | 0-20 dB (uniformly distributed) |
| Number of eNBs | 1 |
| Number of MTCDs | 400,500,600 and 700 |

| | |
|---|---|
| Number of Subframes | 1,000 |
| Number of Slices | 4 |
| Channel Bandwidth (MHz) | 10 MHz ( 50 PRBs) |
| Probability of deadline missing threshold ($P_m^{max}$) | 0.1 |
| Static STIs | 10,20,30,40,50 |
| α | 0.5 |
| Percentage of PRBs to be traded by each slice | 50% |
| Default sample size $h_{dft}$ | 50 samples |
| Standard deviation threshold $Std$ | 0.2 |
| Upper autocorrelation threshold $A_{th,upper}$ | 0.8 |
| Lower autocorrelation threshold $A_{th,lower}$ | 0.7 |

**Table 4.5 Slices Service Level Agreement in simulation 2**

| NV Scheme | Slice 1 Alarm | Slice 2 Cameras | Slice 3 Sensors | Slice 4 Mixed | Slice 5 Shared (PRR only) |
|---|---|---|---|---|---|
| Trading | $t_1^{ref} = 0.2$ | $t_2^{ref} = 0.4$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | NA |
| Resource-based NVS | $t_1^{ref} = 0.2$ | $t_2^{ref} = 0.4$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | NA |
| Data rate-based NVS | $r_1^{ref}$ $Mbps$ | $r_2^{ref}$ $Mbps$ | $r_3^{ref}$ $Mbps$ | $r_4^{ref}$ $Mbps$ | NA |
| PRR | $t_1^{ref} = 0.2$ | $\boldsymbol{t_2^{ref} = 0.3}$ | $t_3^{ref} = 0.2$ | $t_4^{ref} = 0.2$ | $\boldsymbol{t_5^{ref} = 0.1}$ |

### 4.3.2 Experimental Results in the Static STIs Implementation

In the comparison, we focus on the deadline missing ratio, the deadline missing ratio per slice and the throughput for all the STIs values and the number of MTCDs. Figures 4.9 to 4.13 illustrate the deadline missing ratio for the 4 slicing techniques using 10,20,30,40 and 50 STIs respectively. In the trading scheme, the PRBs with the best CQIs are allocated to each slice based on the serving rate that minimizes the probability of deadline missing. Consequently, more data packets arrive before the deadline. For all STI values, the trading scheme performed better than all 3 NV schemes. This applies to all MTCDs numbers that we experimented with. In the worst case of largest STI of 50 TTIs, serving 700 MTCDs: the trading scheme achieved 35% deadline missing ratio, Resource-based NVS achieved 43%, Data rate-based NVS achieved 44%, and PRR achieved 45%. Comparing these resulted when using STI of 10 TTIs: Trading achieved also 35%, Resource-based NVS 36%, Bandwidth-NVS 40%, and PRR achieved 43%.

For both models of NVS and PRR, as the STI decreases, the slice weights are calculated more frequently so the schemes are able to adapt to the changes in traffic better. Unlike, NVS and PRR, the trading scheme performance is not degraded when using large STIs compared to smaller ones. This helps reduces the computational burden on the eNBs.

