NOVEL AI-BASED FRAMEWORK FOR BALANCING TRAFFIC DEMAND AND SERVICE PROCESSES IN UNLICENSED BANDS

INVENTORS:

MOHAMMED HIRZALLAH

BALKAN KECICIOGLU

BELAL HAMZEH

BERNARDO HUBERMAN

Wi-Fi Optimization Using AI Based Traffic Flow Prediction

In a Wi-Fi network, the probability of a STA transmitting is a function of the contention window size *CW.* A collision in the network is dependent on two main factors, the number of stations *n* and the contention window size *CW*…. Closed form solutions that estimate the network throughput and collision probabilities are readily available. Network throughput is given in the equation below:

$$\theta = \frac{P_{\text{succ}}\left(T_{\text{hdr}} + T_{\text{pkt}}\right)}{\left(1 - P_{\text{tr}}\right) T_{\text{idle}} + P_{\text{succ}} T_{\text{succ}} + P_{\text{coll}} T_{\text{coll}}},$$

Which provides the optimum contention window size (CW*) to maximize network performance, and is given by

$$\text{CW}^* = \left(\frac{\text{CW}_{\text{min}}}{2}\right) \times n - 1,$$

where CWmin is defined by the 802.11 standard, and n is the number of active stations.

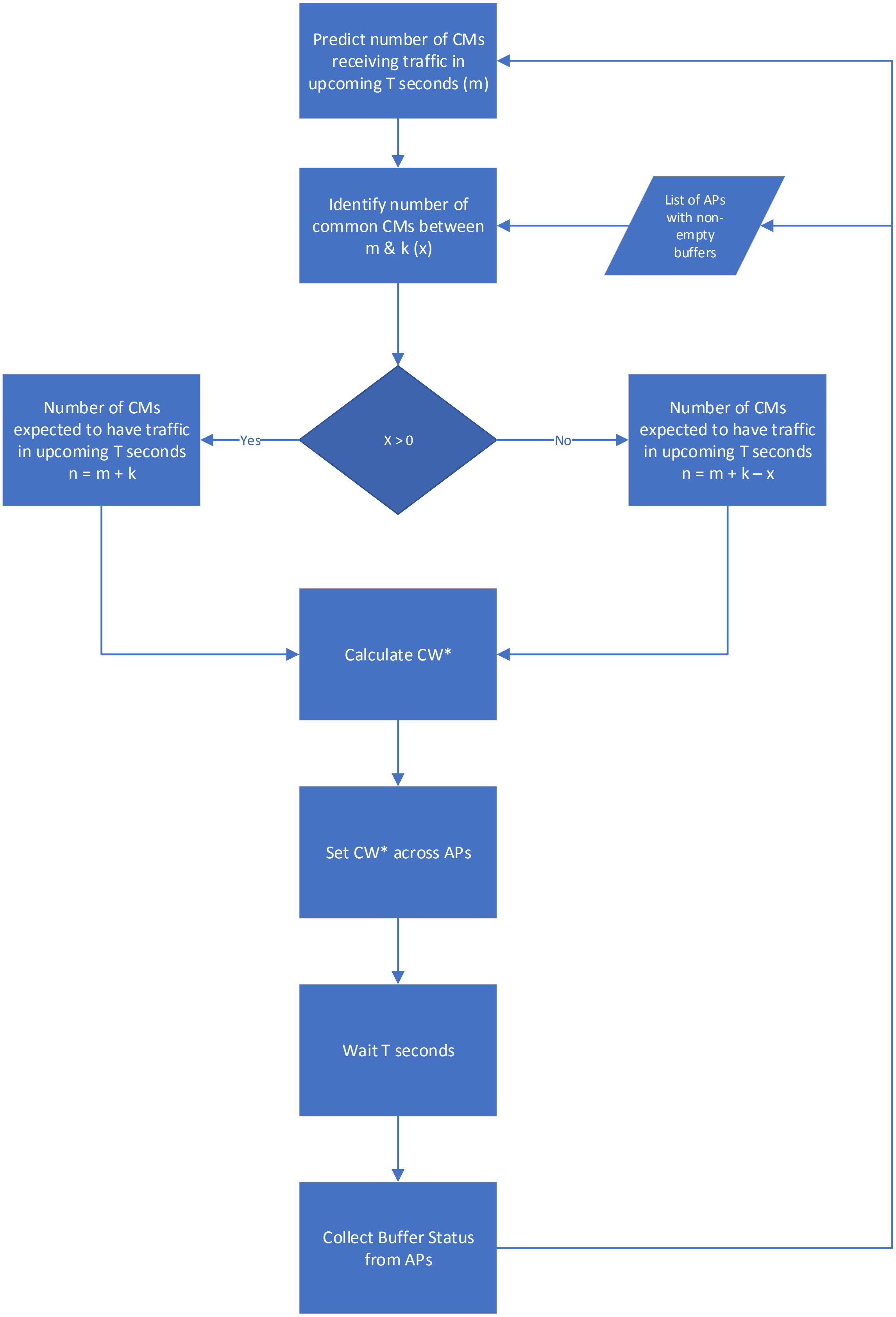
Across a Wi-Fi network, the value of CW will vary across the stations based on their transmission history, thus optimum network performance is near impossible to achieve.

By leveraging the ability to predict when a CM will be receiving data, we can predict the number of the number of stations that will have traffic during an observation window *T*. With accurate prediction of the number of active stations and combined with the analytical model above, CW* can be optimized per each observation window thus reducing the possibility of collision, thus increasing the overall network throughput and QoS.

To implement this, the follownig elements are required:

- Traffic prediction engine that is capable of predicting whether a CM will be receiving traffic in the upcoming T seconds, and capable of doing this across all CMs in area of interest
- CW setting engine which calculates the optimum value of the contention window and communicates it to all APs in area of interest.
- Communication from the APs to the CW setting engine indicating whether they have remaining traffic in their buffers.

