OVERBOOKING IN MOBILE BACKHAUL

Chunming Liu, Bryan Fleming

T-Mobile, Bellevue, USA

{Chunming.Liu, Bryan.Fleming}@t-mobile.com

ABSTRACT

The rapid increase of mobile data subscribers and deployment of broadband radio access networks push mobile operators to upgrade their mobile backhaul from expensive legacy TDM transport to carrier Ethernet for cheaper operational expenditures. The Metro Ethernet Forum has defined a set of Ethernet Virtual Connection services that are adopted to provide scalable Ethernet transport for mobile backhaul. However, these services usually address single cell site backhaul per UNI handoff, neither considering statistical multiplexing gain at a hub site which aggregates backhaul traffics of multiple cell sites, nor supporting Quality-of-Service provisioning in overbooked backhaul pipe. This paper proposed an efficient carrier Ethernet overbooking solution for mobile backhaul. A statistical estimation method has been developed for deriving a safe overbooking factor at a given UNI. Then two efficient transport architectures were proposed to support bandwidth sharing in cellular cluster with overbooked backhaul bandwidth in carrier Ethernet. A novel bandwidth control algorithm has been derived at customer edge device to provide Quality-of-Service for multimedia traffics over overbooked UNI, with SLA policing and protection on high priority services. Experimental data analysis and simulations have showed that our new schemes can benefit mobile operators in resource utilization efficiency, carrier Ethernet cost saving and backhaul performance.

KEYWORDS

Mobile backhaul, overbooking, Carrier Ethernet, UNI handoff, CIR, SLA, Quality of Service

1. INTRODUCTION

In recent years, carrier Ethernet becomes a more popular scheme for the mobile backhaul, because it is packet based transport network and allow flexible Ethernet circuit provisioning with bandwidth scalability and low operational expenditures (OPEX) for mobile operators [1]. The Metro Ethernet Forum (MEF) [9] has defined Ethernet virtual Connection (EVC) services with performance requirements, as a transport association between two or more user-network-interfaces (UNI) [2][9]. There are three types of EVC service including E-Line, E-LAN and E-Tree services, which represent a point-to-point (P2P), multipoint-to-multipoint (MP2MP), and rooted multipoint-to-multipoint EVC service, respectively. Due to simplicity and reliability, P2P EVC services are adopted by most mobile operators to connect their cellular sites to mobile core networks. A carrier Ethernet vendor is responsible to provide such a P2P EVC connectivity following the bandwidth profile and performance requirements defined in a Service Level Agreement (SLA) [2][3] between the carrier vendor and the mobile operator.

Wireless radios become popular to provide mobile backhaul connectivity for cellular sites in cases where Ethernet cable or fiber is not available, or cannot be deployed in a timely or economic manner. Compared to single cell site backhaul that is directly connected with carrier Ethernet cable or fiber, mobile operators normally develop their own wireless radio links, such as microwave [4], to connect multiple cell sites nearby, and deliver aggregated backhaul traffic to a

DOI : 10.5121/ijwmn.2013.5411

hub site which handoff to carrier Ethernet network at UNI. In order to support different backhaul bandwidth requirements for each cell site in cluster, carrier vendors normally assign each mobile cell site an individual EVC circuit with a distinct VLAN ID and Committed Information Rate (CIR), and then implement parallel multiple EVC handoff over a same UNI at the hub site, as shown in Figure 1.

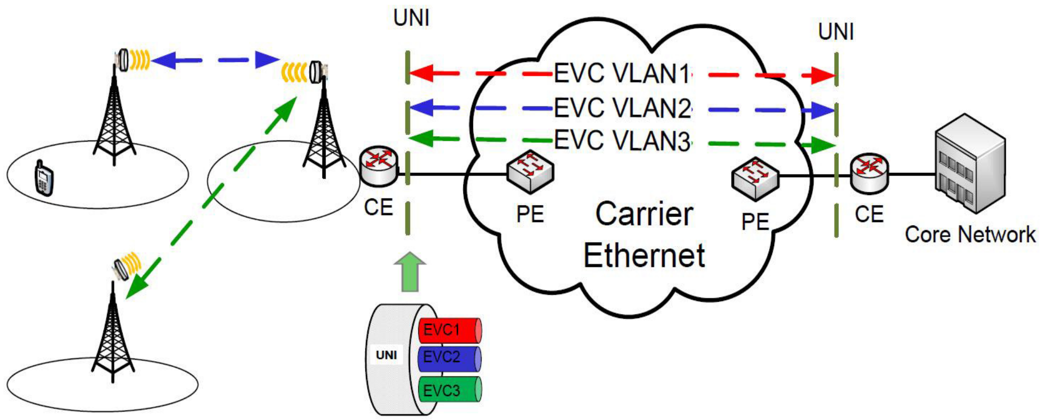


Figure 1. Microwave backhaul with multiple EVC bundling UNI handoffs

However, since all EVC circuits are assigned with fixed CIR values, they cannot take advantages of statistical multiplexing gain at UNI handoff and have no way sharing idle bandwidth among EVCs [1]. Therefore, the overall resource utilization of an UNI is low and mobile operators overpay carrier vendors for unused bandwidth. Authors in [5, 6] proposed delay based methods to decide the minimum transport capacity for which all traffic delay requirements are met. However, multimedia services have different performance requirements and optimal delay performance may result in low bandwidth utilization too [7]. A piecewise linear approach was proposed in [7] to achieve balance between network performance and bandwidth saving. Authors proposed a selective overbooking scheme based on trunk size and usage profile. However, due to different service types and traffic distributions in cell sites, peak bandwidth utilizations at hub site may differ even UNI bandwidth or cluster sizes are same, which means overbooking factors at different hub UNIs may differ from each other, instead of linearly following CIR. A capacity planning scheme for LTE backhaul networks has been proposed in [8] where an overbooking factor calculation method based on traffic forecast, multiplexing gain and peak throughput. However, the overbooking factor is derived from estimation on 50 user peak throughput model, without considering real cellular traffic distributions. Authors neither addressed how to implement carrier Ethernet overbooking scheme, nor investigated bandwidth allocation with QoS for multimedia services when an overbooked UNI becomes congested.