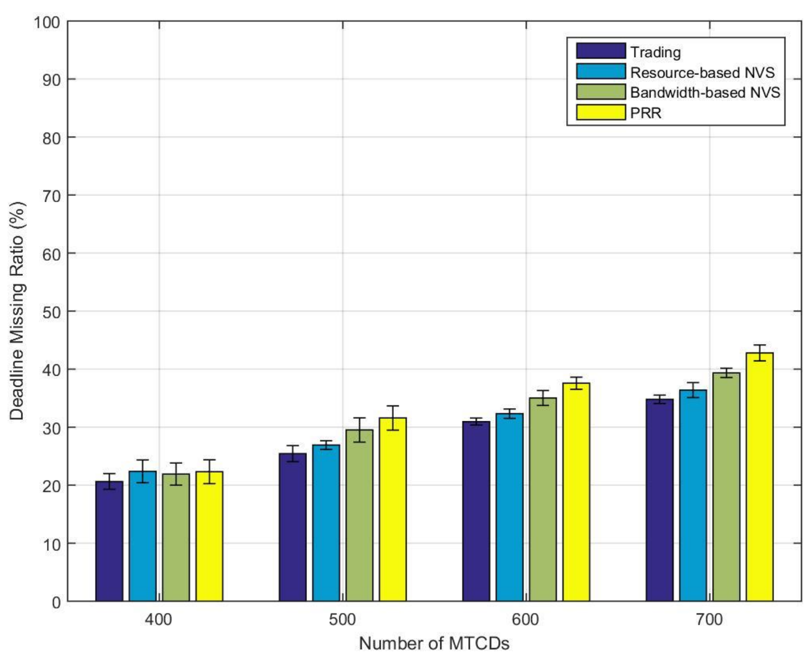


**Figure 4.9 Deadline missing ratio, STI= 10**

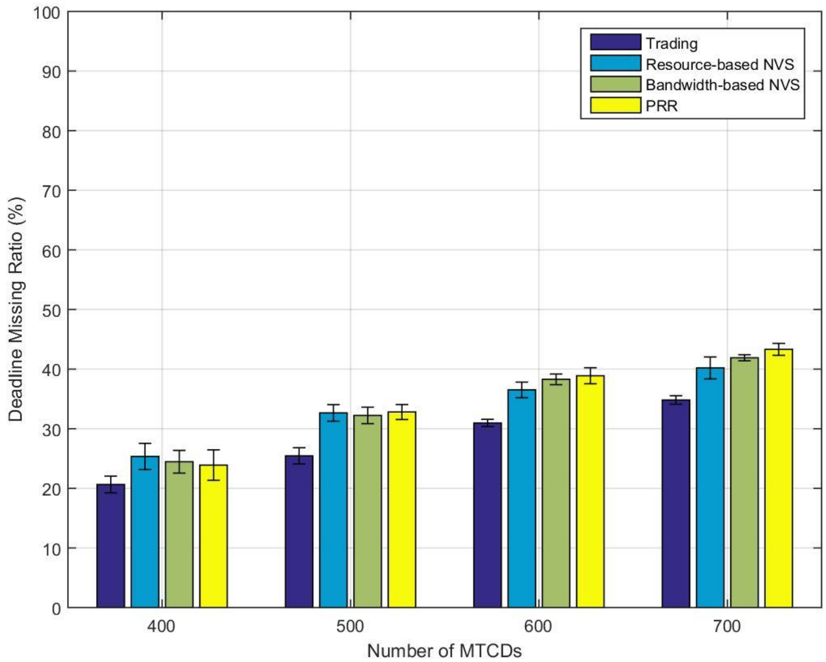


**Figure 4.10 Deadline missing ratio, STI=20**

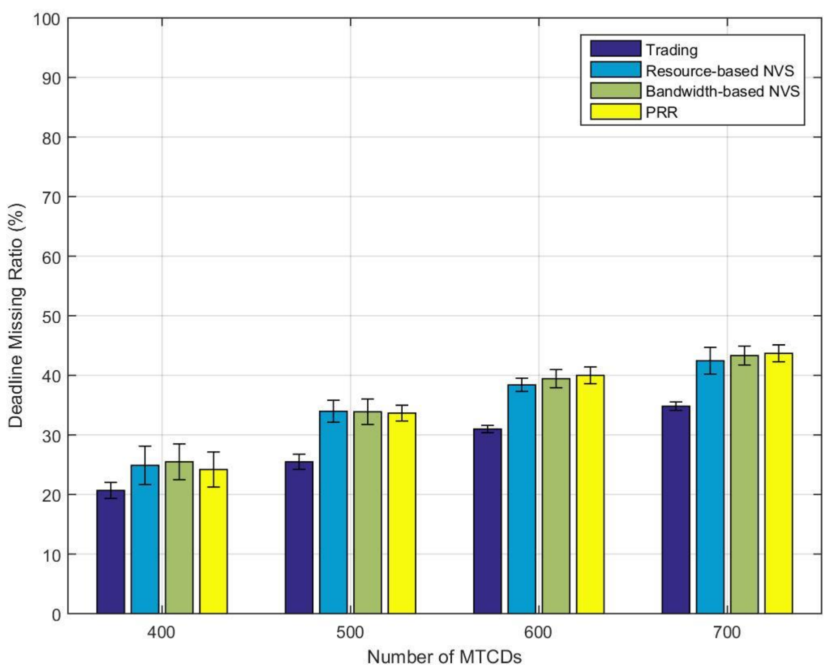


**Figure 4.11 Deadline missing ratio, STI=30**

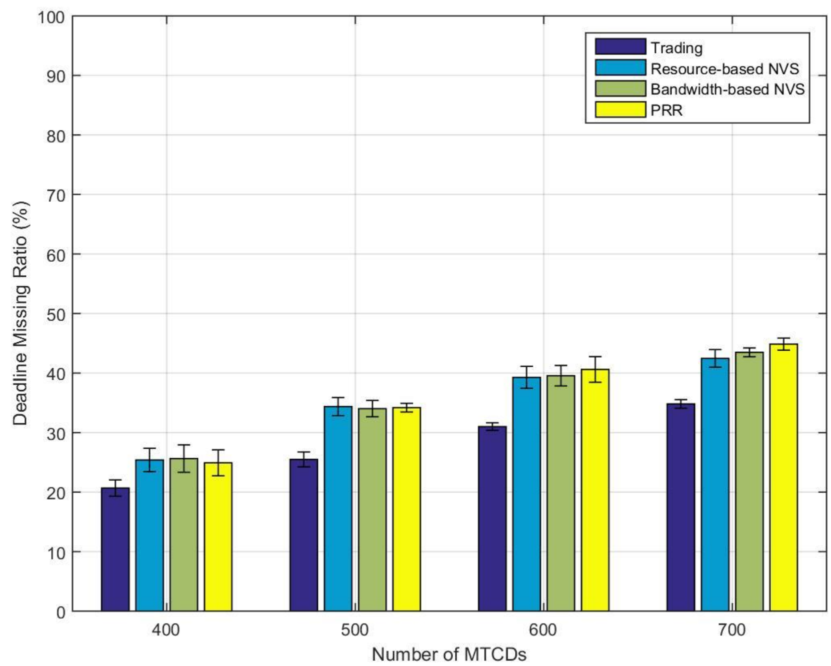


**Figure 4.12 Deadline missing ratio, STI=40**

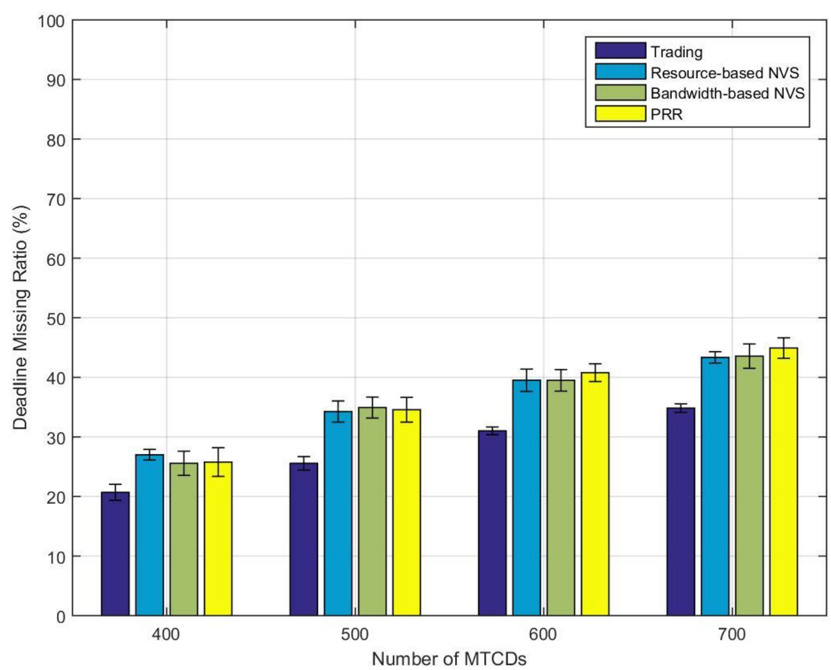


**Figure 4.13 Deadline missing ratio, STI=50**

Figures 4.14 to 4.18 illustrate the deadline missing ratio per slice when serving 700 MTCDs for STI of 10,20,30,40 and 50, respectively. The trading scheme performed better than all NV schemes in slices having rigid delay budgets such as alarm and surveillance camera slices. These results are consistent with the system deadline missing ratio. The trading scheme allocates the most suitable PRBs to the slices based its minimum serving rate (which is a function of the served MTCDs delay budgets). As the achievable throughput by the trades PRBs increased, naturally more packets arrived on time.

For example, when using STI of 50 TTIs to serve alarm MTCDs: the trading scheme deadline missing ratio did not exceed 26 %; whereas in PRR 49%, Data rate-based NVS 40% and Resource-based NVS 45%.

Likewise, when using STI of 50 TTIs to serve surveillance cameras: the trading scheme deadline missing ratio did not exceed 39%; whereas in PRR 47%, Data rate-based NVS 41% and Resource-based NVS 43%.

However, trading scheme did not perform as well as the rest of the slicing schemes in the mixed slice. This is due to the fact that the trading decisions are based on the median value of the PRBs achievable throughput and the median value of the MTCDs minimum serving rates since the serving rates are computed as a function of the MTCDs delay budgets. In slices where the MTCDs budgets vary widely, such as the mixed slice, the computed serving rates may not be accurate. In such cases, we recommend allocating a slice for each type of MTCD. So instead of allocating a single slice of (20% of PRBs) to the MVNO, we can allocate 3 slices ( 6%, 8%, 6% of PRBs) , each slice serves a certain type of MTCD. This allocation ensures that the trading decisions are based on more accurate median values.

Figures 4.19 to 4.23 illustrate the aggregate system throughput for STIs 10,20,30,40 and 50 respectively. As the trading utility is a function of the achievable throughout of the traded PRBs, the GA maximizes the utility by allocating the PRBs that can achieve the highest achievable throughout to the slices. Consequently, the aggregate system throughput increases. This is more evident when larger STIs are used. For comparison, the throughput achieved when serving 700 MTCDs, in the case of an STI of 10 TTIs: the trading, Resource based NVS, Data rate-based NVS and PRR schemes achieved: 21, 20, 20 and 19 Mbps respectively. However, when using 50 TTIs, the NV models performance degraded.

