TCP Burst Resiliency in the IP-20 Platform

Introduction

Demand for network capacity and higher data rates continues to increase unabated. Adding capacity to networks helps operators meet this demand, however, operators need to be aware of several technical limitations, beyond capacity, that interfere with user experience.

While it is logical to think that pushing higher data rates equates to greater throughput, this is not always the case. In fact, forcing on an Ethernet service a data rate that is greater than the service’s capacity will result in a poor user experience as packets may be dropped and re-transmissions may be required resulting in net throughputs much lower than expected. In Long Term Evolution (LTE) networks running the TCP protocol, instantaneous congestion issues can arise frequently resulting in instantaneous, large bursts that contribute further to throughput reduction.

In order to maximize throughput, it is vital to shape the bandwidth offered to the network to match the bandwidth profile of the service level agreement (SLA).

Within its IP-20 platform Quality of Service (QoS) mechanism, Ceragon has developed an integrated approach that solves the pesky problem of instantaneous bursts. The QoS mechanism utilizes three layers starting from the standard QoS layer, continuing with the enhanced packet buffer management layer, and employing Hierarchical QoS (H-QoS) as the final layer. The three layers exploit a huge buffer set aside for storing long bursts as they occur. Integrated with the three QoS layers, Weighted Random Early Detect (WRED) makes bandwidth use highly efficient in order to deliver the best possible user experience.

The Challenge

In some cases, due to bursts, users perceive that they are receiving lower throughput than they expect. In actuality, there are many IP and Application Layer factors that affect the network efficiency when utilizing an Ethernet service and most of these are under the operator’s direct control. Proper use of Ceragon’s IP-20 platform alleviates these problems.
The most important problem discussed in this paper is *rate adaptation*—connecting a network element line with a high data rate limited only by physical interface and cabling, to a radio interface where the data rate is limited by the channel bandwidth (spectrum efficiency). This can result in the dropping of frames. As a consequence, the overall capacity is reduced.

LTE networks provide an excellent example of the need to address rate adaptation scenarios that are characterized by bursty traffic. LTE technology was developed with the objective of high data rates, low-latency and packet-optimized radio access technology.

*Table 1. The peak data rate in the evolution of mobile standards*

<table>
<thead>
<tr>
<th>Level</th>
<th>Standard</th>
<th>Uplink (bps)</th>
<th>Downlink (bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G</td>
<td>GSM</td>
<td>9.6k</td>
<td>14.4k</td>
</tr>
<tr>
<td>2.5G</td>
<td>Edge</td>
<td>384k</td>
<td>513k</td>
</tr>
<tr>
<td>3G</td>
<td>UMTS</td>
<td>384k</td>
<td>2.0M</td>
</tr>
<tr>
<td>3G+</td>
<td>HSDPA/HSUPA</td>
<td>5.8M</td>
<td>14.4M</td>
</tr>
<tr>
<td>3G+</td>
<td>HSPA+</td>
<td>11M</td>
<td>42M</td>
</tr>
<tr>
<td>4G</td>
<td>LTE (Cat 5)</td>
<td>75M</td>
<td>300M</td>
</tr>
<tr>
<td>4G</td>
<td>LTE Advanced (Cat 8)</td>
<td>600M</td>
<td>1200M</td>
</tr>
</tbody>
</table>

Even though it was not originally designed for real-time applications, TCP is the most widely used protocol for wire and wireless networks. In its early days, TCP was applied to wired networks, but later, when applied over wireless channels, it showed poor performance mainly due to high error rates. In order to compensate for these errors, LTE employs many error recovery techniques at the Link Layer; these partially overlap with error recovery performed at the Transport Layer of TCP/IP.
TCP is a connection-oriented protocol that requires handshaking to set up end-to-end communications. One of its major advantages is its end-to-end error detection, recovery and data-flow control. However, TCP exhibits certain disadvantages when it comes to mobile networks, for example, proper adjustment to changes in the link rate due to Adaptive Coding & Modulation (ACM) reaction to microwave fading. In addition, there is some delay involved in setting up a reliable link since a significant handshaking procedure is necessary for starting a session. TCP’s continual acknowledgments and its need to buffer data for potential re-transmission can degrade performance as well (dropped packets in cases of insufficient buffer size). Furthermore, performance can be significantly compromised by transmission latency or frame delay.

The Relationship Between TCP and Bursts

TCP uses a sliding window mechanism which utilizes the network bandwidth efficiently since it allows the sender to transmit multiple packets before waiting for an acknowledgement from the receiver. Considering a sequence of packets to be transmitted, the protocol places a window on the sequence and transmits all packets that lie inside the window.
Figure 2(a) is an example of a sliding window protocol with eight packets in the window. In (b), when an acknowledgement for Packet 1 has been received, the window slides so that Packet 9 can be sent.