The flow of the process is shown in the following flowchart



Predict number of CMs receiving traffic in upcoming T seconds (m)
Identify number of common CMs between m & k (x)
List of APs with non-empty buffers
X > 0
Yes
No
Number of CMs expected to have traffic in upcoming T seconds n = m + k
Number of CMs expected to have traffic in upcoming T seconds n = m + k – x
Calculate CW*
Set CW* across APs
Wait T seconds
Collect Buffer Status from APs

Novel AI-based Framework for Balancing Traffic Demand and Service Processes in Unlicensed Bands

Mohammed Hirzallah, Balkan Kecicioglu, Belal Hamzeh, Bernardo Huberman

Draft rev 1 (last updated on Dec 28th, 2018 by Mohammed)

Abstract

This idea proposes methods for balancing traffic demand and service processes for systems operating in unlicensed bands. Unlicensed channels are shared opportunistically by many heterogeneous systems, prohibiting them from providing guaranteed QoS service. We introduce an artificial intelligence (AI)-based framework for classifying and predicting future traffic demand and service processes as well as controlling channel access parameters to meet QoS requirements. We characterize the traffic demand process based on traffic streaming statistics and characterize the traffic service process based on buffer occupancy and queue dynamics. Predictions of future streaming statistics and buffer occupancy dynamics are leveraged to control various channel access parameters and provision various QoS applications and/or meet a harmonious cross-technology coexistence among systems sharing unlicensed bands.

Description

Motivation

Although enforcing different devices to wait for random durations and following a listen-before-talk (LBT) -based procedure prior to their access to unlicensed channels reduce collision rates, they still reduce the utilization of unlicensed channels. The heterogeneity in applications and their stringent QoS requirements requires adapting channel access parameters continuously overtime. Systems operating in unlicensed bands can control their channel access parameters, including the Initial interframe space (IFS), contention window size (CW), duration of transmit opportunity (TXOP) over time, MAC packet aggregation level, energy and preamble detection thresholds, to improve the quality of experience seen by their customers. However, it is not clear how these parameters should be adapted over time, due to the lack of knowledge about future variations in traffic demand process as well as lack of knowledge about traffic service rate. Traffic service process depends on the conditions of unlicensed channels and the behavior of other systems sharing these channels.

Characterization of Traffic Demand Process:

Traffics flowing through wired and wireless networks depend on human’s interaction with data and internet. Although different applications, such as interactive voice & video, multimedia streaming, online gaming, file transfer, etc., dictate stringent streaming and delivery requirements, they can still be characterized according to the following metrics, as shown in Figure 1:

1. Packet size
2. Packet arrival rate

3. Inter-packet time

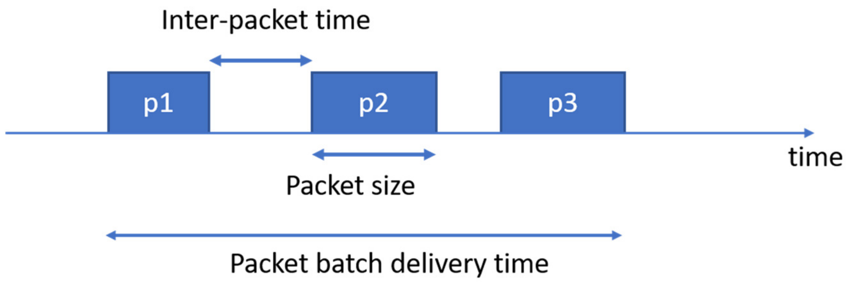


Figure 1 Characterization of traffic demand process.

In order to characterize traffic demand process, we consider a time-based sliding window structure for monitoring traffic demand process over time. This monitoring structure consists of two adjacent sliding windows that pass over traffic received by access points (APs); i.e., Past-Window, Future-Window, as shown in Figure 2. We divide each window into several time slots, and measure the following traffic statistics within each slot:

1. Mean of packet size
2. Variance of packet size
3. Mean of inter-packet delay
4. Variance of inter-packet delay
5. Number of packets

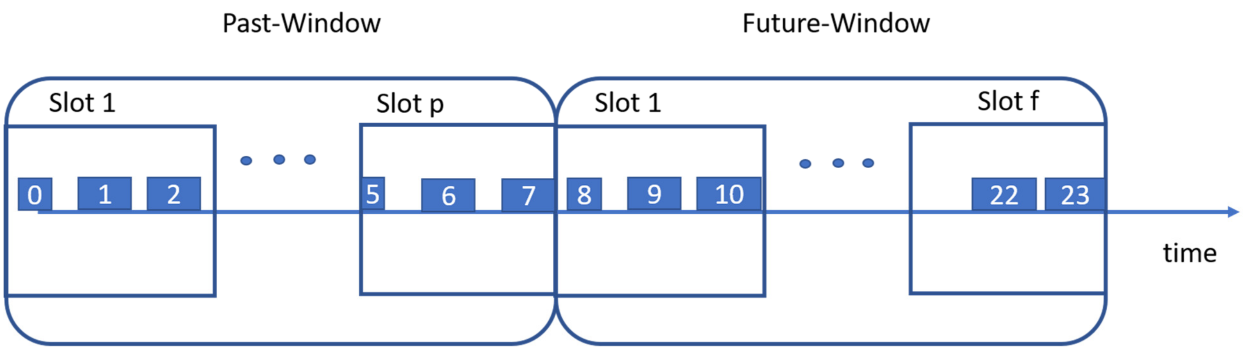


Figure 2 Sliding-window structure used to monitor traffic demand process (blue boxes are packets).

Characterization of Traffic Service Process:

Buffer occupancy and queue dynamics at transmitters provide an insight on channel condition and how successful the transmitter is accessing the unlicensed channel and serving traffic demands, as shown in Figure 3. We characterize buffer occupancy and queue dynamics according to following metrics:

1. Mean of queue length over time

2. Variance of queue length over time
3. Mean of packet queue waiting time
4. Variance of packet queue waiting time
5. Mean of packet service time
6. Variance of packet service time
7. Buffer occupancy

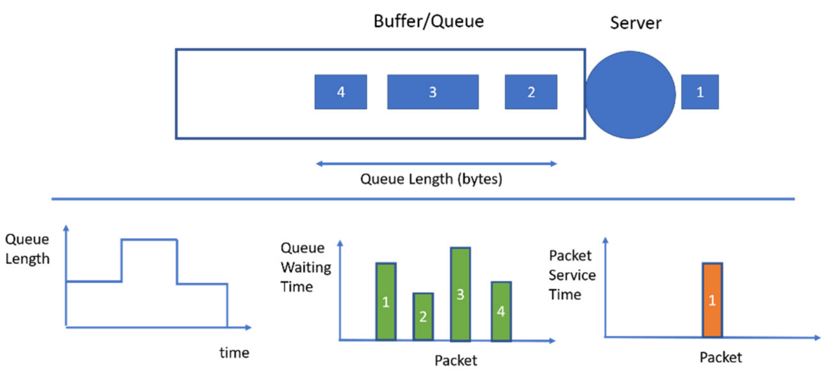


Figure 3 Buffer and queue description at transmitter.