Resource based NVS achieved 17Mbps, Data rate-based NVS achieved 18Mbps and PRR achieved 17Mbps. The trading scheme performance remained the same.

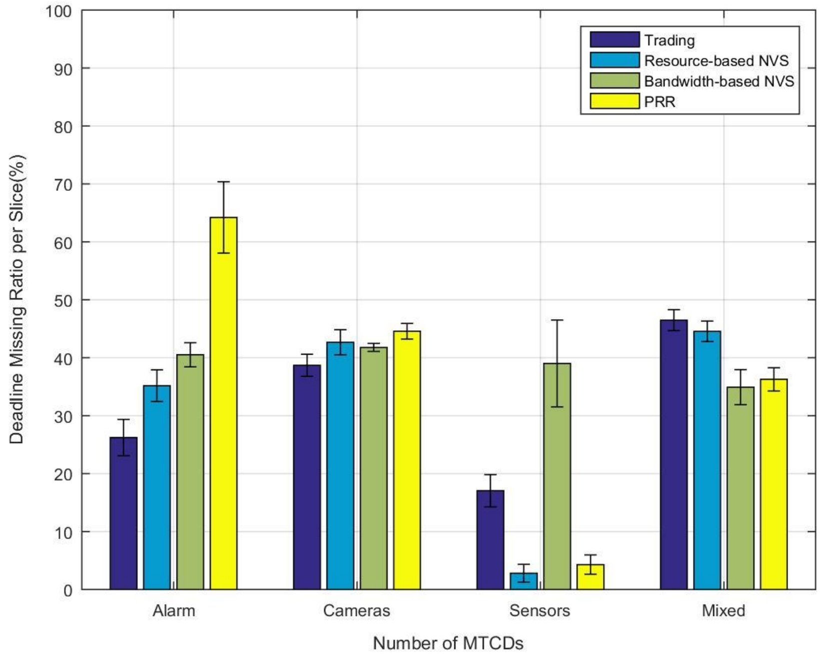


**Figure 4.14 Deadline missing ratio per slice, MTCDs=700, STI=10**

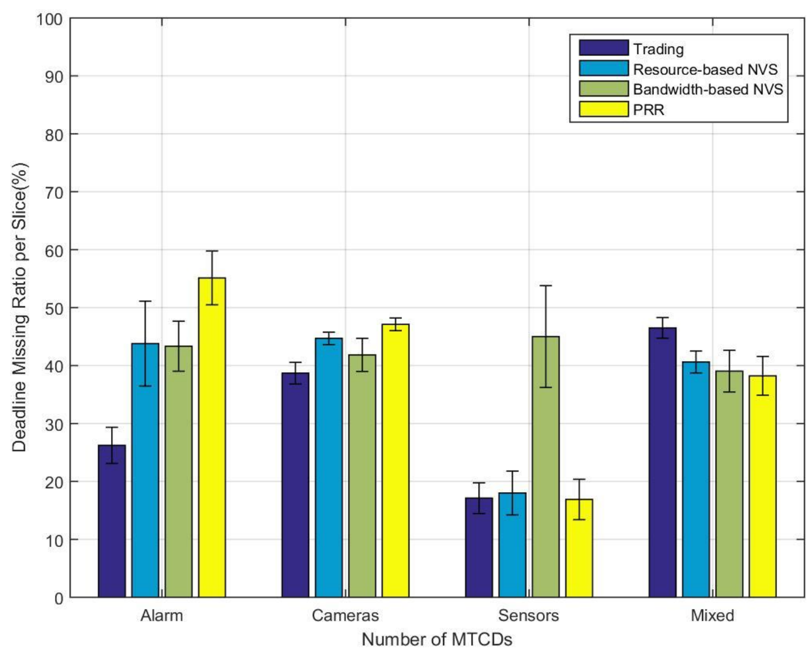


**Figure 4.15 Deadline missing ratio per slice, MTCDs=700, STI=20**

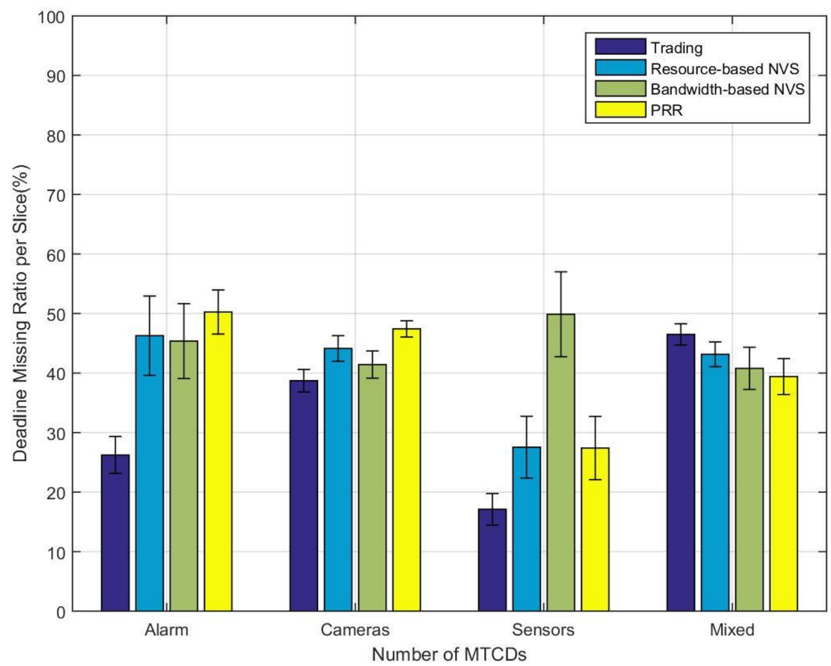


**Figure 4.16 Deadline missing ratio per slice, MTCDs=700, STI=30**

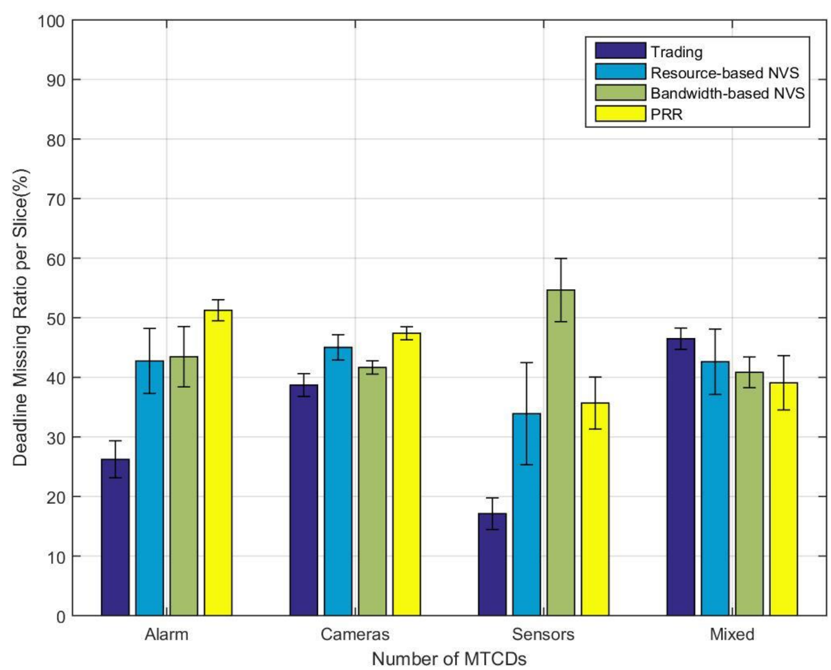


**Figure 4.17 Deadline missing ratio per slice, MTCDs=700, STI=40**

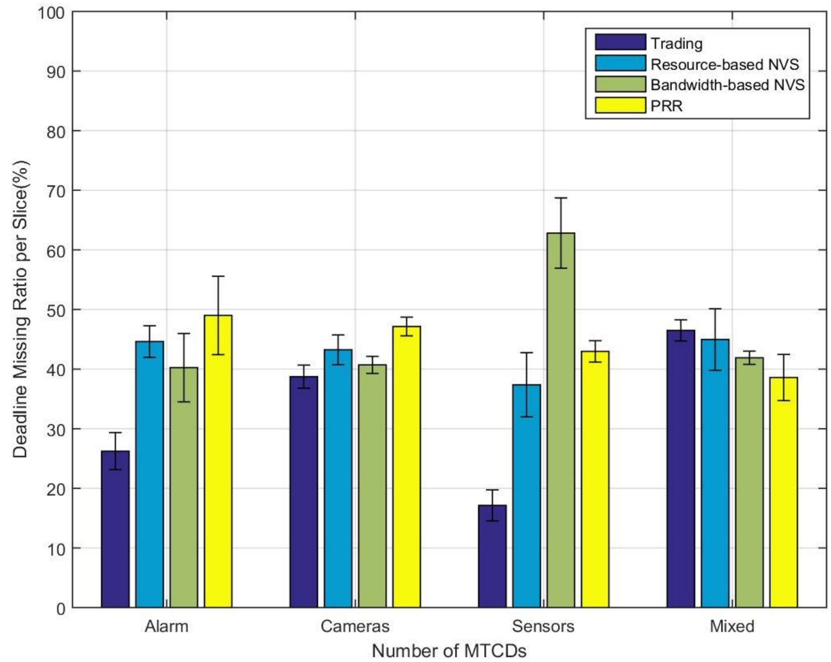


**Figure 4.18 Deadline missing ratio per slice, MTCDs=700, STI=50**

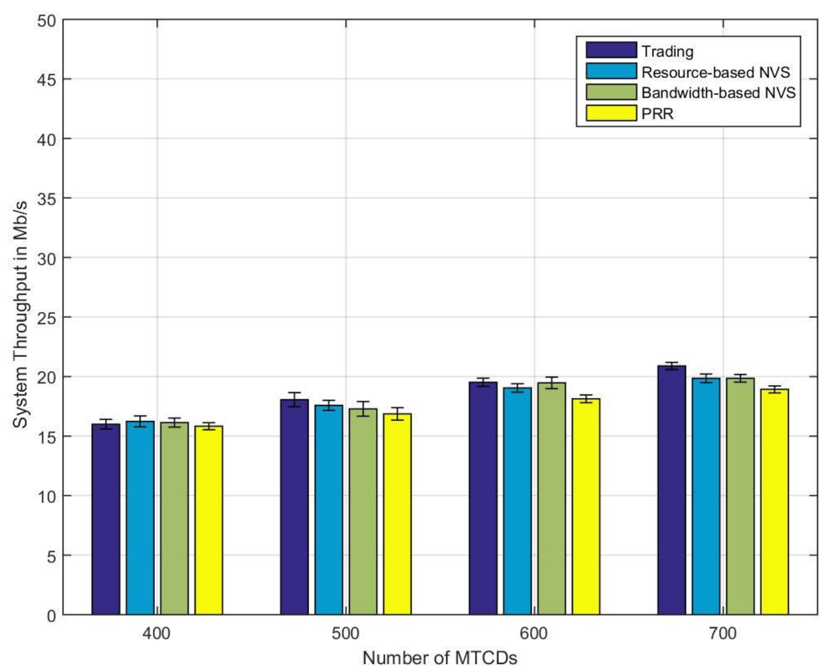


**Figure 4.19 System Throughput, STI=10**

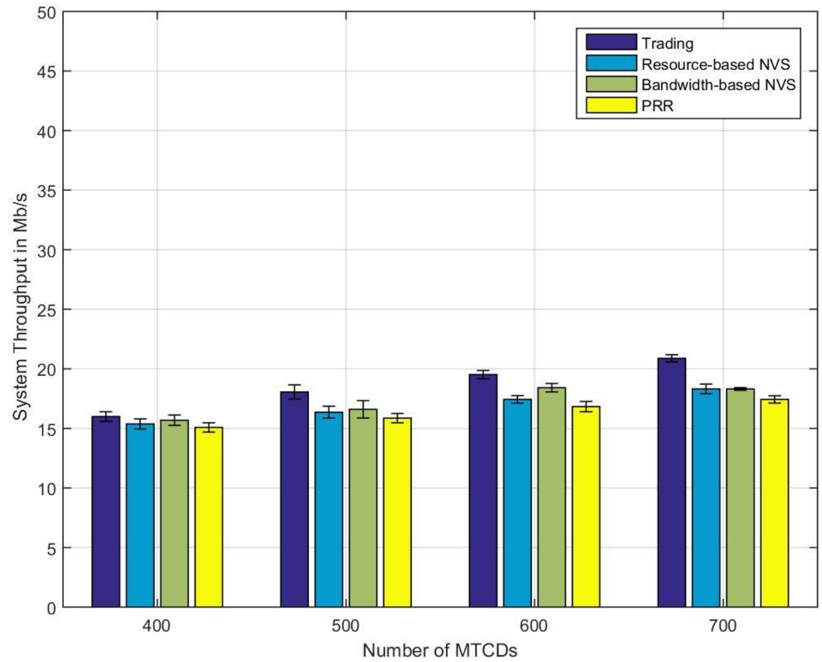