The number of packets that can be transmitted without an acknowledgement is constrained by the window size. TCP allows the window size to vary over time. Each acknowledgement, which specifies how many octets have been received, contains a window advertisement that specifies how many additional octets of data the receiver is prepared to accept, i.e., the receiver’s current buffer size. In response to an increased window advertisement, the sender increases the size of its sliding window and proceeds to send octets that have not been acknowledged. In response to a decreased window advertisement, the sender decreases the size of its window and refrains from sending octets beyond the boundary. The advantage of using a variable size window is that it provides flow control as well as reliable transfer. Having the capability to build proper transmit buffers is crucial since the transmitter cannot be stopped all at once due to the fact that the flow control is dependent upon closing the window.

We make the following observations:

- TCP network traffic is bursty in nature. The burstiness is derived directly from a flow-control definition allowing the transmission of a segment of octets (window size) in a given time.

- A varying window size may grow so large that, when the transmitter sends all the octets corresponding to the current size of the window, it causes instantaneous congestion and fast bursts in cases when the transmit buffer does not have ample space. To avoid this potential problem, buffers of sufficient size, i.e., large enough to accommodate all the data, must be allocated.
• Initially, the TCP burst size grows linearly as the bit rate and latency increase. Later, the increment may be exponential (depending on the particular TCP algorithm). However, large bursts are never recommended as they may affect other flows that will need to wait for the whole burst to be transmitted.

• As the number of TCP flows increases, network efficiency improves since, when several TCP flows share the same limited bandwidth (radio), the throughput of each flow is smaller and likewise the burst length that it can create. In addition, when there are several TCP flows, statistically, the timing of transmit and receive of each flow is different so bursts don’t necessarily occur at the same time. As the number of TCP flows decreases, the effect of large bursts and concomitant congestion increases. The worst case of inefficiency due to congestion is when only one TCP flow is used. When traffic is dropped due to congestion of a single stream, a large quantity of data may be lost and even cause a TCP ‘slow start’. In the case of multiple streams, the loss is merely several packets that will be retransmitted on the fly (fast path without slow start).

An additional significant factor which determines the perceived bandwidth performance is the Round Trip Time (RTT or link delay), the amount of time it takes for bits to travel across the network and back. Actual network utilization measures depend on RTT.

Since TCP transmission requires acknowledgements, a sender needs to buffer at least enough bits to continue sending until an acknowledgement is received. In order to keep a congested link as busy as possible so as to maximize the throughput of the network, the buffer required is the same as the TCP receive window and equals:

\[
(1) \quad \text{Buffer Size [bits]} = \text{Flow Bandwidth [bps]} \times \text{RTT [sec]}
\]

IP-20 handles the bursty traffic employing three levels in its enhanced QoS mechanism as described in the next section.

**Ceragon’s Solution for Overcoming Burst Situations**

It is recommended not to inject into the network a throughput rate that is higher than the given bandwidth capabilities. Therefore, shaping the traffic prior to injecting it into the network element is always the preferable solution and is, in fact, an MEF (10.2) recommendation. Since traffic shaping is not always possible, alternatives must be found within the capabilities of the network element. For such cases, one should exploit a solution based on IP-20’s enhanced QoS capabilities.

Exploiting IP-20’s enhanced QoS features such as policers, classification, advanced scheduling, monitoring and other techniques, provides several benefits. These and other enhanced QoS features enable operators to provide differentiated services with strict SLAs
while maximizing network resource utilization. Three levels of the enhanced QoS capabilities are directly related to overcoming burst problems.

**Quality of Service**

Quality of Service deals with the way frames are handled within the switching fabric. QoS is required in order to deal with many different network scenarios such as traffic congestion, packet availability and delay restrictions. FibeAir IP-20’s personalized QoS enables operators to handle a wide and diverse range of scenarios. Its smart QoS mechanism operates from the frame’s ingress into the switching fabric until the moment the frame egresses via the destination port. The basic QoS diagram is described in Figure 3:

![Figure 3: QoS concept](image)

The basic mode of the IP-20 QoS mechanism provides eight transmission queues per port with the following attributes:

- Fixed 1-Mbit buffer per queue
- Total of 8 Mbits of buffer space for the 8 queues

Correct traffic engineering and proper use of the QoS feature should allow users to overcome and cope with the TCP bursts allocated to each of the queues.