To monitor buffer occupancy and queue dynamics, we introduce time-based sliding window structure, consisting of two adjacent windows, as shown in Figures 4-7. We divide each window into several time slots and monitor the previously mentioned metrics in each time slot.

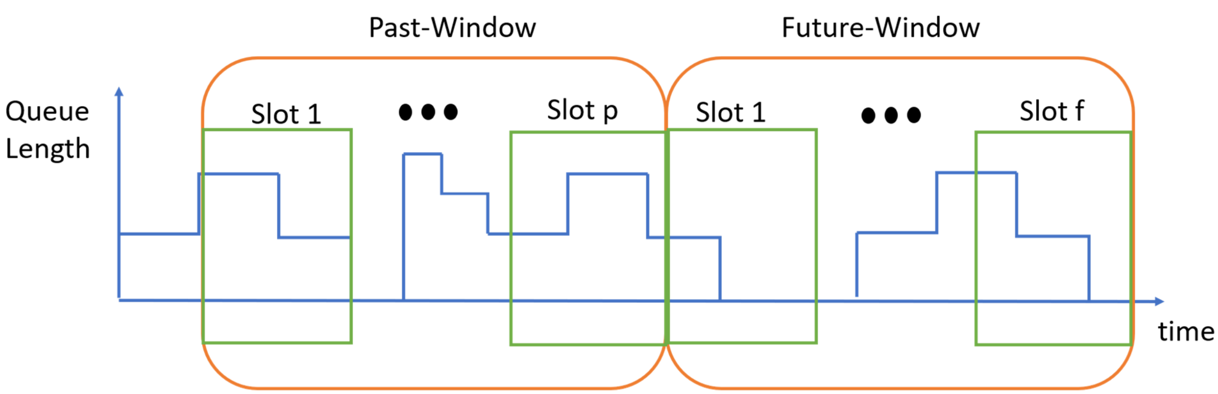


Figure 4 Sliding-window structure used to monitor queue length over time.

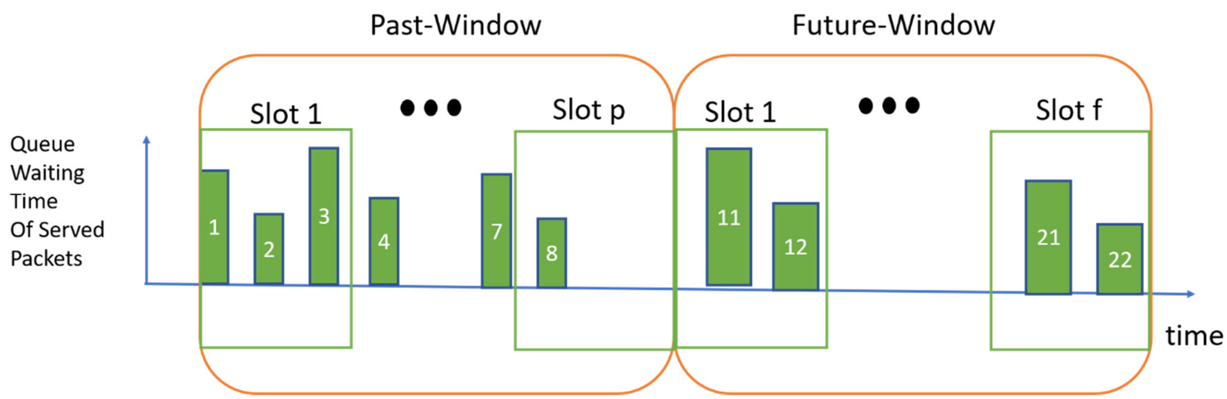


Figure 5 Sliding-window structure used to monitor queue waiting time for served packets over time.

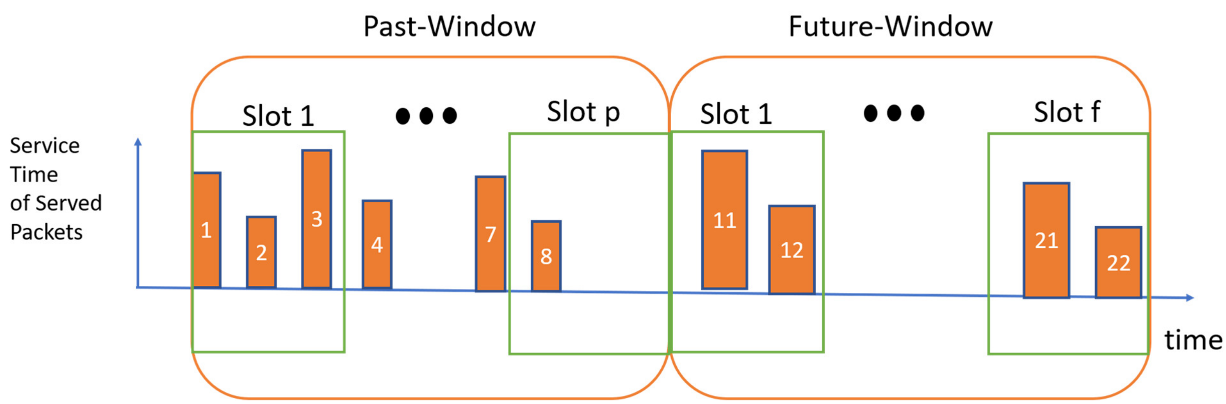


Figure 6 Sliding-window structure used to monitor packet service time for served packets over time.

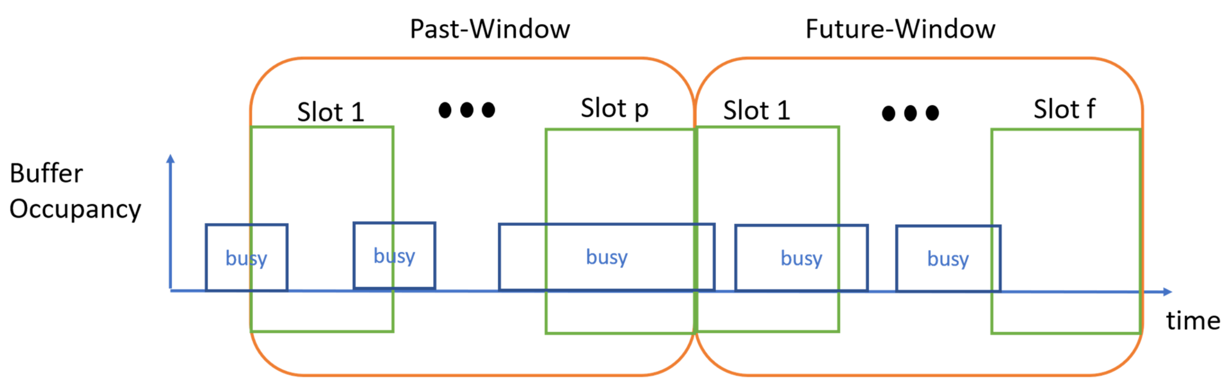


Figure 7 Sliding-window structure used to monitor buffer occupancy over time.