**Figure 4.20 System Throughput, STI=20**

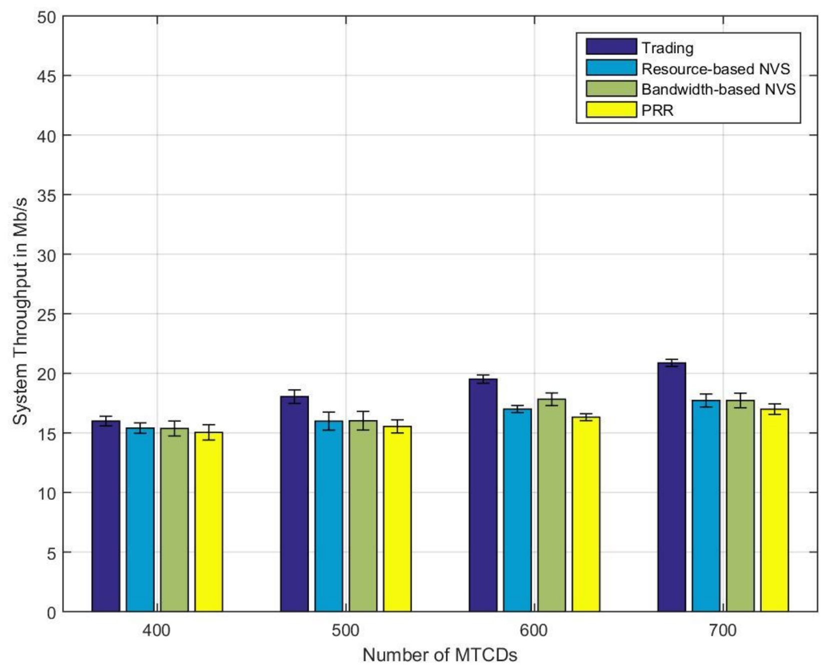


**Figure 4.21 System Throughput, STI=30**

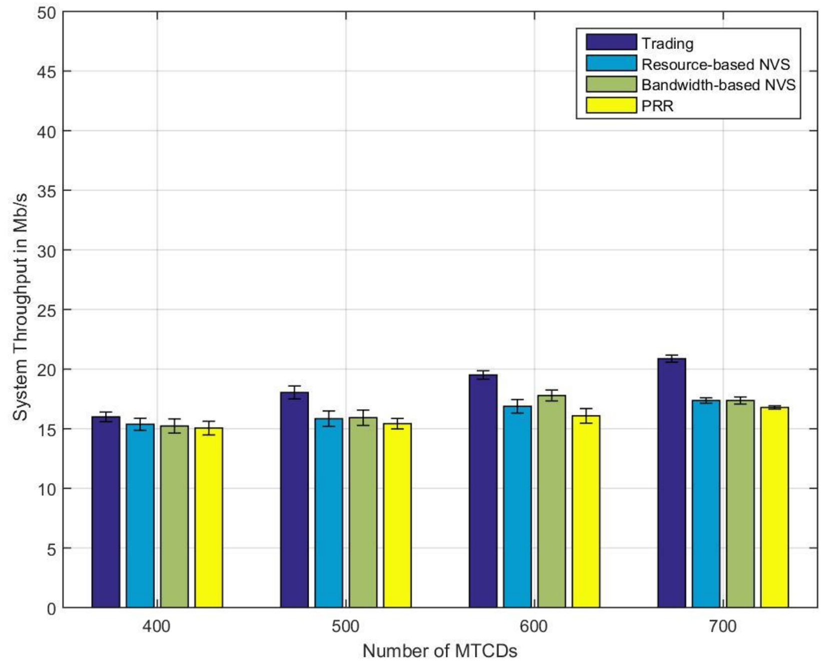


**Figure 4.22 System Throughput, STI=40**

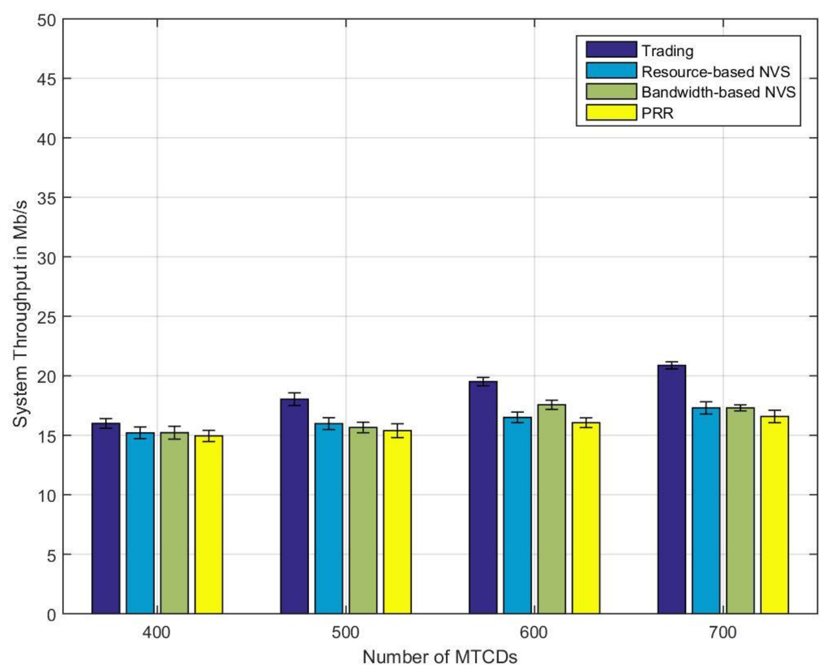


**Figure 4.23 System Throughput, STI=50**

### 4.3.3 Experimental Results in the Dynamic STI Implementation

We now evaluate the same parameters for dynamic STIs. Starting with the deadline missing ratio. Consistently with the static results, the trading scheme performs better than Resource-based, Data rate-based and PRR. As shown in Figure 4.24, the trading scheme achieved the lowest deadline missing ratios across all MTCDs numbers.

Figures 4.25 to 4.28 illustrate the deadline missing per slice when serving 400,500,600 and 700 MTCDs respectively. The deadline missing per slice, when using the trading scheme, is mostly smaller than the rest of the slicing schemes especially in the cases of alarm and surveillance cameras slices.

As explained earlier, the deadline missing ratio is higher in the mixed slice due to the large difference in delay budgets between the various MTCDs types; this that led to inaccurate trading decisions.

Lastly, Figure 4.29 illustrates the aggregate system throughput when serving 100,200,300 and 400 MTCDs using dynamic STI calculations. Similar to static STIs, the trading scheme achieved the highest throughput among all the slicing scheme.

These result show that regardless of the STI size, the trading scheme is better adapted to M2M traffic.