Another of IP-20’s capabilities for improving throughput under TCP is *Weighted Random Early Detect* (WRED). Used with tail-drop for congestion management, the WRED profile determines the buffer depth for the IP-20. WRED can increase capacity utilization of TCP traffic by eliminating the phenomenon of global synchronization that occurs when TCP flows sharing bottleneck conditions receive loss indications at around the same time. This phenomenon can result in periods during which link bandwidth utilization drops significantly as a consequence of simultaneous falling of all the TCP flows to *slow start*. WRED eliminates the occurrence of traffic congestion peaks by restraining the transmission rate of the TCP flows. Each queue’s occupancy level is monitored by the WRED mechanism.
and randomly selected frames are dropped before the queue becomes overcrowded. For the basic mode of a fixed queue buffer size limit, only tail drop is implemented.

Enhanced Packet Buffer Management

To strengthen the handling of burst effects, usually for the best-effort class of service, IP-20’s Enhanced Packet Buffer Management\(^1\) is used. The feature includes a large buffer allocation mechanism which allows specification of the buffer size of each of the eight priority queues according to the specific service. The total amount of memory dedicated for queue buffers is 4Gbit while the size of each queue can be set up to 64Mbit in granularities of 64kbit. The mode which enables a configurable (non-default) queue buffer size limit for Green and Yellow frames also enables WRED. The default queue buffer size limit is 1Mbits for Green frames and 0.5 Mbits for Yellow frames.

The buffer size must be calculated carefully based on two considerations:

- **Latency.** If low latency is required (dropping frames in the queue is preferable to increasing latency), small buffer sizes are preferred. The actual effective buffer size of the queue can be set to the minimum.

- **Throughput immunity to fast bursts.** When the traffic is characterized by fast and/or long bursts, it is recommended to increase the buffer sizes of the priority queues to prevent packet loss. Of course, this will have an effect on latency. The increase in delay is experienced only in the case of full buffers. If the buffers don’t fill up, then there isn’t any delay penalty.

Burst sizes can be configured as a tradeoff between the latency and immunity to bursts according to application requirements.

Hierarchical QoS

For networks that require large numbers of services, IP-20’s H-QoS\(^2\) supplies a virtual pipe experience per each service. Each service gets its own personalized treatment as the H-QoS mechanism can sub-classify between “traffic types” (Classes of Service - CoS) from separate customers / services.

H-QoS adds service bundles with dedicated queues to interfaces. Without this feature, only the default eight queues per port are supported. In this mode, users can associate services from the service model to configurable groups of eight transmission queues (service bundles), from a total 8K queues. In H-QoS mode, IP-20 performs QoS in a hierarchical

\(^1\) License Required
\(^2\) License Required
manner where the egress path is managed on three levels: ports, service bundles and specific queues. This enables users to fully distinguish between streams, thus providing a true service level to customers. Dynamic buffer management can be accomplished according to the following:

- Queue buffering – total memory of up to 4Gbit
- Up to 8K Queues with up to 64Mbit per queue

![Figure 4: H-QoS concept](image)

In addition, H-QoS helps to confront additional aspects of the TCP bursts phenomena:

**Over Subscription**
Users can define maximum buffer space per queue even if the total amount exceeds the total memory size (4Gbit). The assumption is that the transmission queues are not synchronized and do not transmit at the same time. Users may also work in the conservative mode, of course.

**Burst Isolation**
One of H-QoS’s main strengths is the ability to differentiate services—in part, its separate treatment of bursts related to each of the data services. H-QoS makes sure that a large burst on one service won’t occupy the whole pipe preventing other services from being transmitted. Actually, this deals with the *Head of a Line Blocking* phenomena where lower-priority frames affect high-priority frames that won’t be transmitted prior to an empty transmission pipe. The burst isolation feature is applicable in many applications such as:
• RAN Sharing – in order to assure fairness and burst isolation of operators sharing the same radio infrastructure

• Diverse mobile technologies – 2G, 3G, LTE, LTE-A, etc. sharing the same radio infrastructure, when large bursts are expected for the latter (larger uplink/downlink capacities). Burst isolation guarantees that none of the different technologies will be starved and all will be transmitted equally.

Figure 5 shows an abstract example where two operators (services) sharing the same CoS are joined together in an aggregation point. One of the operators transmits a burst of data which fills the transmission pipe. With standard QoS, Operator 2’s data is starved and cannot be transmitted. Since, in H-QoS even the same CoS are separated into different services, fairness, proper queue management and optimum throughput utilization can be guaranteed.

