                    Pseudowire Congestion Considerations

Abstract

   Pseudowires (PWs) have become a common mechanism for tunneling
   traffic and may be found in unmanaged scenarios competing for network
   resources both with other PWs and with non-PW traffic, such as TCP/IP
   flows. Thus, it is worthwhile specifying under what conditions such
   competition is acceptable, i.e., the PW traffic does not
   significantly harm other traffic or contribute more than it should to
   congestion. We conclude that PWs transporting responsive traffic
   behave as desired without the need for additional mechanisms. For
   inelastic PWs (such as Time Division Multiplexing (TDM) PWs), we
   derive a bound under which such PWs consume no more network capacity
   than a TCP flow. For TDM PWs, we find that the level of congestion
   at which the PW can no longer deliver acceptable TDM service is never
   significantly greater, and is typically much lower, than this bound.
   Therefore, as long as the PW is shut down when it can no longer
   deliver acceptable TDM service, it will never do significantly more
   harm than even a single TCP flow. If the TDM service does not
   automatically shut down, a mechanism to block persistently
   unacceptable TDM pseudowires is required.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   A pseudowire (PW) (see [RFC3985]) is a construct for tunneling a
   native service, such as Ethernet or TDM, over a Packet Switched
   Network (PSN), such as IPv4, IPv6, or MPLS. The PW packet
   encapsulates a unit of native service information by prepending the
   headers required for transport in the particular PSN (which must
   include a demultiplexer field to distinguish the different PWs) and
   preferably the 4-byte Pseudowire Emulation Edge-to-Edge (PWE3)
   control word.

   PWs have no bandwidth reservation or control mechanisms, meaning that
   when multiple PWs are transported in parallel, and/or in parallel
   with other flows, there is no defined means for allocating resources
   for any particular PW, or for preventing the negative impact of a
   particular PW on neighboring flows. The case where the service
   provider network provisions a PW with sufficient capacity is well
   understood and will not be discussed further here. Concerns arise
   when PWs share network capacity with elastic or congestion-responsive
   traffic, whether that capacity sharing was planned by a service
   provider or results from PW deployment by an end user.

   PWs are most often placed in MPLS tunnels, but we herein restrict
   ourselves to PWs in IPv4 or IPv6 PSNs; MPLS PSNs are beyond the scope
   of this document. There are several mechanisms that enable
   transporting PWs over an IP infrastructure, including:

   o UDP/IP encapsulations as defined for TDM PWs [RFC4553] [RFC5086]
      [RFC5087],

   o PWs based on Layer 2 Tunneling Protocol (L2TPv3) [RFC3931],

   o MPLS PWs directly over IP according to RFC 4023 [RFC4023], and

   o MPLS PWs over Generic Routing Encapsulation (GRE) over IP
      according to RFC 4023 [RFC4023].

   Whenever PWs are transported over IP, they may compete for network
   resources with neighboring congestion-responsive flows (e.g., TCP
   flows). In this document, we study the effect of PWs on such
   neighboring flows, and discover that the negative impact of PW
   traffic is generally no worse than that of congestion-responsive
   flows [RFC2914] [RFC5033].

   At first glance, one may consider a PW transported over IP to be
   considered as a single flow, on par with a single TCP flow. Were we
   to accept this tenet, we would require a PW to back off under
   congestion to consume no more bandwidth than a single TCP flow under

such conditions (see [RFC5348]). However, since PWs may carry
traffic from many users, it makes more sense to consider each PW to
be equivalent to multiple TCP flows.

The following two sections consider PWs of two types:

Elastic Flows:
    Section 3 concludes that the response to congestion of a PW
    carrying elastic (e.g., TCP) flows is no different from the
    aggregated behaviors of the individual elastic flows, had they not
    been encapsulated within a PW.

Inelastic Flows:
    Section 4 considers the case of inelastic constant bit rate (CBR)
    TDM PWs [RFC4553] [RFC5086] [RFC5087] competing with TCP flows.
    Such PWs require a preset amount of bandwidth, that may be lower
    or higher than that consumed by an otherwise unconstrained TCP
    flow under the same network conditions. In any case, such a PW is
    unable to respond to congestion in a TCP-like manner; although
    admittedly the total bandwidth it consumes remains constant and
    does not increase to consume additional bandwidth as TCP rates
    back off. For TDM services, we will show that TDM service quality
    degradation generally occurs before the TDM PW becomes TCP-
    unfriendly. For TDM services that do not automatically shut down
    when they persistently fail to comply with acceptable TDM service
    criteria, a transport circuit breaker [CIRCUIT-BREAKER] may be
    employed as a last resort to shut down a TDM pseudowire that can
    no longer deliver acceptable service.