The contributions of this paper are follows: First, we developed a statistical regression method for overbooking factor estimation, based on UNI peak utilization data, cellular traffic distributions, and statistical service outage performance. Then we designed two efficient carrier Ethernet architectures to help bypass CIR binding in EVC and implement mobile backhaul bandwidth sharing among a cellular cluster. Specifically, a mobile operator can adopt either an Ethernet switching or routing scheme to complete backhaul transport in its cellular cluster, while using a single VLAN EVC pipe in carrier Ethernet for bandwidth overbooking. Then to overcome network congestion caused by aggregated burst traffics at UNI, a new bandwidth control algorithm was proposed to provide Quality of Service and fairness for resource allocation. Solid experimental networking data and simulation results showed that the proposed schemes can benefit mobile operators in resource utilization efficiency, backhaul cost saving and quality of service provisioning.

This paper is organized as follows: in section 3, we present the UNI capacity estimation and overbooking derivation methods. Then we describe two efficient Carrier Ethernet architectures for implementing UNI overbooking with carrier Ethernet. In section 4, we present a new bandwidth control scheme at an overbooked UNI to support QoS. Network performance data and simulation results are shown in section 5 with analysis. Finally, we conclude this paper with a brief summary.

2. ACRONYMS

CBS	Constant Burst Size
CE	Customer Edge device
CIR	Committed Information Rate
CoS	Class of Service
EVC	Ethernet Virtual Connection
MBS	Maximum Burst Size
MEF	Metro Ethernet Forum
PE	Provider Edge device
PIR	Peak Information Rate
SLA	Service Level Agreement
UNI	User-Network-Interface

3. UNI CAPACITY AND OVERBOOKING DERIVATION

Consider a microwave cluster containing total K cell sites and traffic intensity of a cell site i, X_i, the aggregated traffic intensity at UNI, Y_{uni}, is represented as,

$$Y_{uni} = A_{uni} \cdot \sum_{i=1}^{K} X_i \tag{1}$$

where A_{uni} is the statistical multiplexing gain value at the UNI.

Assume backhaul traffic intensities of cell sites are independent, and there are two major traffic patterns: the Poisson based model and self-similar model [7]. The Poisson-based traffic model has been intensively used to represent cellular voice connections, while the self-similar model is used to represent data services with burst throughputs. We further assume there are M types of voice services, and N types of self-similar services. We adopt an ON-OFF source model to analyze the peak throughput of a voice connection, where the ON and OFF states represent the active and silent conditions of the voice connection, respectively. Both ON and OFF state intervals are assumed to be exponentially distributed, and R_j is a constant packet generation rate of voice class j in the ON state. Due to packet burst characteristics and CIR throttle on backhaul capacity, the throughput ξ_j of self-similar service class j follows truncated Pareto distribution with following probability distribution function [5],

$$f(\xi_j) = \frac{\alpha_j E_j^{\alpha_j} \xi_j^{-\alpha_j - 1}}{1 - (E_j / H_j)^{\alpha_j}} \tag{2}$$

where α_j denotes shape parameter, E_j denotes the minimal traffic rate, and H_j denotes the maximum traffic rate of service class j. Then the aggregated throughput ξ_{uni} at the UNI is denoted as,

$$\xi_{uni} = \sum_{i=1}^{K}\sum_{j=1}^{M}\sum_{k=0}^{l_{i,j}}\xi_p(i,j,k) + \sum_{i=1}^{K}\sum_{j=M+1}^{M+N}\sum_{k=0}^{l_{i,j}}\xi_s(i,j,k)$$

$$\leqslant \sum_{i=1}^{K}\sum_{j=1}^{M} l_{i,j} R(j) + \sum_{i=1}^{K}\sum_{j=M+1}^{M+N}\sum_{k=0}^{l_{i,j}}\xi_s(i,j,k) \quad (3)$$

where $\xi_p(i,j,k)$ represents the throughput of connection k of Poisson-based service type *j* in cell *i*, $\xi_s(i,j,k)$ represents throughput of connection k of self-similar service type *j* in cell *i*, and $l_{i,j}$ represents total connection number of class *j* in cell *i*.

From the equation (3), the peak aggregation throughput at UNI is determined by the sum of self-similar traffics following truncated Pareto distribution function, which can be approximated as a Gaussian distribution [5][7], with mean value μ_{peak}, and standard deviation σ_{peak}, which can be derived from peak throughputs at the UNI. When an overbooking ratio O_{uni} is applied at UNI bandwidth which equal to $\sum_{i=1}^{N} CIR(i)$, it is expected to achieve low bandwidth outage probability, i.e., the probability of the case that overbooked bandwidth cannot transport aggregated peak throughput is smaller than or equal to a threshold service outage ratio ε, $0 < \varepsilon < 1$. Then we can get,

$$P(\xi_{uni} \le O_{uni} \cdot \sum_{i=1}^{N} CIR(i)) \ge \varepsilon \quad (4)$$

When equality holds in (4), the overbooking ratio O_{uni} is minimized, and the overbooked UNI bandwidth can be minimized. Since ξ_{uni} follows normal distribution, then the following relationship is satisfied,

$$O_{uni} = \frac{\sigma_{peak} \cdot Q^{-1}(\varepsilon) + \mu_{peak}}{\sum_{i=1}^{N} CIR(i)} \quad (5)$$

where $Q^{-1}(x)$ is the inverse function of normal distribution *Q(x)*, and $Q(x) = \frac{1}{2\pi}\int_x^{\infty} e^{-\frac{s^2}{2}} ds$.

