

## Chapter 2

# Single-User Allocation

As discussed in chapter 1, the proper allocation of network resources is essential for providing the desired QoS. Both the network manager and the user are interested in obtaining the desired QoS as efficiently as possible. However, determining the amount of resources required for VBR sources is difficult, due to their bursty and unpredictable behavior, as seen in figure 1.1. Resource allocation becomes more difficult if reliable source information is not available before transmission begins.

While many different types of network resources exist, this thesis will focus on link bandwidth. Previous work in the area of single-user bandwidth allocation is presented and discussed in this chapter. Methods will be differentiated based on when information is collected, as well as what type of information is used for making allocation decisions.

### 2.1 Classifications and Previous Work

As seen in figure 2.1, there are many different categories of single-user bandwidth allocation. The categories are based on whether calculations are performed before or during transmission (off-line versus on-line). Off-line methods can be further divided into static (single allocation) and dynamic (varying allocation) methods. On-line methods are divided into groups depending on whether source measurements or QoS measurements are taken. Each of these categories have their own unique advantages and disadvantages. Their practicality is governed by the type and amount of a priori information required.

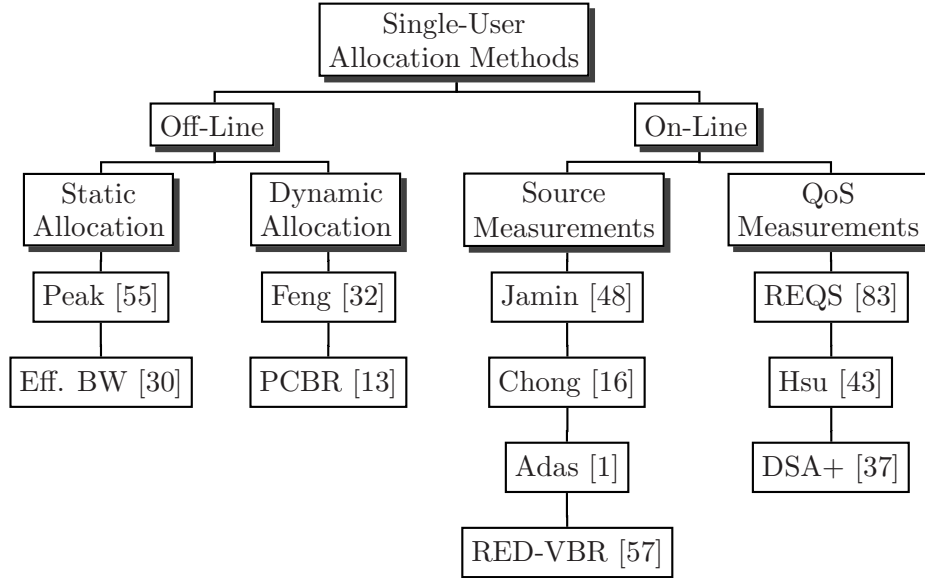


Figure 2.1: Categories of single-user bandwidth allocation.

### 2.1.1 Off-Line Allocation

Conventional approaches of bandwidth allocation rely on predetermined traffic characteristics. In the case of MPEG-compressed video, an off-line approach would require the entire bandwidth trace. From this information statistical values, such as the peak-to-mean ratio, can be determined. These methods are called “off-line” because calculating the appropriate bandwidth allocation occurs before transmission begins. These methods can be categorized as static or dynamic. Static methods allocate a single amount for the duration of the session, while dynamic methods adjust the allocated amount over time.

The earliest methods of resource allocation were based on the peak rate of the source [55]. Bandwidth statically allocated using this value provides strict deterministic guarantees. The drawback of this policy is the underutilization of resources. This is especially true for VBR sources which may have long periods of little traffic. For this reason, peak rate allocation is considered an upper bound for bandwidth allocation.

Effective bandwidth is another static resource allocation method. The effective bandwidth of a source can be defined as the minimum bandwidth required to satisfy a desired QoS. Formulas were originally developed to determine this value for a two-state

fluid flow model. Elwalid expanded the theory to include general Markovian models and defined the effective bandwidth to be the maximal real eigenvalue of a matrix derived from source characteristics and admission criterion [30]. Chou and Chang broadened the effective bandwidth theory to include video traces with limited success [17]. However it has been demonstrated that a VBR source may not adhere to these models [41, 44]. Yet more complex models, e.g. Hidden Markov Models, may be computationally intractable [12].

Dynamic off-line allocation approaches change the allocation amount over time [13, 32]. These methods are targeted for the transmission of stored video. Using the complete source trace, a bandwidth delivery plan can be created for the stored video that takes advantage of the client-side buffer. Using this buffer, large bursts can be prefetched before they occur, which results in a smooth video stream. For example, the off-line method developed by Feng, et al. determines the minimum number of renegotiations (increases or decreases in resource allocation) required for the playback of a previously-stored MPEG video [32]. The video is transmitted at the calculated rates to prevent buffer overflow and underflow. Off-line methods can provide the desired QoS with few renegotiations; however, these methods require the bandwidth trace to determine the optimal interval or segment length. For this reason, they are not viable if the trace is not available. On-line approaches attempt to manage such a case.