AI-Framework

We introduce a unified framework to monitor the traffic demand and service processes, as well as to control the different channel access parameters used by systems operating in unlicensed bands. Our framework consists of the following functional block, as shown in Figure 8.

Monitoring Traffic Demand Correlation and Stationarity Block

This block monitors the changes in traffic demand process by measuring the correlation and stationarity of the arriving traffic streams. This block computes a metric λ_d that describes the level of correlation and stationarity in the traffic demand process.

Monitoring Traffic Service Correlation and Stationarity Bock

This block monitors the changes in traffic service process by measuring the correlation and stationarity of queue dynamics and buffer occupancy over time. This block computes a metric λ_s that describes the level of correlation and stationarity in the traffic service process.

Experience Sharing between Monitors of Stationarity in Traffic Demand and Service Processes Block

This block translates the stationarity metrics between the monitors of stationarity for traffic demand and service processes. It provides a mapping between λ_d and λ_s.

Classifier Selector for Traffic Demand Classification Block

This block selects a proper ensemble of classifiers for classifying traffic demand process based on the value of stationarity metric λ_d. It also decides and adapts the structure of the sliding windows presented in Figure 2, including the length of windows and the number of their time slots. It also decides which metrics should be measured in each time slot by coordinating with the Dimensionality Reduction Block.

Classifier Selector for Traffic Service Classification Block

This block selects a proper ensemble of classifiers for classifying traffic service process based on the value of stationarity metric λ_s. It also decides and adapts the structure of the sliding windows presented in Figure 4-7, including the length of windows and the number of their time slots. It also decides which metrics should be measured in each time slot by coordinating with the Dimensionality Reduction Block.

Experience Sharing between Classifier Selectors Block

This block translates the choice made by Classifier Selector Blocks in the traffic demand and service processes. It helps capturing any correlation between the choice of classifier selection made for classifying the demand and service processes.

Classifying Traffic Demand Process Block

This block employs clustering algorithms for classifying traffic demands into several clusters, where patterns within the same cluster indicate a certain profile of traffic demand. Traffic demand patterns collected in the Future-Window in Figure 2 are used to build PDFs characterizing future traffic demands.

Classifying Traffic Service Process Block

This block employs clustering algorithm for classifying traffic services into several clusters, where patterns within the same cluster indicate a certain profile of traffic service process. This provides an indication on the condition of unlicensed channel and congestion level. Traffic service patterns collected in the Future-Window in Figures 4-7 are used to build PDFs characterizing future traffic services.

Experience Sharing between Traffic Demand and Service Classifiers Block

This block translates the experience of classification for the traffic demand and traffic service processes. It helps capturing any correlation between classification for traffic demand and service processes.

Predicting Future Traffic Demand Block

This block predicts future streaming statistic of traffic demand process based on the outcomes provided by the Traffic Demand Classifier Block. With help of traffic demand PDFs, it predicts the values of metrics characterizing future traffic demand. It also provides minimum bandwidth and latency requirements of future traffic demand.

Predicting Future Traffic Service Block

This block predicts future buffer occupancy and queue dynamics of the traffic service process based on the outcomes provided by the Traffic Service Classifier Block. With help of traffic service PDFs, it predicts the future behavior of traffic service process. This block helps the Controller Block make a better decision on whether any of the channel access parameters should be controlled to meet QoS requirements.

Experience Sharing between Traffic Demand and Service Predictors Block

This block translates the experience of prediction between the the traffic demand and traffic service predictors. It helps capturing any correlation between predictions of future traffic demand and service processes.

Experience Database Block

This block saves logs of previous predictions of traffic demand and service processes associated with their controls and resultant observations. This block helps the Controller Block making a better control decision in the future by consulting with previous control experience.

APs Coordinator Block

This block helps APs communicate their predictions and control for traffic demand and service processes.

Dimensionality Reduction Block

This block decides the number of metrics that should be used for classifying and predicting traffic demand and traffic service processes. It works jointly with Classifier Selector Block.

Controller Block

This block controls the different values of channel access parameters, including:

1. Contention window (CW) size
2. Initial Inter-Frame Space (IFS)
3. Duration of transmit opportunity (TXOP)
4. MAC packet data unit (MPDU) aggregation level
5. Negative ACK threshold for doubling CW value
6. Energy and preamble detection thresholds

This block works and interacts jointly with Predictor Blocks, Experience DB Block, and AP Coordination Block. This block produces a balance between the traffic demand and traffic service processes and ensure meeting QoS requirements.