Applying O_{uni} on existing UNI bandwidth, we get the overbooked UNI bandwidth for the whole cluster as,

$$B_o = O_{uni} \cdot \sum_{i=1}^{N} CIR(i) \quad (6)$$

Because full utilization results in high latency and congestion [8], the real UNI bandwidth after overbooking is set to be larger than overbooked bandwidth derived in equation (6) by dividing a maximum utilization ratio η. Therefore, the feasible UNI bandwidth B_{uni} is set as,

$$B_{uni} = B_o / \eta \quad (7)$$

where $0.5 \leqslant \eta < 1$ and it is normally defined by mobile operators based on backhaul performance requirements [7].

4. EFFICIENT UNI HANDOFF ARCHITECTURE

Since current MEF E-line service does not support overbooking at UNI with parallel EVC circuits, we need to improve the carrier Ethernet transport architecture to implement bandwidth sharing inside the cellular cluster. Considering a CE device at a UNI is either a layer 2 switch or a cell site router (CSR) which supports layer 3 networking functionalities, we proposed two UNI handoff architectures to implement the above goal.

4.1. Q-in-Q Tunneling Architecture

In scenarios that the CE at UNI only supports Layer 2 VLAN switching, a mobile operator can order a single EVC pipe with its contracted carrier vendor for mobile backhaul of a target cell cluster. The EVC pipe is implemented with IEEE 802.1ad stacked VLAN bridging technology (Q-in-Q tunneling) [2], in which the Ethernet carrier configures and delivers P2P mobile backhaul traffic between two PEs based on outer VLAN service tag (S-tag) in carrier Ethernet frames. After the carrier PE handoffs traffic to associated CE in mobile network, the CE implement switching based on inner customer VLAN tag (c-tag) for frame delivery to the destination recipient cell, as shown in Figure 2.

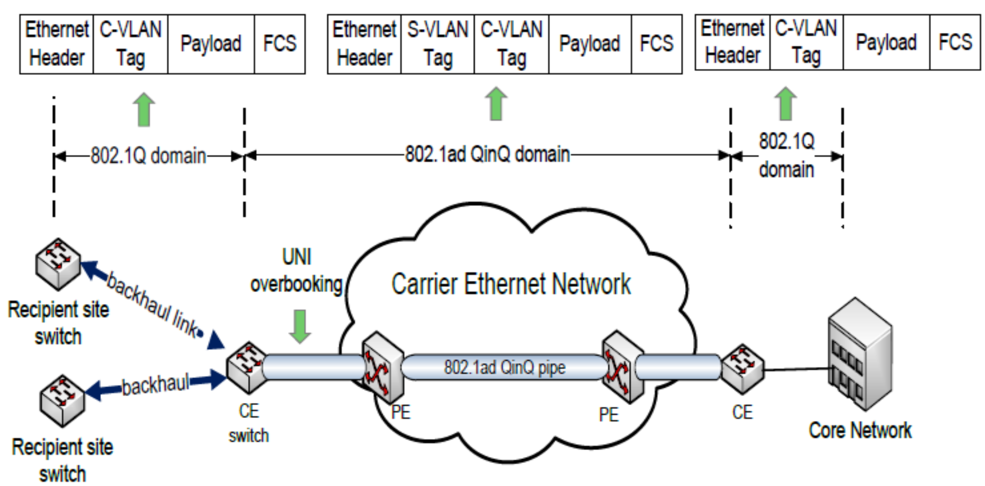


Figure 2. Q-in-Q Tunneling Carrier Backhaul Architecture

The EVC capacity is determined by the overbooked UNI bandwidth defined in equation (7). The mobile operator predefines and encapsulates a VLAN ID associated with one of their destination cell sites with inner customer-tag (C-tag) without necessarily notifying the carrier vendor CIR value of each cell site, and only contract with the vendor with the CIR value of ordered EVC circuit, which tunnel backhaul traffic between MSO and the hub site. The tunneling backhaul scheme goes as follows: any backhaul Ethernet frame generated in MSO or in a cellular cluster is attached with a C-tag VLAN ID which works as an ID marker of the backhaul of a destination cell site. When aggregated backhaul frames arrive at CE devices at UNI, the CE attaches S-tag VLAN ID in each frame as an indicator of the pre-specified carrier Ethernet EVC circuit, and handoffs the Q-in-Q frame to carrier Ethernet network. When the PE device in the carrier network receives the frame, it checks S-tag VLAN ID in the frame and finds out the associated carrier EVC circuit for delivery. Once a CE on the other side of the EVC circuit receives Q-in-Q frame from its associated PE, it detaches the outer S-tag, checks the inner VLAN ID in C-tag for destination site, and delivers the frame to next hop that follows same layer 2 switching scheme based on VLAN ID until it reaches the destination cell site.