Thus, in both cases, pseudowires will not inflict significant harm on
neighboring TCP flows, as in one case they respond adequately to
congestion, and in the other they would be shut down due to being
unable to deliver acceptable service before harming neighboring
flows.

Note: This document contains a large number of graphs that are
necessary for its understanding, but could not be rendered in ASCII.
It is strongly suggested that the PDF version be consulted.

2. Terminology

   The following acronyms are used in this document:

   AIS   Alarm Indication Signal [G775]

   BER   Bit Error Rate [G826]

   BW   Bandwidth

   CBR   Constant Bit Rate

   ES   Errored Second [G826]

   ESR   Errored Second Rate [G826]

   GRE   Generic Routing Encapsulation [RFC2784]

   L2TPv3 Layer 2 Tunneling Protocol Version 3 [RFC3931]

   MOS   Mean Opinion Score [P800]

   MPLS   Multiprotocol Label Switching [RFC3031]

   NSP   Native Service Processing [RFC3985]

   PLR   Packet Loss Ratio

   PSN   Packet Switched Network [RFC3985]

   PW   Pseudowire [RFC3985]

   SAToP   Structure-Agnostic TDM over Packet [RFC4553]

   SES   Severely Errored Seconds [G826]

   SESR   Severely Errored Seconds Ratio [G826]

   TCP   Transmission Control Protocol

   TDM   Time Division Multiplexing [G703]

   UDP   User Datagram Protocol

3. PWs Comprising Elastic Flows

   In this section, we consider Ethernet PWs that primarily carry
   congestion-responsive traffic. We expand on the remark in Section 8
   (Congestion Control) of [RFC4553], and show that the desired
   congestion avoidance behavior is automatically obtained and
   additional mechanisms are not needed.

   Let us assume that an Ethernet PW aggregating several TCP flows is
   flowing alongside several TCP/IP flows. Each Ethernet PW packet
   carries a single Ethernet frame that carries a single IP packet that
   carries a single TCP segment. Thus, if congestion is signaled by an
   intermediate router dropping a packet, a single end-user TCP/IP
   packet is dropped, whether or not that packet is encapsulated in the
   PW.

   The result is that the individual TCP flows inside the PW experience
   the same drop probability as the non-PW TCP flows. Thus, the
   behavior of a TCP sender (retransmitting the packet and appropriately
   reducing its sending rate) is the same for flows directly over IP and
   for flows inside the PW. In other words, individual TCP flows are
   neither rewarded nor penalized for being carried over the PW. An
   elastic PW does not behave as a single TCP flow, as it will consume
   the aggregated bandwidth of its component flows; yet if its component
   TCP flows backs off by some percentage, the bandwidth of the PW as a
   whole will be reduced by the very same percentage, purely due to the
   combined effect of its component flows.

   This is, of course, precisely the desired behavior. Were individual
   TCP flows rewarded for being carried over a PW, this would create an
   incentive to create PWs for no operational reason. Were individual
   flows penalized, there would be a deterrence that could impede
   pseudowire deployment.

   There have been proposals to add additional TCP-friendly mechanisms
   to PWs, for example by carrying PWs over DCCP. In light of the above
   arguments, it is clear that this would force the PW down to the
   bandwidth of a single flow, rather than N flows, and penalize the
   constituent TCP flows. In addition, the individual TCP flows would
   still back off due to their endpoints being oblivious to the fact
   that they are carried over a PW. This would further degrade the
   flow's throughput as compared to a non-PW-encapsulated flow, in
   contradiction to desirable behavior.

We have limited our treatment to the case of TCP traffic carried by
Ethernet PWs (which are by far the most commonly deployed packet-
carrying pseudowires), but it is not overly difficult to show that
our result is equally valid for other PW types, such as ATM or frame-
relay pseudowires.

4. PWs Comprising Inelastic Flows

Inelastic PWs, such as TDM PWs [RFC4553] [RFC5086] [RFC5087], are
potentially more problematic than the elastic PWs of the previous
section. As mentioned in Section 8 (Congestion Control) of
[RFC4553], being constant bit rate (CBR), TDM PWs can't incrementally
respond to congestion in a TCP-like fashion. On the other hand,
being CBR, TDM PWs do not make things worse by attempting to capture
additional bandwidth when neighboring TCP flows back off.

Since a TDM PW consumes a constant amount of bandwidth, if the
bandwidth occupied by a TDM PW endangers the network as a whole, it
might seem that the only recourse is to shut it down, denying service
to all customers of the TDM native service. Nonetheless, under
certain conditions it may be possible to reduce the bandwidth
consumption of an emulated TDM service. A prevalent case is that of
a TDM native service that carries voice channels that may not all be
active. The ATM Adaptation Layer 2 (AAL2) mode of [RFC5087] (perhaps
along with connection admission control) can enable bandwidth
adaptation, at the expense of more sophisticated native service
processing (NSP).