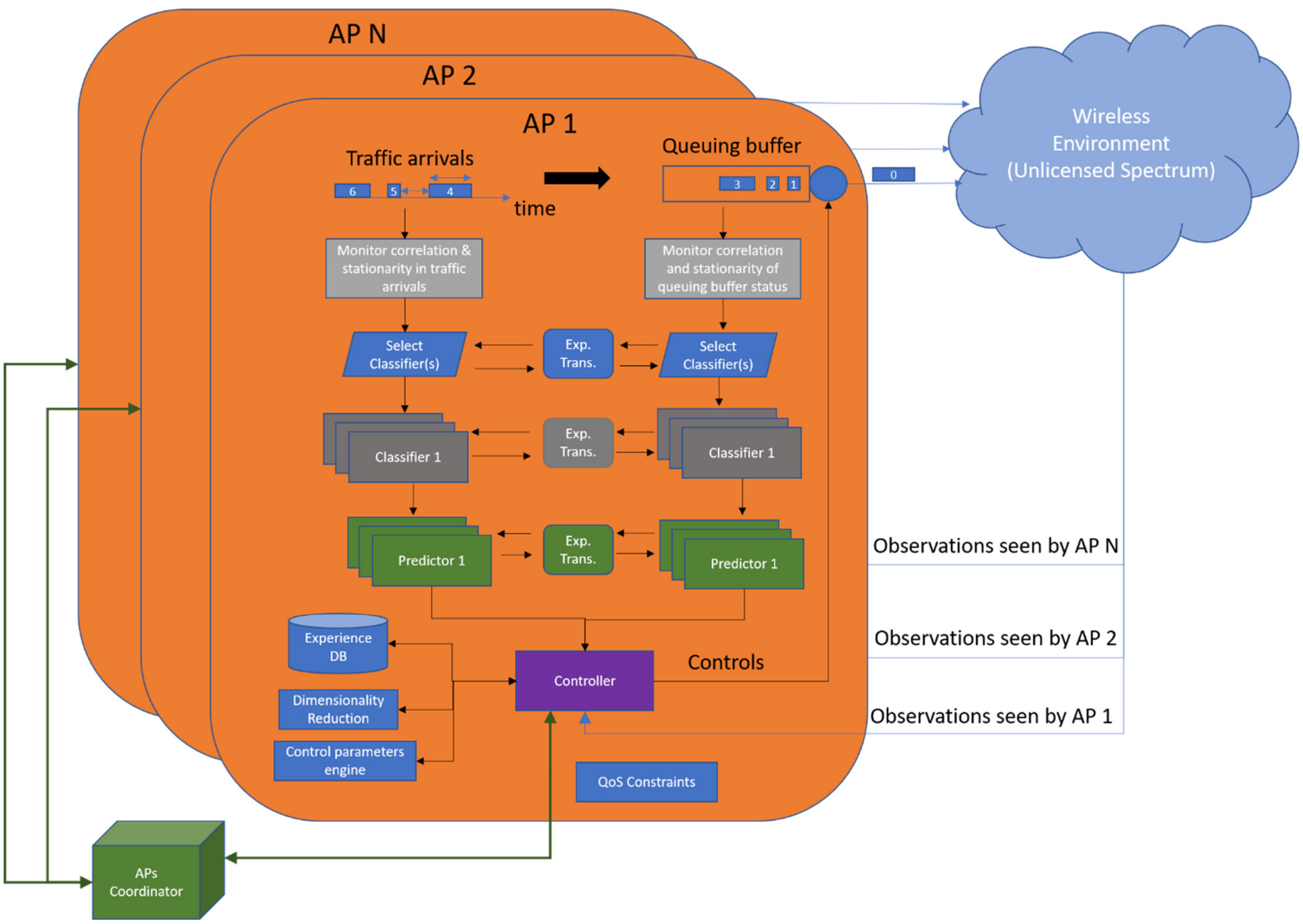
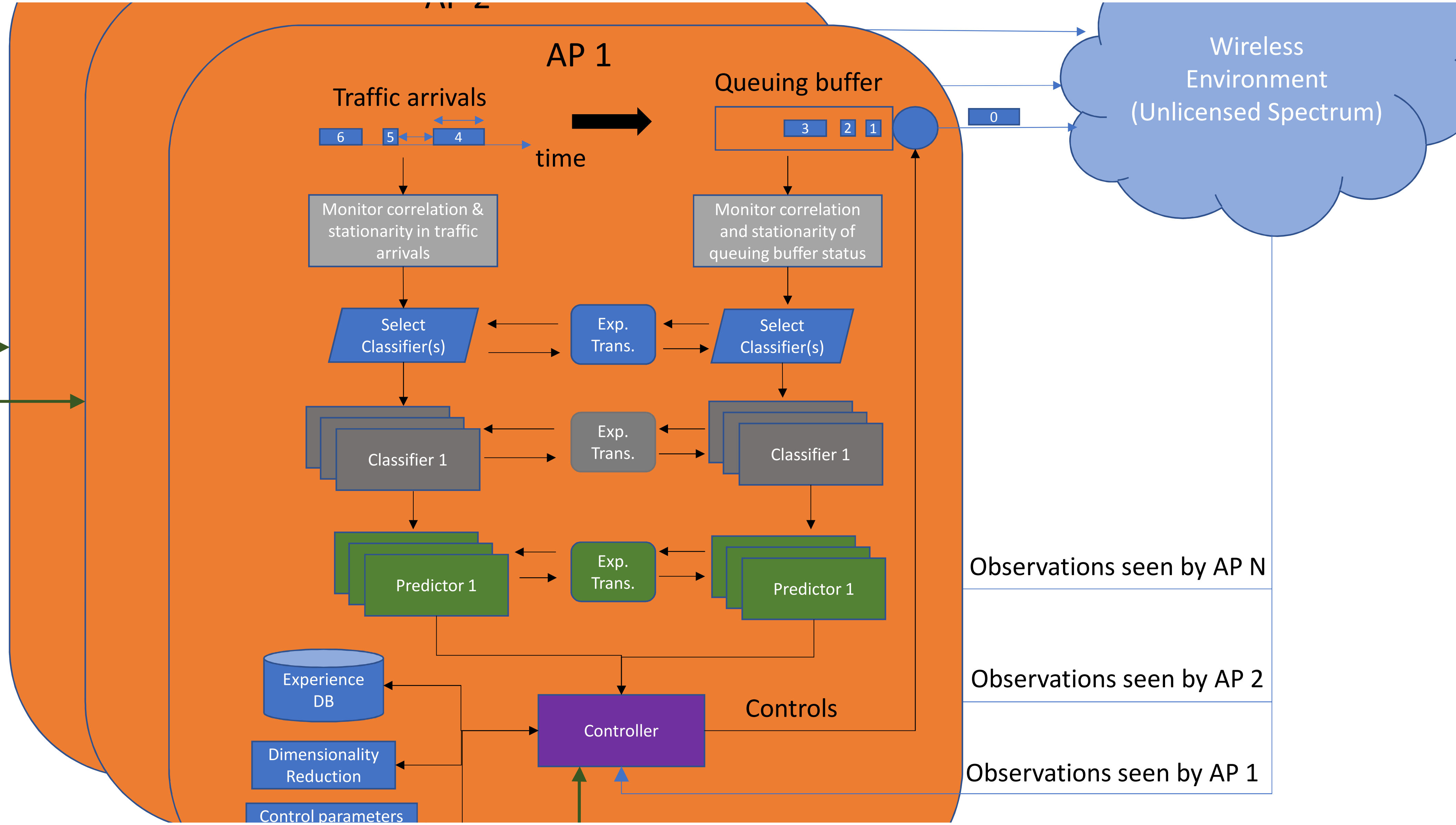