4.2. Cross-layer Backhaul Architecture

In scenarios that the CE at UNI is a cell site router (CSR) which supports IP layer functionalities, we adopt a cross-layer solution for cellular cluster backhaul, i.e., a carrier vendor configures a single EVC pipe from PE associated with the MSO to the edge PE associated with a CE at the hub site with overbooked bandwidth. After a UNI handoff, the router at hub site will deliver the payload packet to the destination site by checking IP address in frame payload. Then in the cellular cluster, the backhaul packets are delivered through routing, which implements an end-to-end cross-layer backhaul scheme, as shown in Figure 3.

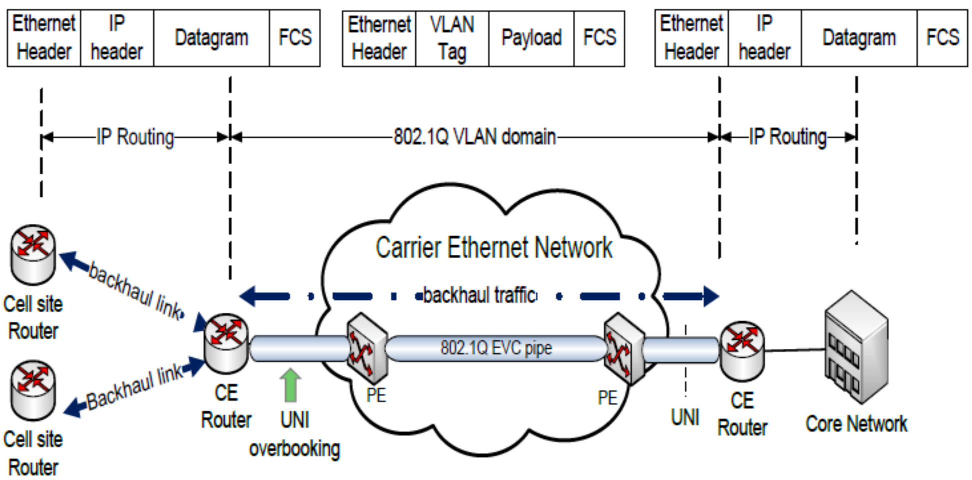


Figure 3. Cross-layer backhaul architecture

The traffic routing in cellular cluster can be implemented by either static routing or dynamic routing. If the cluster size is small, static routing is preferred and the routing table is configured manually in the CSR at hub site by specifying next-hop interface IP address towards the destination cell site. However, when the cluster size increase, the static routing is neither fault tolerant nor efficient for networking configuration and management. Dynamic routing can fix these problems by supporting real-time routing path construction and selection. It is required that the recipient sites have their own routers so that they can communicate with the CSR at hub site with common dynamic routing protocols, such as Open Shortest Path First (OSPF) or Routing Information Protocol (RIP) [1][2], for creation, maintenance, and updating routing paths in their routing table.

4.3. Advantages of Proposed Architectures

Advantages of our proposed architectures: the UNI is implemented as a backhaul aggregation pipe, instead of a group of parallel EVC circuits. And the UNI bandwidth profile is determined by the multiplication product value of the derived overbooking ratio and the sum of CIR values of all cell sites in the cluster. Therefore, the statistical multiplexing gain is used for bandwidth sharing among the cell site cluster, and no more dependency on carrier EVC provisioning. The mobile operator can adjust bandwidth profile more easily based on cluster size and the peak utilization at the UNI. The scalability of mobile backhaul also get improved because it is not necessary for the mobile operator to request additional EVC circuits for newly added cell sites in the cluster.

Furthermore, carrier vendors reduce the complexity in networking configurations and EVC maintenance, through a single EVC circuit provisioning for whole cellular cluster.

5. Bandwidth Control at Overbooked UNI

Due to burst throughput and overbooking at UNI bandwidth, a traffic policing and rate shaping scheme are required for providing stable QoS level even when UNI becomes congested [12,13], as well as maximizing bandwidth utilization [9, 10]. MEF has defined a two-rate, three-color marker (trTCM) algorithm for CE VLAN CoS which can be implemented via two token buckets scheduling [3]. One bucket is used to determine in-profile service frame rate per CoS, following CIR value, while the other bucket is to determine excess service frame rate per CoS, following EIR value. However, in the MEF scheme, each bucket size is a fixed value and egress point does not differentiate service priorities for multimedia traffics. Due to cluster traffic aggregation and UNI bandwidth overbooking, the bandwidth scheduling and rate enforcement need to be re-calculated in real time. Based on service performance requirements, traffics generated in mobile network are classified into three types which are marked with P bits in the Priority Code Point (PCP) field of IEEE 802.1Q tag in Ethernet frames. When aggregated traffics arrive, The CE at UNI checks p bits in frames and puts traffic of each class into an individual forwarding class (FC) queue which implement egress policy with specified CIR, Peak Information Rate (PIR), Constant Burst Size (CBS) and Maximum Burst Size (MBS). Policing rule of each FC queue is set as follows: the specified CIR rate is enforced when egress to the PE device in carrier network, and service traffics above CIR are marked as out-of-profile. The PIR is the maximum rate at which frames are allowed to burst over the CIR and delivered as best efforts. Traffic shaping is implemented through absorbing the traffic burst up to MBS and dropping traffic over PIR after MBS. The CBS in kilobyte is generally determined by the product of maximal tolerable latency of the service specified in the SLA and the assigned CIR rate, as follows:

$$CBS_i = \frac{CIR_i \cdot L_i}{1000 \times 8} \tag{8}$$

where L_i denotes the required latency value of service i, and $i \in \{voice, video, data\}$. Similarly, the MBS in kilobyte is denoted as the product of maximal tolerable latency of the service and the assigned PIR rate,

$$MBS_i = \frac{PIR_i \cdot L_i}{1000 \times 8} \tag{9}$$

The scheduling priorities: the voice service is assigned with highest priority for rate allocation, due to its most strict requirements on throughput, latency, jitter and FLR. The next priority is given to video stream service, due to its high requirements on throughput, latency and FLR. The data service priority is lowest due to best-effort traffic characteristics. However, the data service is sensitive to the minimum throughput to avoid starvation and network management blocking. To implement dynamic rate allocation based on aggregated traffic intensities at UNI, we further divide time into a sequence of slots, in which each slot contains same time length $t_{slot_}$second(s), and the scheduler assign CIR and PIR rates for each service at the beginning of the slot.