![Diagram of H-QoS burst isolation](image)

*Figure 5: H-QoS burst isolation*
Practical Examples

Below are two examples that demonstrate the use of IP-20’s capabilities to maximize throughput despite TCP burstiness.

Example 1
The first step is to calculate the maximum buffer size required according to Formula (1) (as shown previously):

\[ Buffer\ Size\ [bits] = Flow\ Bandwidth\ [bps] \times RTT\ [sec] \]

The example considers an LTE-A service and assumes three different classes of service (traffic types): premium (gold), silver and bronze. It demonstrates the use of Formula (1). The example shows how to buffer the TCP data while maintaining different services.

Common values for the LTE-A service suggest (for one TCP flow of end to end connection):

\[ Link\ Delay < 60ms,\ TCP\ bandwidth = 1000Mbps \]

Applying the maximum latency scenario, substituting (2) into (1) yields:

\[ Buffer\ Size = 60Mbit \]

The large buffer size calculated in (3) is easily supported by the IP-20’s buffers which can accommodate up to 4Gbit of memory and 64Mbit per queue. Planning for burst handling is as follows:

- A 60Mbit buffer size will be allocated to the TCP flow to fill the data for three separate queues each with a different CoS.

- If needed (not always required), the remainder can be allocated between up to five queues (with higher priorities) and may be dedicated for management protocols, synchronization packets, Pseudowire services (voice and signaling), etc.
**Example 2**

Assuming RAN sharing between two network operators:

1. Operator 1: TCP Flow = 150Mbps, RTT = 100ms
2. Operator 2: TCP Flow = 200Mbps, RTT = 50ms
3. Shared radio capacity = 100Mbps only!

The required buffers can be calculated according to (1):

\[
\begin{align*}
(4) \text{ Operator 1 Buffer} &= 15 \text{ Mbit} \\
(5) \text{ Operator 2 Buffer} &= 10 \text{ Mbit}
\end{align*}
\]

The network planner’s goal is to make sure that each of the operators will enjoy an equal portion of the aggregated radio capacity (fairness). However, when using standard QoS, this cannot be the case, i.e., TCP flows with different round-trip times that share the same bottleneck link will not obtain equal portions of the available bandwidth. Actually, the ratio between their throughputs should be inversely proportional to the ratio of their RTTs:

\[
\begin{align*}
(6) \frac{\text{Throughput}_i}{\text{Throughput}_j} &= \frac{\text{RTT}_j}{\text{RTT}_i}
\end{align*}
\]
In this case, Operator 2’s throughput would be higher by product of 2 than Operator 1’s – 66.6Mbps compared to 33.3Mbps (remember: radio throughput is only 100Mbps!). Only H-QoS can resolve this case by separating the two operators into two separate service bundles and, in addition:

1. Defining the buffer required per each TCP flow
2. Configuring equal weight per each service bundle
3. 

![Figure 7: H-QoS for fairness](image)

Generally, an additional advantage for using a scheme such as the aforementioned is the fact that we can use smaller buffers (with all the benefits explained previously) than in standard QoS, as long as the strict SLA is attained per service.

**Summary**

Bursty traffic is an emerging impediment to maximizing throughput in multi-carrier LTE-Advanced networks that employ the TCP protocol. In this Technical Brief, we defined the nature of the problem of bursts and explained how they occur in TCP. Ceragon has solved the problem in its IP-20 platform with its innovative technology of H-QoS and enormous buffer sizes which allow network planners to handle any type of TCP-oriented traffic in order to maximize the utilization of network resources. User experience is enhanced.
References


J. Lee, S. Bohacek, J.P. Hespana, K. Obrazcka, “A Study of TCP Fairness in High-Speed Networks”


About Ceragon

Ceragon Networks Ltd. (NASDAQ: CRNT) is the #1 wireless backhaul specialist. We provide innovative, flexible and cost-effective wireless backhaul and fronthaul solutions that enable mobile operators and other wired/wireless service providers to deliver 2G/3G, 4G/LTE and other broadband services to their subscribers. Ceragon’s high-capacity solutions use microwave technology to transfer voice and data traffic while maximizing bandwidth efficiency to deliver more capacity over longer distances under any deployment scenario. Based on our extensive global experience, Ceragon delivers turnkey solutions that support service-provider profitability at every stage of the network lifecycle enabling faster time to revenue, cost-effective operation and simple migration to all-IP networks. As the demand for data drives the ever-expanding need for capacity, Ceragon is committed to serving the market with unmatched technology and innovation, ensuring effective solutions for the evolving needs of the marketplace. Our solutions are deployed by more than 430 service providers in over 130 countries.