### **2.1.2 On-Line Allocation**

On-line schemes require minimal information about the source, since real-time observations are made to predict future bandwidth requirements. On-line methods also renegotiate allocation amounts over time based on these measurements. As seen in figure 2.1, off-line approaches can be categorized based on the information measured. Some techniques measure the source characteristics in order to predict future demands. It is important to note these methods do not attempt to provide any particular QoS. Alternatively, on-line methods that measure current QoS adjust allocation amounts to maintain a targeted QoS (no more or less).

## Source Measurements

In [1], Adas used adaptive and non-adaptive least mean square error predictors to predict MPEG frame sizes. The MPEG source was divided into I B and P frames [63] then separate filters were used to predict future MPEG frames. Bandwidth was allocated based on this information. Similarly, Chong et al. examined on-line bandwidth allocation of MPEG-compressed video based on the previous frames observed [16]. Neural networks as well as filters were used to predict future frame sizes. A drawback of this method is the complexity associated with neural networks, and the significant amount of information required to make accurate predictions. Other similar renegotiation schemes which predict future allocation demands include [28, 91].

RED-VBR is a method for supporting VBR video, with an off-line or on-line allocation technique. RED-VBR is based upon the D-BIND model, that attempts to capture the property that a source may have different rates at different time intervals [57]. The allocation algorithm stores the currently reserved D-BIND parameters and calculates the D-BIND parameters for the last M frames. A renegotiation take place when a difference exists between the reserved and measured D-BIND parameters. RED-VBR does not use nor measure the QoS for allocation. QoS is an issue when sources are multiplexed together and is provided on a “per-segment” basis as described in [104].

Another renegotiation algorithm specifically for MPEG-compressed video was introduced by Reininger et al. [86]. This method renegotiated the usage parameter control (UPC) values based on the targeted and observed quantization of MPEG frames. When a new set of UPC parameters are determined, the user adapts (scales) transmission of the video until the new parameters are granted. Adapting the transmission of the video reduces the cell loss during the renegotiation. Saito, et al. proposed a bandwidth renegotiation scheme for ATM virtual paths [92]. Bandwidth was altered in fixed increments depending on the availability and the state of other virtual paths.

As previously stated, these on-line techniques only predict future source demands. There is no attempt to determine the appropriate bandwidth allocation that results in a desired QoS. For this reason, these methods may over allocate bandwidth for a desired QoS.

In addition, none of these methods attempt to reduce the number of renegotiations required for transmission of a given source.

### QoS Measurements

On-line methods that measure QoS attempt to allocate bandwidth to maintain a desired QoS (no more or less). Hsu and Walrand proposed a method that dynamically allocates the bandwidth of a connection in order to provide a desired cell loss probability [43]. Their technique consists of a dynamic (adaptive) algorithm that renegotiates bandwidth over fixed length intervals. The renegotiation amount is determined using a simple difference error function and previous loss measurements. The algorithm is proven theoretically to converge to a steady state rate using techniques developed in [4]. This proof is valid only for Markov-modulated fluid sources and no investigation into actual VBR traffic was performed. Other dynamic or feedback approaches include [29, 102]. Although these on-line renegotiation methods provide better resource utilization and can manage real-time traffic, they rely on multiple allocation changes. In most cases allocation changes are done at fixed intervals, which may be problematic over long spans of time or during bursty periods. It is desirable to reduce the number of allocation changes, since contention for more resources may occur.

REQS (Resource Efficient Quality of Service) is another on-line resource allocation method that measures actual QoS [83]. REQS attempts to converge to a steady state rate based on QoS observations over varying time intervals. As the algorithm converges, the time intervals are increased. As a result the number of allocation changes required is dramatically reduced. This method was shown to be robust with small convergence times under various conditions using a Markov Modulated Bernoulli Process (MMBP). However no proof of convergence was attempted and the effects of actual traffic, specially those having long range dependencies, were not investigated.

## 2.2 Chapter Summary

As mentioned in the introduction, the proper allocation of link bandwidth is essential for providing QoS in computer networks. Users and network managers are interested in providing QoS as efficiently as possible. Efficiency represents the amount of link bandwidth allocated as well as the number of renegotiations required.

There are many different methods of bandwidth allocation. These methods differ on the amount of a priori information and the number of renegotiations required. Off-line methods compute the bandwidth allocation required for a certain QoS, but require accurate source information. These methods may determine a single bandwidth amount for the entire session (effective bandwidth [30]) or may renegotiate over time (PCBR [13] and Feng's algorithm [32]). While these methods perform well for stored media, they are unable to efficiently allocate bandwidth for live or interactive media.

In contrast, on-line methods allocate bandwidth using little a priori information. Source or QoS measurements are used to allocate bandwidth, which may change over time. Using source measurements, on-line methods attempt to predict future demands. Yet, these methods do not determine the appropriate allocation for a desired QoS. Alternatively, on-line methods that measure actual QoS seek to allocate bandwidth to provide a desired QoS. These methods try to reduce the amount of resources required; however, few attempt to reduce the number of renegotiations. In the next chapter, a new on-line bandwidth allocation method based on measured QoS is introduced. This method will actively seek to reduce the bandwidth allocated as well as reduce the number of renegotiations required. This method will also successfully allocate link bandwidth for various VBR sources, which is difficult due to their unpredictable behavior.