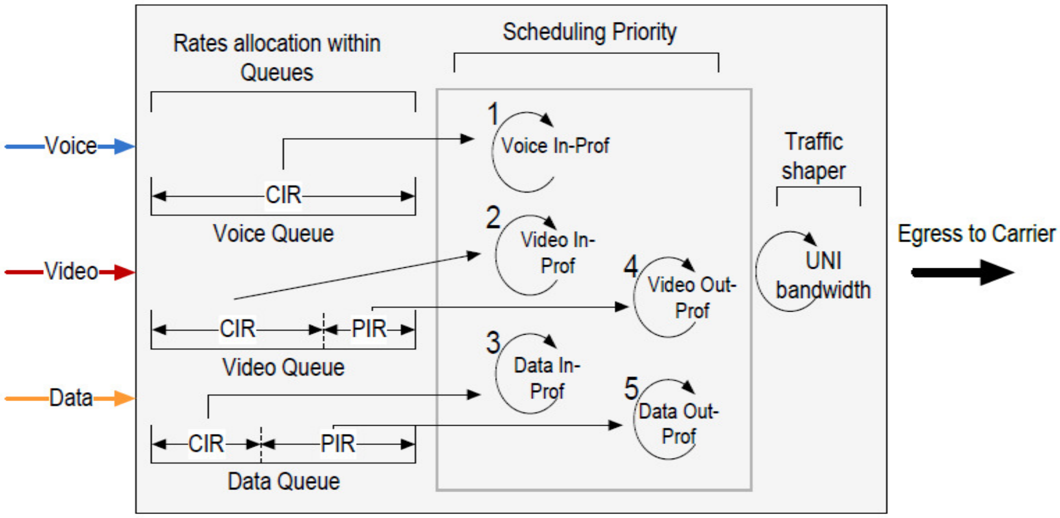


Figure. 4 Egress Traffic Scheduling at Overbooked UNI

Consider $N_{voice}(i)$ voice connections in cluster at time slot i, total CIR of voice stream at CE is assigned with full capacity as follows:

$$CIR_{voice}(i) = PIR_{voice}(i) = N_{voice}(i)R_o \tag{10}$$

Since Ro is the packet rate of voice ON state, there is no overbooking adopted for voice traffics and all voice traffics at CE should handoff to the carrier without frame loss. Therefore, there is no excess packet rate over CIR of voice, and PIR_{voice} equal to CIR_{voice}.

Due to bursty video traffics, it is difficult to assign video FC queue with same allocation rate all the time. However, we can assign CIR bandwidth as much as possible to video bursts, if QoS policy of voice and data service can be satisfied. Therefore, the total CIR of video stream at CE at slot i is shown as follows:

$$CIR_{video}(i) = max\ \{B_{uni} - CIR_{voice}(i) - g_{uni}\ B_{uni},\ 0\} \tag{11}$$

where g_{uni} is a guard ratio defined by a mobile operator, normally $0 < g_{uni} < 0.2$, as a percentage of total UNI bandwidth reserved for data service to provide minimum throughput, which help avoiding the data starvation caused by video bursts. The video bursts over CIR will be scheduled for PIR rate which equals to the whole available UNI bandwidth.

For data service, we assign available UNI bandwidth after CIR rate allocations on voice and video service queues, as well as the whole available UNI bandwidth for the data PIR rate allocation:

$$CIR_{data}(i) = B_{uni} - CIR_{voice}(i) - CIR_{video}\ (i) \tag{12}$$

The PIRs allocations for video and data traffic can only provide best-effort access for video and data services after CIR scheduling, and denotes as,

$$PIR_{video}(i) = PIR_{data}(i) = B_{uni} \quad (13)$$

The above rate allocations follow a strict round robin sequence from voice to BE data shown in Figure 4. CIR assignment can guarantee bandwidth assignment while PIR make use of maximum available resource left after CIR assignments. All idle bandwidth left in CIR rate allocation step will become available bandwidth for the PIR allocation in the same slot.

6. PERFORMANCE EVALUATIONS

In this section, we first evaluate the network performance on overbooked UNI with ns-2 simulation tool [11-13], which has been widely adopted for network simulations and performance evaluations. A dumbbell topology is adopted to simulate the traffic on UNI, as shown in Figure 5. The network scenarios are built from the following elements: (1) Hub CE connects with multiple connections, containing at least one voice, video, and data connections; (2) We assume that congestion only happens when traffic egress from CE to PE; (3) CE supports the trTCM scheduling frame where our proposed dynamic bandwidth control algorithm in section 4 for performance comparison.

Table 1. System Parameters for Simulation

t_{slot}	B_{uni}	η	L_{voice}	L_{video}	L_{data}	FLR_{video}
1s	20Mbps	0.8	8ms	20ms	100ms	10^{-3}

We adopt 3GPP voice model with ON/OFF time interval following exponential distributions: mean active time as 600 millisecond and mean idle time as 400 milliseconds, and active data rate is a constant rate as 24 Kbytes per second. We use a contributed MPEG4 video traffic generator model which is derived from a video trace, with initial seed as 0.3 and rate factor as 60. We adopt FTP application model with TCP transport and file size 100 mega bytes for data service. The frame size of all three traffic model is set as 1024 bytes.