In the following, we will focus on structure-agnostic TDM PWs
[RFC4553] although similar analysis can be readily applied to
structure-aware PWs (see Appendix B). We will show that, for many
cases of interest, a TDM PW, even when treated as a single flow, will
behave in a reasonable manner without any additional mechanisms. We
also show that, at the level of congestion when a TDM PW can no
longer deliver acceptable TDM service, a single unconstrained TCP
flow would typically still consume more capacity than a whole TDM PW.
Therefore, to ensure that a TDM PW does not inflict significantly
more harm than a TCP flow, it suffices to shut down a TDM PW that is
persistently unable to deliver acceptable TDM service. This shutting
down could be accomplished by employing a managed transport circuit
breaker, by which we mean an automatic mechanism for terminating an
unresponsive flow during persistently high levels of congestion
[CIRCUIT-BREAKER]. Note that a transport circuit breaker is intended
as a protection mechanism of last resort, just as an electrical
circuit breaker is only triggered when absolutely necessary.

For the avoidance of doubt, the above does not say that a TDM PW
should be shut down when it becomes TCP-unfriendly. It merely says
that the act of shutting down a TDM PW that can no longer deliver
acceptable TDM service ensures that the PW does not contribute to
congestion significantly more than a TCP flow would. Also, note that
being unable to deliver acceptable TDM service for a short amount of
time is insufficient justification for shutting down a TDM PW. While
TCP flows react within a round-trip time, service commissioning and
decommissioning are generally time-consuming processes that should
only be undertaken when it becomes clear that the congestion is not
transient.

In order to quantitatively compare TDM PWs to TCP flows, we will
compare the effect of TDM PW traffic with that of TCP traffic having
the same packet size and delay. This is potentially an overly
pessimistic comparison, as TDM PW packets are frequently configured
to be short in order to minimize latency, while TCP packets are free
to be much larger.

There are two network parameters relevant to our discussion, namely
the one-way delay (D) and the packet loss ratio (PLR). The one-way
delay of a native TDM service consists of the physical time-of-flight
plus 125 microseconds for each TDM switch traversed, and is thus very
small as compared to typical PSN network-crossing latencies. Since
TDM services are designed with this low latency in mind, emulated TDM
services are usually required to have similar low end-to-end delay.
In our comparisons, we will only consider one-way delays of a few
milliseconds.

Regarding packet loss, the relevant RFCs specify actions to be
carried out upon detecting a lost packet. Structure-agnostic
transport has no alternative to outputting an "all-ones" Alarm
Indication Signal (AIS) pattern towards the TDM circuit, which, when
long enough in duration, is recognized by the receiving TDM device as
a fault indication (see Appendix A). TDM standards (such as [G826])
place stringent limits on the number of such faults tolerated.
Calculations presented in Appendix A show that only loss
probabilities in the realm of fractions of a percent are relevant for
structure-agnostic transport. Structure-aware transport regenerates
frame alignment signals, thus avoiding AIS indications resulting from
infrequent packet loss. Furthermore, for TDM circuits carrying voice
channels, the use of packet loss concealment algorithms is possible
(such algorithms have been previously described for TDM PWs).
However, even structure-aware transport ceases to provide a useful
service at about 2 percent loss probability. Hence, in our
comparisons we will only consider PLRs of 1 or 2 percent.

TCP Friendly Rate Control (TFRC) [RFC5348] provides a simplified
formula for TCP throughput as a function of round-trip delay and
packet loss ratio.

$$X = \frac{S}{R \left(\sqrt{2p/3} + 12 \sqrt{3p/8}\ p\ (1+32p^2) \right)}$$

where:

   X is the average sending rate in bytes per second,

   S is the segment (packet payload) size in bytes,

   R is the round-trip time in seconds,

   p is the packet loss probability (i.e., PLR/100).

We can now compare the bandwidth consumed by TDM pseudowires with
that of a TCP flow for a given packet loss ratio and one-way end-to-
end delay (taken to be half the round-trip delay R). The results are
depicted in the accompanying figures (available only in the PDF
version of this document). In Figures 1 and 2, we see the
conventional rate vs. packet loss plot for low-rate TDM (both T1 and
E1) traffic, as well as TCP traffic with the same payload size (64 or
256 bytes respectively). Since the TDM rates are constant (T1 and E1
having payload throughputs of 1.544 Mbps and 2.048 Mbps
respectively), and Structure-Agnostic TDM over packet (SAToP) can
only faithfully emulate a TDM service up to a PLR of about half a
percent, the T1 and E1 pseudowires occupy line segments on the graph.
On the other hand, the TCP rate equation produces rate curves
dependent on both one-way delay and packet loss.