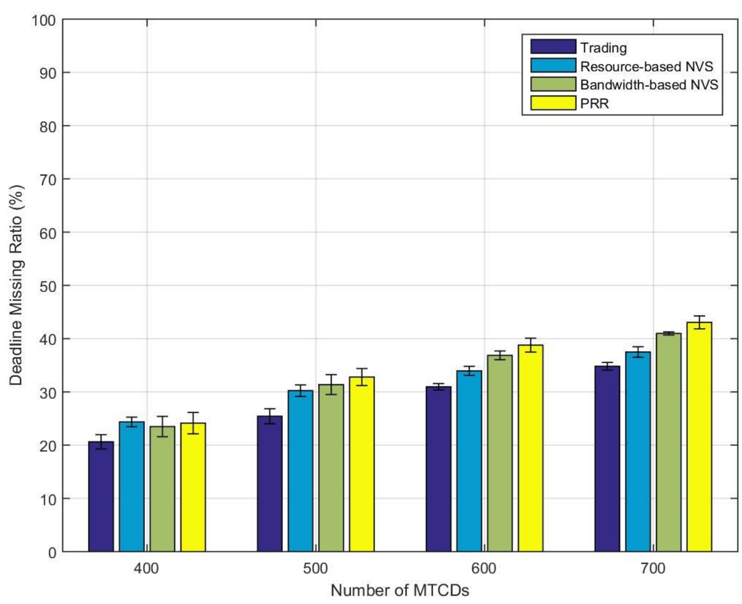


**Figure 4.24 Deadline missing ratio, dynamic STI**

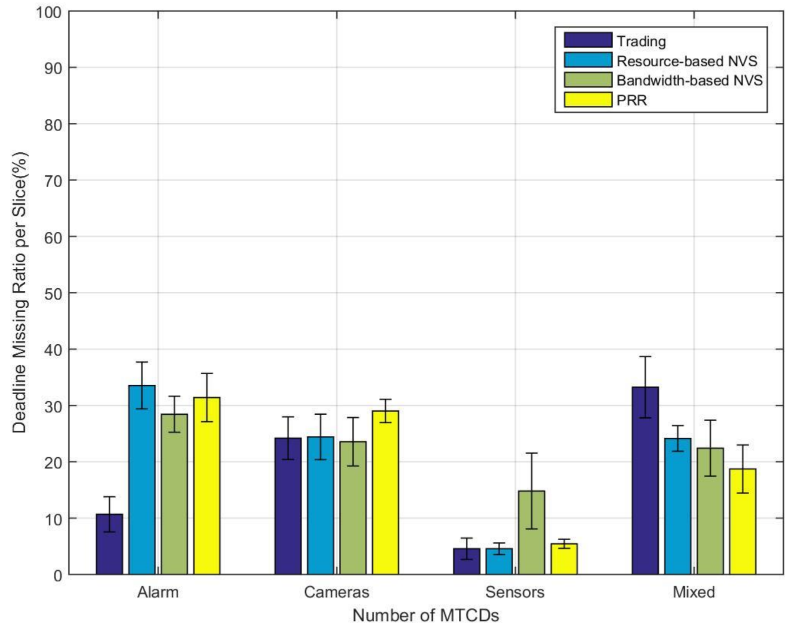


**Figure 4.25 Deadline missing ratio per slice, MTCDs=400, dynamic STI**

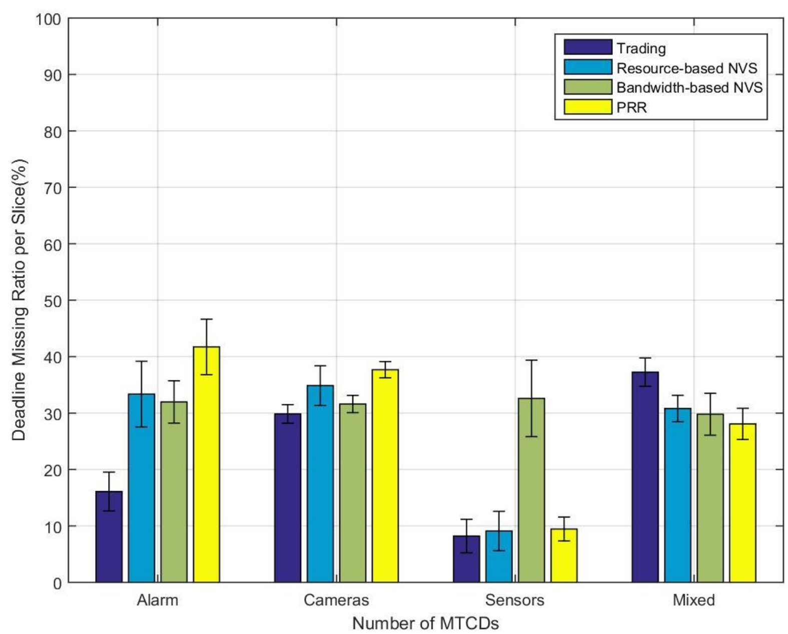


**Figure 4.26 Deadline missing ratio per slice, MTCDs=500, dynamic STI**

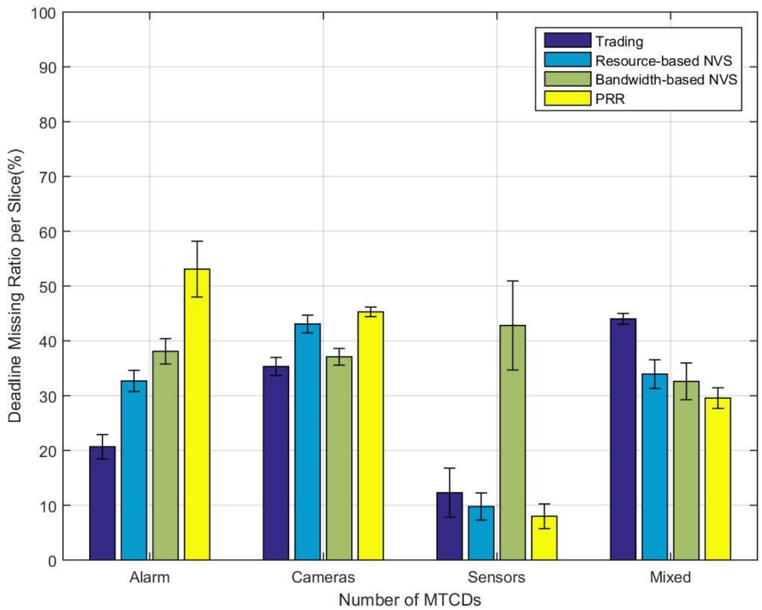


**Figure 4.27 Deadline missing ratio per slice, MTCDs=600, dynamic STI**

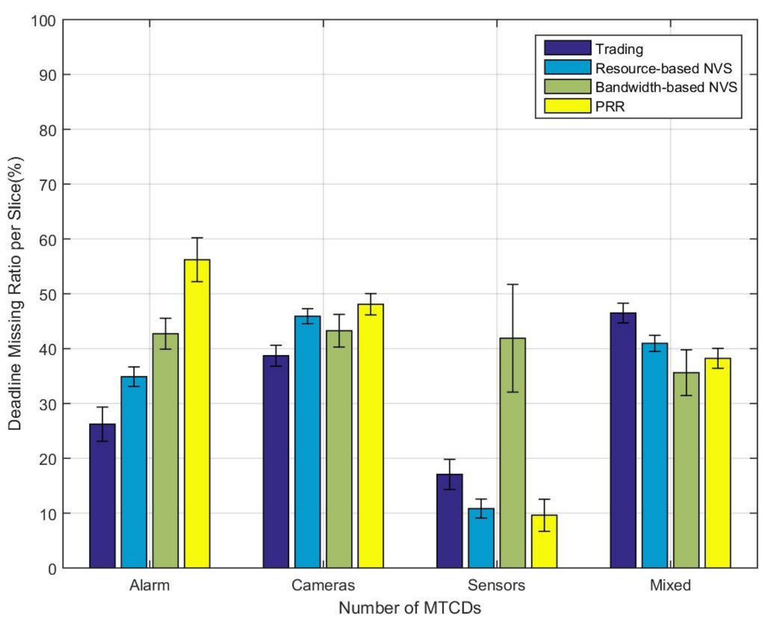


**Figure 4.28 Deadline missing ratio per slice, MTCDs=700, dynamic STI**

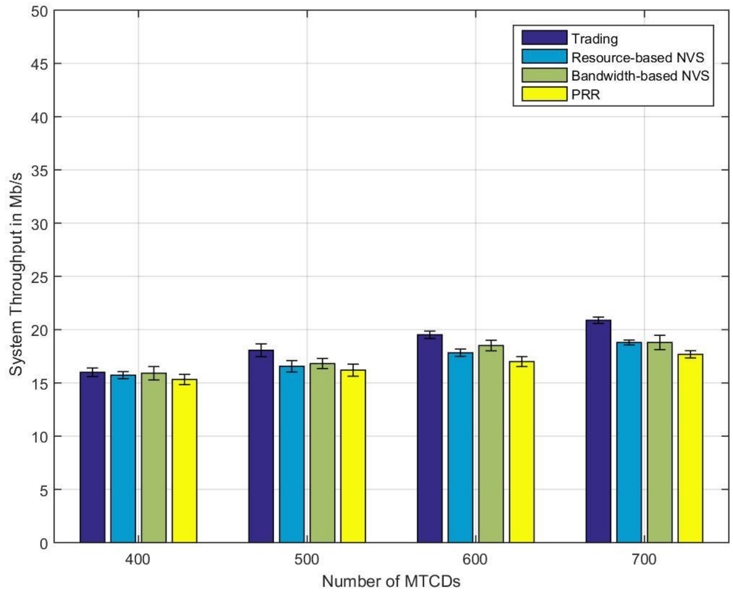


**Figure 4.29 Overall system throughput, dynamic STI**

## 4.4 Chapter Summary

In this chapter, we presented the experimental results of simulating Resource-based, Data rate-based and PRR and comparing them to the static reservation of resources. These results showed that NV is a good candidate for M2M deployments with minor modifications on the actual schemes. In the second part of the chapter, we discussed the experimental results of the proposed trading scheme compared to the 3 adapted NV schemes. The trading scheme showed promising results compared to the other slicing schemes from deadline missing perspective. Thanks to its ability to allocate the most suitable PRBs to the sharing slices; their performance have significantly improved.

# Chapter 5: Conclusion and Future Work

## 5.1 Conclusion

Many studies predict that IoT is the future of wireless communication. Billions of IP-aware devices are expected to be connected to the internet in the near future. The proliferation of the IoT technology presents multiple challenges. The most important challenge is to adapt to the dramatic increase of the mobile date traffic. The other challenge is adapting the current technologies to the needs of the M2M traffic.

Network virtualization and radio resources slicing are potential candidates for M2M deployment. They provide solutions for the sharing virtual operators to share their radio resources efficiently without affecting the QoS they offer to their induvial users. However, the known NV schemes such as NVS and PRR were designed for H2H traffic. M2M traffic is event-driven and delay-sensitive.

In this thesis, we addressed this problem in two stages.

First, we presented approaches to adapt Resource-based and Data rate-based NVS models and PRR to MTC. In Resource-based NVS, the portion of PRBs allocated to each slice depended on its delay budget and packet size. In Data rate-based NVS, the bandwidth allocated to each slice was calculated as a function of the probability of deadline missing. Both approaches succeeded to improve the performance of NVS from deadline missing perspective. To adapt PRR to MTC, the MTCDs with rigid delay budgets were given higher priority to be assigned PRBs from the shared slice. PRR showed promising results compared to NVS and SR.

The simulation results have shown that NVS and PRR can be easily adapted to M2M traffic with minor modifications on the current schemes. Resource-based NVS, Data rate-based NVS and PRR achieved lower deadline missing ratio and higher throughput compared to static reservation. In terms of resource utilization, the simulation showed that the PRBs were best utilized by the NV schemes thanks to their ability to reallocate unused resources from one slice to the other slice.

Second, we presented a novel trading scheme that allows the sharing operators to trade their resources. We used GA optimization to trade the PRBs between the slices based on their reported CQIs while maintaining the same number of PRBs allocated to each slice. The trading decisions were based on the ability of the traded resources to minimize the probability of deadline missing. We compared the performance of the trading scheme to the 3 adapted NV schemes. The trading scheme showed strong results.