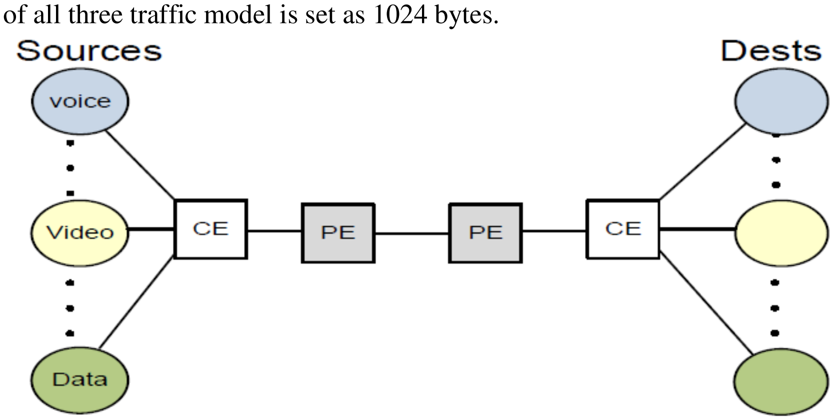


Figure 5. Dumbbell Topology for Network Simulation

Figure 6 shows the result of voice throughput when aggregated video connections increase at the UNI egress point. We can find that total voice throughputs for 5 voice connections and 10 voice connections are about 51 Kbps and 98 Kbps respectively, even when the UNI becomes congested due to large number of video bursts. This is because each voice connection is assigned full bandwidth based on its active state rate at UNI, which guarantee the high priority voice throughput and related Key Performance Index (KPI) performances.

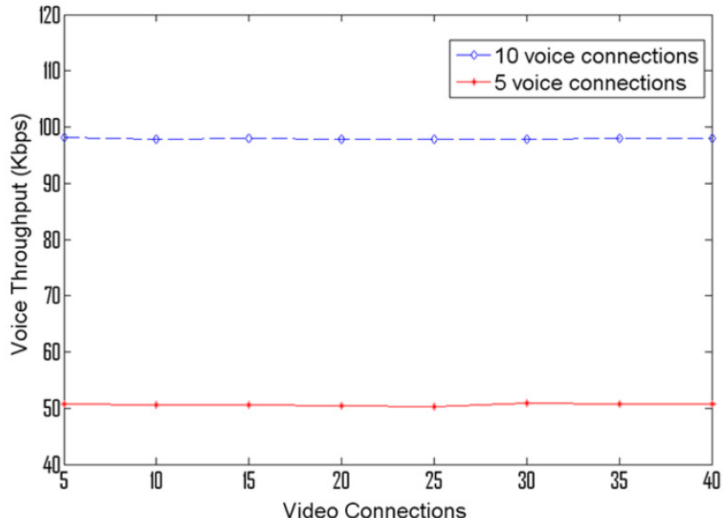


Figure 6. Voice Throughput Performance

Since video service has higher priority for resource allocation over data service, it has been shown in Figure. 7 that data throughput shrinks when video connections increase. However, due to minimum guard CIR bandwidth reserved as 20% of total 20Mbps UNI bandwidth, the data throughput is always above 4 Mbps, which satisfy minimum data throughput requirement of mobile operator for network management and signaling purposes.

Figure 8 shows the frame loss ratio of video traffic when total video connections increase. There are total 5 voice connections and 1 data connection. For static trTCM scheme, we set CIR ratios for voice, video and data as 20%, 40%, and 40% of the UNI bandwidth, respectively. PIR ratio of voice is same as its CIR ratio, and PIR ratios of video and data are both equal to 100% of UNI bandwidth. It is found that video frame loss ratio for static trTCM are almost same as our dynamic bandwidth allocation scheme when video traffic intensity is small. However, the frame loss ratio of static trTCM increases more quickly than our dynamic scheme when heavy video bursts happen. This is because our dynamic scheduling method can make use of most idle bandwidth in each slot for video burst bandwidth allocation after it satisfies requirements of voice bandwidth and minimum data throughput, while in static trTCM, video connections have fixed ratio on CIR scheduling round and cannot compete for idle resource efficiently.