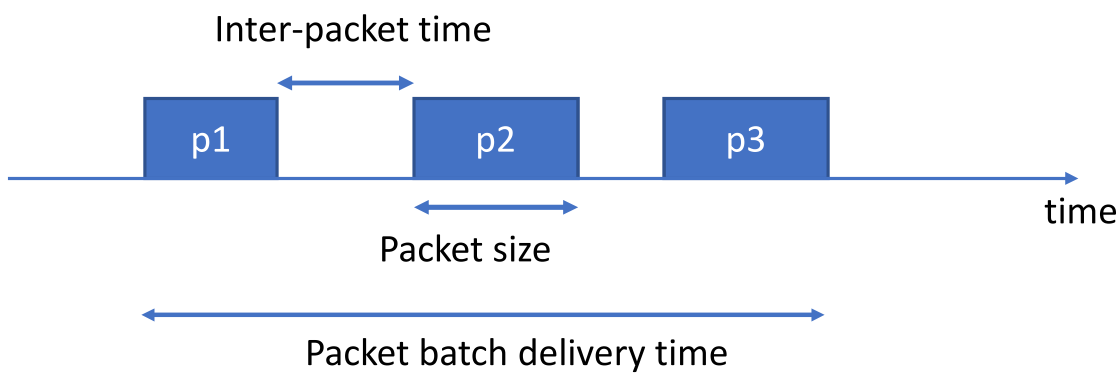
Figure 8 Artificial intelligence (AI)-enabled framework for monitoring and balancing traffic demands and service processes.

Unified AI-based Framework for Unlicensed Spectrum Operation

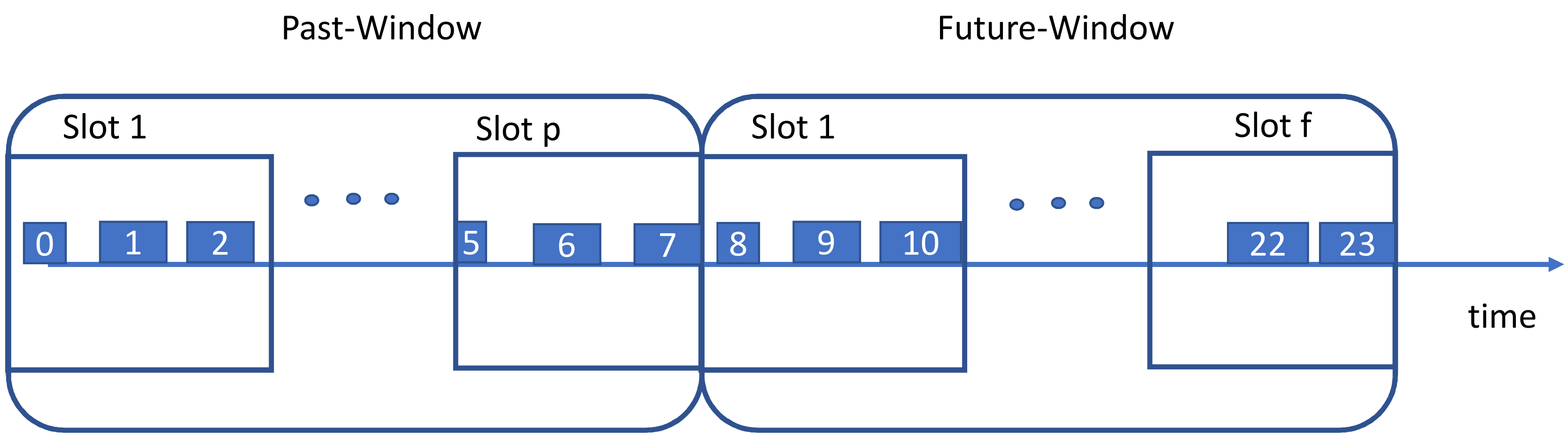
Mohammed Hirzallah, Balkan Kecicioglu, Belal Hamzeh, and Bernardo Huberman


AP 1
Traffic arrivals
6
5
4
time
Queuing buffer
3
2
1
0
Wireless Environment (Unlicensed Spectrum)
Monitor correlation & stationarity in traffic arrivals
Monitor correlation and stationarity of queuing buffer status
Select Classifier(s)
Exp. Trans.
Select Classifier(s)
Classifier 1
Exp. Trans.
Classifier 1
Predictor 1
Exp. Trans.
Predictor 1
Experience DB
Dimensionality Reduction
Control parameters
Controller
Controls
Observations seen by AP N
Observations seen by AP 2
Observations seen by AP 1