Compared to Resource-based NVS, Data rate-based NVS and PRR, the trading scheme has achieved lower deadline missing ratio and higher throughput. These results are applicable on all STI values and MTCDs configurations. It is worth mentioning that unlike NVS and PRR, the trading schemes maintain the quality of its performance even if the STIs are large. This quality significantly improves the computation load on the eNB.

We also presented a solution to allow the virtual operators to initiate the trade at dynamic time intervals. The slicing time intervals were computed as a function of the auto-correlation of the traffic. In times where the traffic was changing slowly, the slicing interval was long reducing the computational load on the eNB. Whereas, in times where the traffic was changing rapidly; the slicing interval was small to allow the slices to evaluate their performances more frequently and adapt accordingly. The results of the dynamic implementation were consistent with the static implementation. The trading scheme has performed better than the other 3 slicing schemes in terms of deadline missing and throughput.

We therefore provided evidence that allowing multiple operators to share and trade resources is a promising approach for accommodating large-scale IoT deployments.

## 5.2 Future Work

The slicing can be further investigated from the below perspectives:

- Service Level Agreement estimation

  In the adapted network virtualization schemes, we calculated the reserved bandwidth based on the sum of the minimum serving rates of the deployed MTCDs. However, in many M2M implementations, the MTCDs can stay idle for long periods of time. We need further investigation to calculate the reserved bandwidth as a sum of the minimum serving rates of the active machines. This requires further analysis of the traffic pattern of the deployed MTCDs.

- End-to-end slicing

  In this work we focus on RAN slicing. As 5G will enable end-to-end slicing, we need to expand this work to investigate how end-to-end slicing and resource trading can be better adapted to M2M needs.

- MTCDs scheduling

  The thesis scope is only the high-level slice scheduling i.e. how the PRBs are allocated to the slice. The schedulers used within each slice to allocate the PRBs to the MTCDs may also considered in future work and hence creating what is considered as two-tier resource allocation.

The trading scheme also needs further investigation from the below perspectives:

- Dynamic slice reservation

  The trading scheme is based on static reservation. Each virtual operator reserves a permanent portion of PRBs. The trading scheme maintains the same number of PRBs allocated to each slice. We suggest in the future work that the slice size allocated to each virtual operator changes periodically. This will improve the radio resources utilization by avoiding reserving unused PRBs to one slice.

- Power constraints

  As MTCDs have power constraint due to their limited battery life. We suggest looking into trading that also optimizes the power consumption.

# References


[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022," 2019.

[2] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *Communications Magazine*, vol. 51, no. 7, pp. 27-35, 2013.

[3] 3GPP TS 22.951, "Service Aspects and Requirements for Network Sharing," 2012.

[4] V. G. T. N. Vidanagama, D. Arai and T. Ogishi, "Service Environment for Smart Wireless Devices: An M2M Gateway Selection Scheme," *IEEE Access*, vol. 3, pp. 666-677, 2015.

[5] R. Kokku, R. Mahindra, Z. Honghai and S. Rangarajan, "NVS: A Substrate For Virtualizing Wireless Resources in Cellular Networks," *IEEE/ACM Transactions Networking*, vol. 20, no. 5, pp. 1333 - 1346, 2012.

[6] T. Guo and R. Arnott, "Active LTE RAN Sharing with Partial Resource Reservation," in *IEEE 78th Vehicular Technology Conference*, Las Vegas, 2013.

[7] S. Gendy and Y. Gadallah, "LTE-based Network Virtualization Schemes Adaptation for M2M Deployments," in *Proceedings of IEEE BlackSeaCom*, Sochi, Russia, 2019.