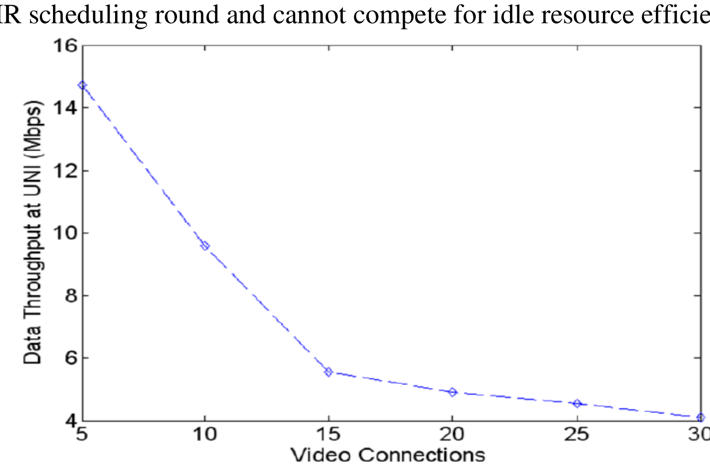


Figure 7. Data Throughput

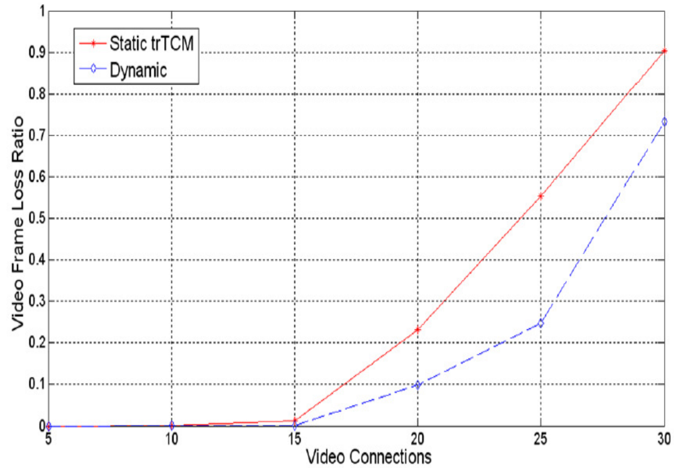


Figure 8. Video FLR

Similarly, Fig. 9 shows the results of video frame latency performance with 5 voice, 10 video and 1 data connections at UNI. We can see that the latency range of our dynamic scheme is smaller than static trTCM method. This is because video connection can get more bandwidth allocation rate in the proposed dynamic scheduling scheme with more CIR and PIR bandwidths. Therefore, the average waiting time in service queue is smaller than the static trTCM method, which determines the video frame latency performance.

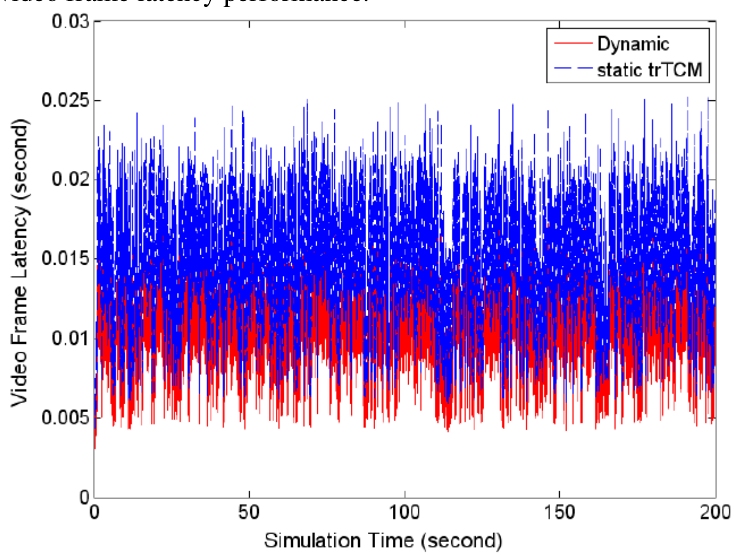


Figure 9. Video Frame Latency

Then we evaluate the overbooking ratio selection with several production microwave clusters in a national cellular network. For proprietary information protection, we neglect locations and names of the selected clusters and only show related performance data, as shown in table 2. For each cluster, the service outage threshold ε is set as 0.001%, and maximum allowable utilization η in equation (7) is set as 0.8. μ_{peak} and σ_{peak} of each UNI are calculated through daily peak UNI throughput data over one year. The peak throughput with 99% confidence interval based on estimation is denoted as T_{peak}. Compared to a linear piecewise relationship between UNI bandwidth and overbooking ratio in [7], we found that peak UNI backhaul throughputs of clusters are different to each other and not strictly follows a monotonically non-increasing relationship with total CIR values of a cell cluster. This demonstrates that traffic intensity and statistical multiplexing gains of clusters are different from each other, even they have same CIR values or cluster size. So it is more appropriate to adopt our statistical estimation method for cluster based overbooking estimation, rather than using the linear piecewise based scheme which derived overbooking ratio only based on UNI CIR value.

Table 2. Real Network Utilization Performance & Overbooking Ratio

Cluster	C1	C2	C3	C4	C5
K	2	3	3	4	4
$\sum_{i=1}^{K} CIR(i)$	100	150	150	200	200
T_{peak}	40.7	33.1	76.3	21.9	80.0
O_{uni}	40.11%	23.76%	52.62%	10.85%	44.43%
B_{uni}	60	50	100	30	110

We further investigate cost saving through overbooking. Figure.10 shows a pricing example from the Carrier Ethernet Market website [14] which is monotonically increasing with UNI bandwidth. It is worthy to notice that some carrier vendors provide discrete UNI bandwidth options for UNI handoff. Once we apply overbooking, the UNI bandwidth can be reduced without deteriorating service quality. The new UNI bandwidth may in same or different pricing range, compared to original UNI bandwidth. For example, when UNI bandwidth drops from 200 Mbps to 100 Mbps, the new pricing can be saved by 40.1%, as shown in Figure 11. In this case, the mobile operator can save transport cost from UNI overbooking. But when UNI bandwidth changes from 50 Mbps to 35 Mbps, the new pricing is same as the old one. Then for the cost perspective, the mobile operator would rather keep original UNI bandwidth from carrier for the cellular cluster than do UNI overbooking.