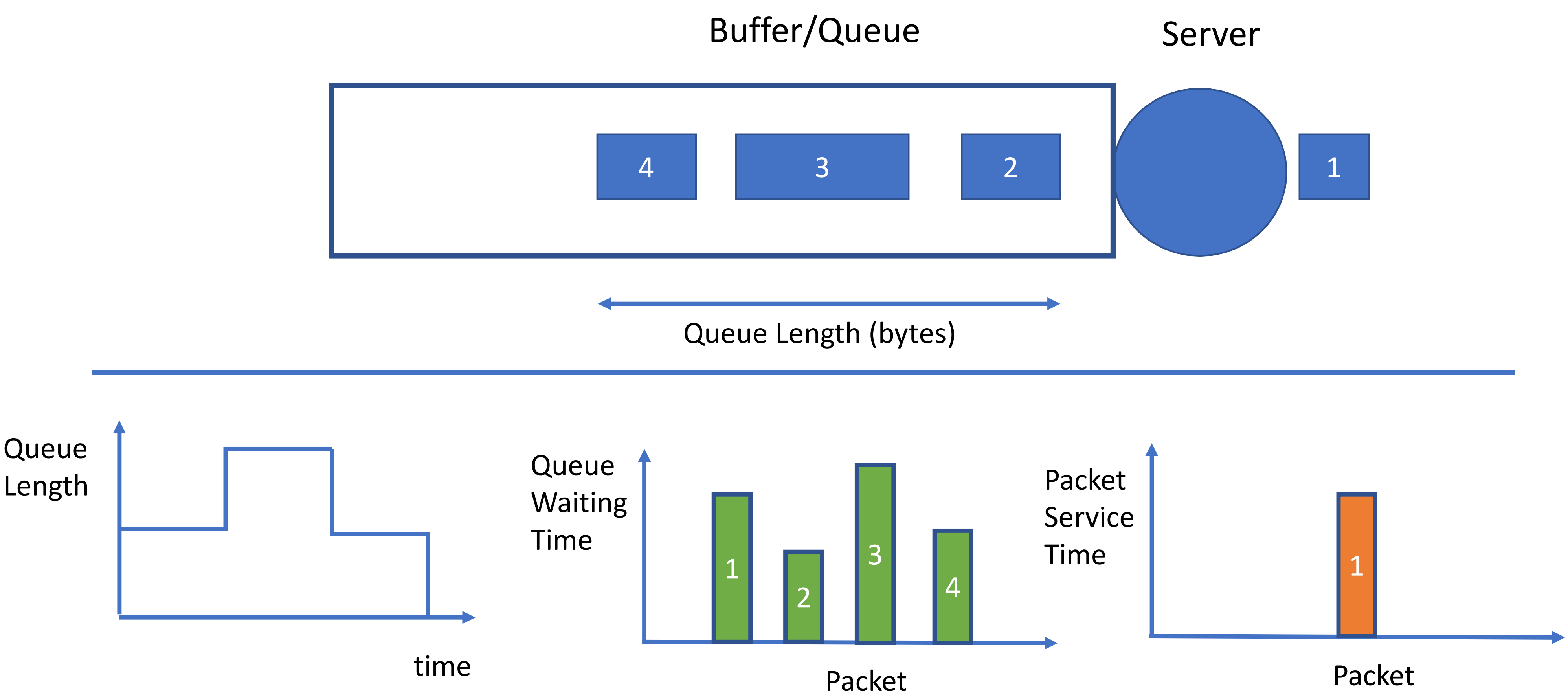
Traffic Arrivals Process



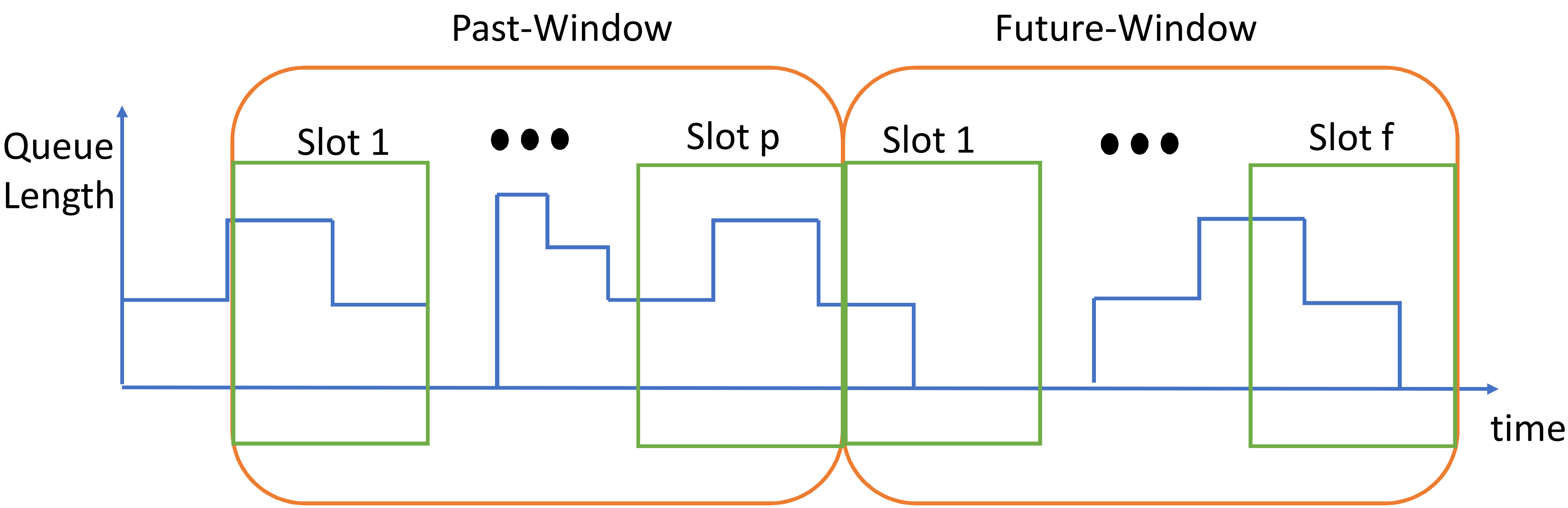
Monitoring of Traffic Demand

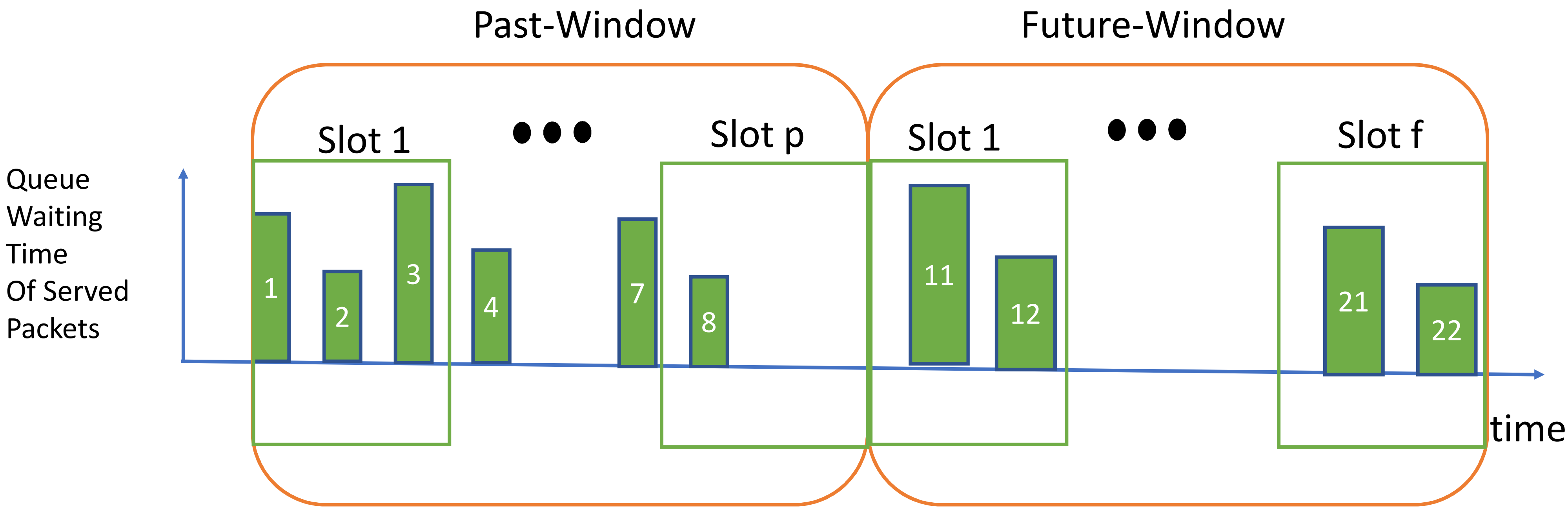


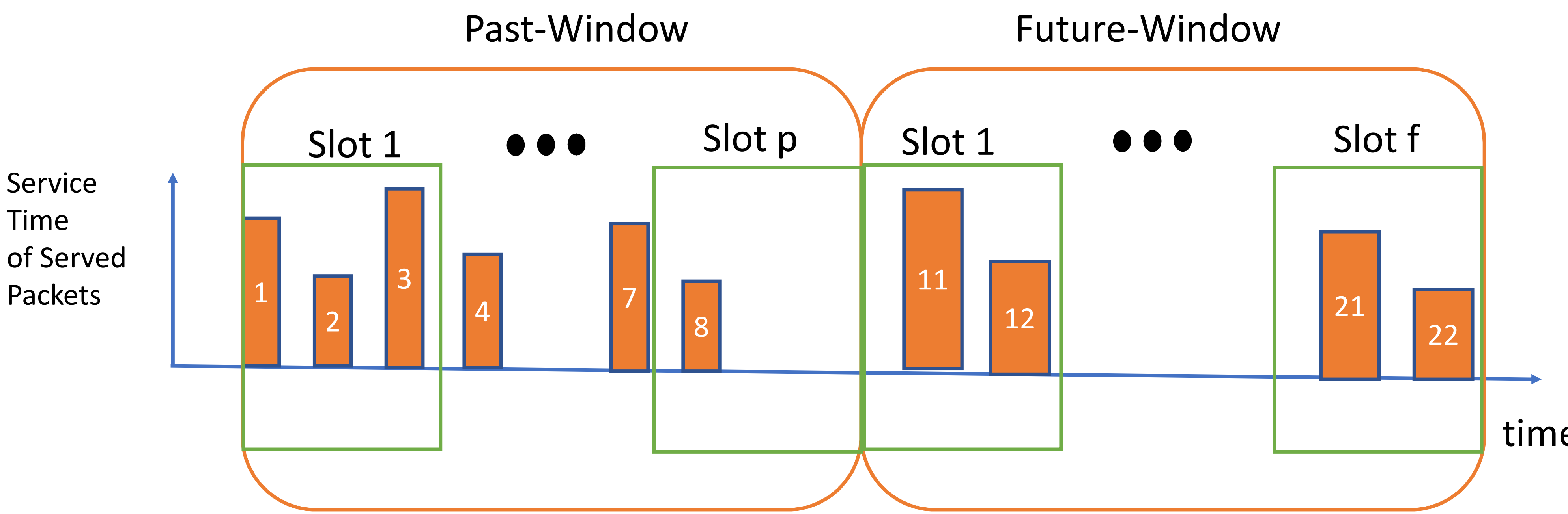
Queueing Buffer Process



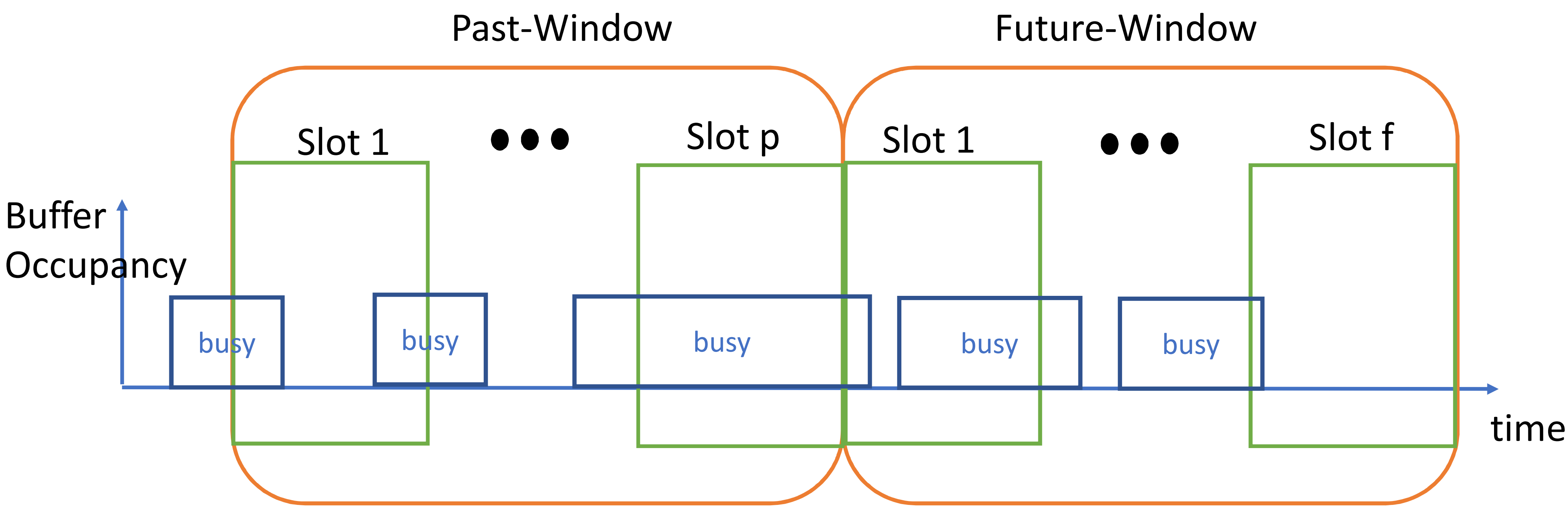
Classification & Prediction Structure




Past-Window
Future-Window
Slot 1
Slot p
Slot 1
Slot f
Queue Waiting Time Of Served Packets
1
2
3
4
7
8
11
12
21
22
time



Past-Window
Future-Window
Slot 1
Slot p
Slot 1
Slot f
Service
Time
of Served
Packets
1
2
3
4
7
8
11
12
21
22
time


Past-Window
Future-Window
Slot 1
Slot p
Slot 1
Slot f
Buffer
Occupancy
busy
busy
busy
busy
busy
time