[8] S. Gendy and Y. Gadallah, "A Resouce Trading Scheme in Slice-Based LTE M2M Communications," *To Appear in Proceedings of VTC 2019-Fall*, Honlulu, USA, September 2019.

[9] C. Cox, An Introduction to LTE,LTE-Advanced,SAE,VoLTE and 4G Mobile Communications, John Wiley & Sons, 2014.

[10] A. Najah , A. M. Taha, M. Salah and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1239 - 1265, 2013.

[11] ITU, "M.2083: IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," 07 December 2017. [Online]. Available: https://www.itu.int/rec/R-REC-M.2083. [Accessed 29 July 2019].

[12] A. Khlifi and R. Bouallegue, "HYBRID LS-LMMSE CHANNEL ESTIMATION Technique for LTE Downlink System," *International Journal of Next-Generation Networks*, vol. 3, no. 4, pp. 1-13, 2011.

[13] Y. Gadallah, M. H. Ahmed and E. Elalamy, "Dynamic LTE resource reservation for critical M2M Deployments," *Pervasive and Mobile Computing*, vol. 40, pp. 541-555, 2017.

[14] 3GPP , "Physical layer procedures (Release 10)," TR 36.213, version 10.1.0, 2011.

[15] M. A. Mehaseb, Y. Gadallah, A. Elhamy and H. Elhenawy, "Classification of LTE Uplink Scheduling Techniques: An M2M Perspective," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1310 - 1335, 2016.

[16] A. E. Mostafa and Y. Gadallah, "A Statistical Priority-Based Scheduling Metric for M2M Communications in LTE Networks," *IEEE Access*, vol. 5, pp. 8106 - 8117, 2017.

[17] K. Zheng, F. Hu, W. Wang, W. Xiang and M. Dohler, "Radio Resource Allocation in LTE-Advanced Cellular Networks with M2M Communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184 - 192, 2012.

[18] 3GPP TR 22.852, "3GPP System Architecture Working Group 1 (SA1) RAN Sharing Enhancements Study Item.," 2012.

[19] G. Tseliou, F. Adelantado and C. Verikoukis, "Resources negotiation for network virtualization in LTE-A networks," in *IEEE International Conference on Communications*, Sydney, 2014.

[20] G. Tseliou, F. Adelantado and C. Verikoukis, "Scalable RAN Virtualization in Multi-Tenant LTE-A Heterogeneous Networks," *IEEE Transactions on Vehicular Technology,*, vol. PP, no. 99, p. 1, 2015.

[21] M. Li, L. Zhao, X. Li, X. Li, Y. Zaki, A. Timm-Giel and C. Görg, "Investigation of Network Virtualization and Load Balancing Techniques in LTE Networks," in *IEEE 75th Vehicular Technology Conference*, Yokohama, 2012.

[22] R. Kokku, R. Mahindra, H. Zhang and S. Rangarajan, "CellSlice: Cellular Wireless Resource Slicing for Active RAN Sharing," in *Fifth International Conference on Communication Systems and Networks*, Bangalore, 2013.

[23] Y. Gadallah, M. M. Ahmed and E. Elalamy, "Dynamic LTE resource reservation for critical M2M Deloyements," *Pervasive and Mobile Computing*, vol. 40, pp. 541-555, 2017.

[24] D. Sattar and A. Matrawy, "Optimal Slice Allocation in 5G Core Networks," *IEEE Networking Letters,*, vol. 1, no. 2, pp. 48-51, 2019.

[25] T. Guo and A. Suárez, "Enabling 5G RAN Slicing With EDF Slice Scheduling," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2865 - 2877, 2019.

[26] B. Nour, A. Ksentini, N. Herbaut, P. A. Frangoudis and H. Moungla, "A Blockchain-Based Network Slice Broker for 5G Services," IEEE Networking Letters, 2019.

[27] Q. Feng, D. He, S. Zeadaly, M. K. Khan and N. Kumar, "A Survey on Privacy Protection in Blockchain System," *The Journal of Network and Computer Applications*, vol. 126, pp. 45-58, 2019.

[28] M. Lopez-Martınez, J. J. Alcaraz, J. Vales-Alonso and J. Garcia-Haro, "Automated spectrum trading mechanisms: understanding the big," *Wireless Networks*, vol. 21, no. 2, p. 685–708, 2015.

[29] X. Zhou and H. Zheng, "TRUST: A General Framework for Truthful Double Spectrum Auctions,," in *IEEE INFOCOM 2009*, Rio de Janeiro, 2009.

[30] X. Zhou, S. Gandhi, S. Suri and H. Zheng, "eBay in the Sky: Strategy-Proof Wireless Spectrum Auctions," in *MobiCom08*, San Francisco, 2008.

[31] S. Wang, P. Xu, X. Xu, X. Tang and x. Liu, "TODA: Truthful Online Double Auction for Spectrum Allocation in Wireless Network," in *IEEE Symposium on New Frontiers in Dynamic Spectrum*, Singapore, 2010.

[32] M. Shih, K. D. Huang, C. Yeh and H. Wei, "To Wait or To Pay: A Game Theoretic Mechanism for Low-Cost M2M and Mission-Critical M2M," *IEEE Transactions on Wireless Communications,* vol. 15, no. 11, pp. 7314 - 7328, 2016.

[33] T. M. Ho, N. H. Tran, S. M. Kazmi and C. S. Hong, "Dynamic pricing for resource allocation in wireless network virtualization: A Stackelberg game approach," in *International Conference on Information Networking (ICOIN)*, Da Nang, 2017.

[34] H. von Stackelberg, Market Structure and Equilibrium, Berlin: Springer, 2011.

[35] M. E. Tarerefa, O. Falowo and N. Ventura, "Application-Aware Game Theoretic Pricing Algorithm for Cellular Machine-to-Machine Communications," in *14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol, 2018.

[36] M. Pan, S. Liang, H. Xiong, J. Chen and G. Li, "A Novel Bargaining Based Dynamic Spectrum Management Scheme in Reconfigurable Systems," in *2006 International Conference on Systems and Networks Communications*, Tahiti, 2006.

[37] Y. Yan, J. Huang and J. Wang, "Dynamic Bargaining for Relay-Based Cooperative Spectrum Sharing," *IEEE Journal on Selected Areas in Communications,,* vol. 31, no. 8, pp. 1480-1493, 2013.

[38] 3GPP R1-120056, "Analysis on traffic model and characteristics for MTC and text proposal," TSG-RAN Meeting, WG1#68, Dresden, 2012.

[39] G. Basilashvili, "Study of Spectral Efficiency for LTE Network," *American Scientific Research Journal for Engineering, Technology, and Sciences,* vol. 29, no. 1, pp. 21-32, 2017.

[40] M. Y. Abdelsadek, Y. Gadallah and M. H. Ahmed, "An LTE-Based Optimal Resource Allocation Scheme for Delay-Sensitive M2M Depoyments Coexistent with H2H Users," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Atlanta, 2017.

[41] F. Meng and S. Wu, "Research of Genetic Algorithm in function optimization based on HCI," in *IEEE International Symposium on IT in Medicine and Education*, Xiamen, 2008.