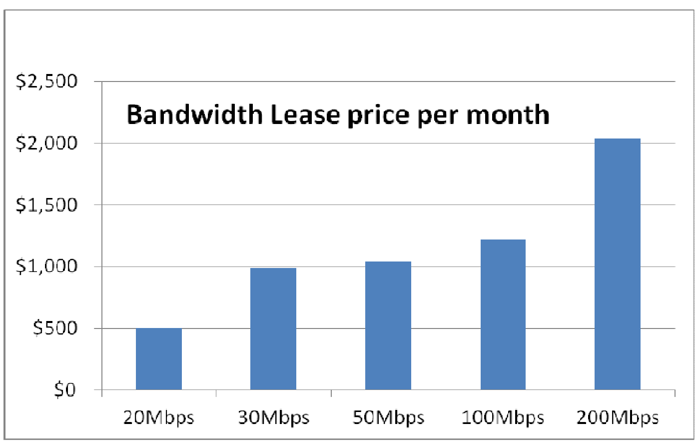


Figure 10. UNI Pricing

Full UNI (Mbps)	UNI band overbooking (Mbps)	Cost Saving (%)
200	100	40.1%
100	50	14.8%
50	35	0%

Figure 11. Cost Saving Map

7. CONCLUSION

This paper proposes an efficient overbooking framework for mobile backhaul. We first develop a statistical estimation method to derive a safe overbooking factor for a given UNI. Then two novel transport architectures are proposed to help cellular cluster backhauls with overbooked UNI bandwidth in carrier Ethernet. A novel bandwidth control algorithm has been derived at customer edge device to provide Quality-of-Service for multimedia traffics over overbooked UNI, with SLA policing and protection on high priority services. Experimental networking data analysis and simulations show that our new schemes can benefit mobile operators in resource utilization efficiency, carrier Ethernet cost saving and backhaul performance.

REFERENCES

[1] P. Briggs, R. Chundury, and J. Olsson, "Carrier ethernet for mobile backhaul," Communications Magazine, IEEE, vol. 48, no. 10, pp. 94–100, 2010.

[2] K. Fouli and M. Maier, "The road to carrier-grade ethernet," Communications Magazine, IEEE, vol. 47, no. 3, pp. S30–S38, 2009.

[3] F. Brockners, Metro Ethernet Services and standardization, Metro Ethernet Forum Std., 2005.

[4] S. Little, "Is microwave backhaul up to the 4g task?" Microwave Magazine, IEEE, vol. 10, no. 5, pp. 67–74, 2009.

[5] A. W. M. Stasiak, M. Glabowski and P. Zwierzykowski, Modelling and dimensioning of mobile wireless networks: from GSM to LTE. John Wiley & Sons.

[6] X. Li, et al., "Dimensioning of the lte s1 interface," in Proc. IEEE Wireless and Mobile Networking Conference, 2010 Third Joint IFIP, 2010, pp. 1–6.

[7] F. Huang, C. Deccio, R. Ball, M. Clement, and Q. Snell, "A piecewise linear approach to overbooking," in Proc. IEEE High Performance Switching and Routing Conference (HPSR'04), 2004.

[8] A. Checko, L. Ellegaard, and M. Berger, "Capacity planning for carrier ethernet lte backhaul networks," in Proc. IEEE Wireless Communications and Networking Conference (WCNC'12), 2012.

[9] Metro Ethernet Forum, "Metro Ethernet Services – A technical Overview", http:/ / metroethernetforum.org/metroethernet-services.pdf

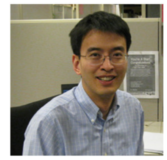
[10] C. Liu, C. Zhou, N. Pissinou and K. Makki, "Adaptive Quality of Service Provisioning in IEEE 802.11 Wireless Local Access Network," International Journal of Sensors, Wireless Communications and Control, Vol. 2, pp.117-129 2012.

[11] ISI, "The network simulator – ns2," http://www.isi.edu/nsnam/ns/

[12] C. Liu, C. Zhou, "An improved interworking architecture for UMTS-WLAN tight coupling," IEEE Wireless Communications and Networking Conference, Vol. 3, pp. 1690-1695, 2005.

[13] C. Liu, C. Zhou, "Providing quality of service in IEEE 802.11 WLAN," IEEE International Conference on Advanced Information Networking and Applications, vol. 1, pp. 817-824, 2006.
[14] Carrier Ethernet Market, http://www.carrier-ethernet-market.com/en/

Authors



Chunming Liu is currently a Principal Engineer at T-Mobile, Bellevue, WA, United States, and working on mobile network strategy, design and standards. He received Ph.D. degree in Electrical Engineering from the Florida International University in 2006. From 2006 to 2008, he worked as a senior software engineer at Motorola for 4G WiMAX product design and software development. From 2008 to 2010, he worked for Microsoft as a software development engineer on Windows Phone Platform. He has over 10 years R&D experience in areas of wireless communications, mobile computing, and telecommunications, with 18 publications in refereed journals, conferences and book chapters. He has invented 6 US / International patents (pending). He has been Editorial Board Member of Journal of Networks, International Journal of Wireless and Mobile Networks, and International Journal of Privacy and Health Information Management. He is an IEEE Senior Member.

Bryan Fleming is Vice President of Technical Systems and Business Operations, T-Mobile USA. He is responsible for the Carrier Management, Engineering & EIT Procurement organizations at T-Mobile. His teams direct telecoms spend, infrastructure purchases and support engineering and supply chain operations. His team led T-Mobile's the FTTC/Ethernet program which supports America's Largest 4G Network® covering over 220 million POPs. With Bryan's 22 years of experience in Telecom, he has proven that he is an innovative Telecom leader executing high ROI solutions that fuel multibillion-dollar revenues, driving excellence in operational performance, and sustaining superior customer relationships. Throughout his career, Bryan has held various leadership positions in RF engineering, operations, finance, procurement, and strategy. Bryan holds a Bachelor's degree in Business Management from Columbia College and an MBA from Auburn University.