For large packet sizes, short one-way delays, and low packet loss
ratios, the TDM pseudowires typically consume much less bandwidth
than TCP would under identical conditions. For small packets, long
one-way delays, and high packet loss ratios, TDM PWs potentially
consume more bandwidth, but only marginally. Furthermore, our
"apples to apples" comparison forced the TCP traffic to use packets
of sizes smaller than would be typical.

Similarly, in Figures 3 and 4 we repeat the exercise for higher rate
E3 and T3 (rates 34.368 and 44.736 Mbps respectively) pseudowires,
allowing delays and PLRs suitable for these signals. We see that the
TDM pseudowires consume much less bandwidth than TCP, for all
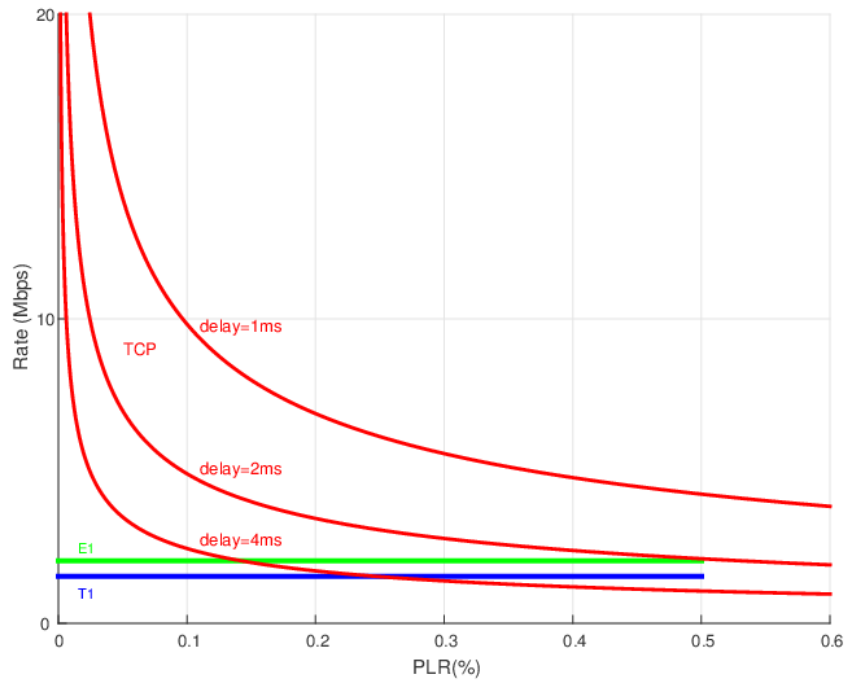reasonable parameter combinations.

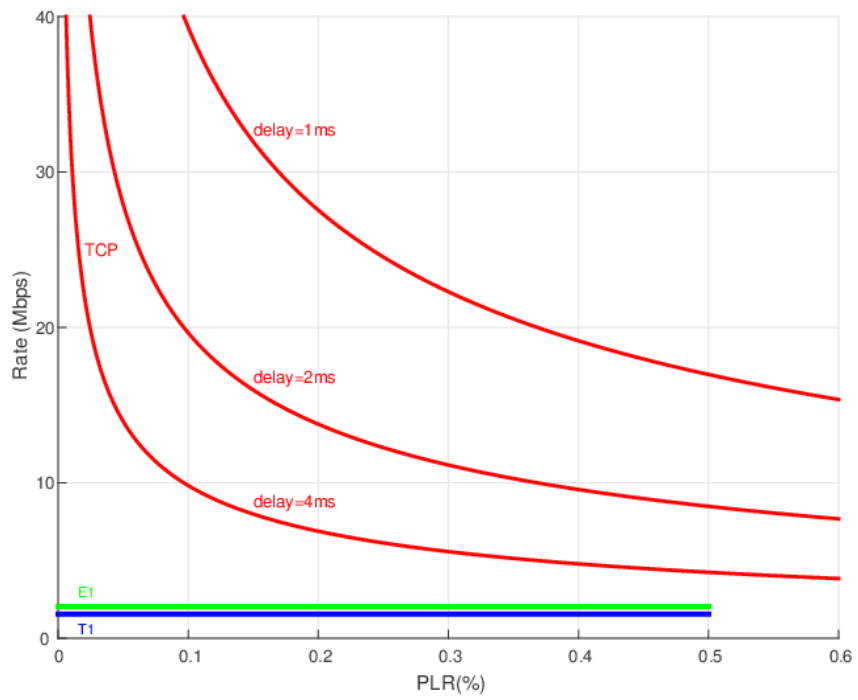Figure 1: E1/T1 PWs vs. TCP for Segment Size 64B
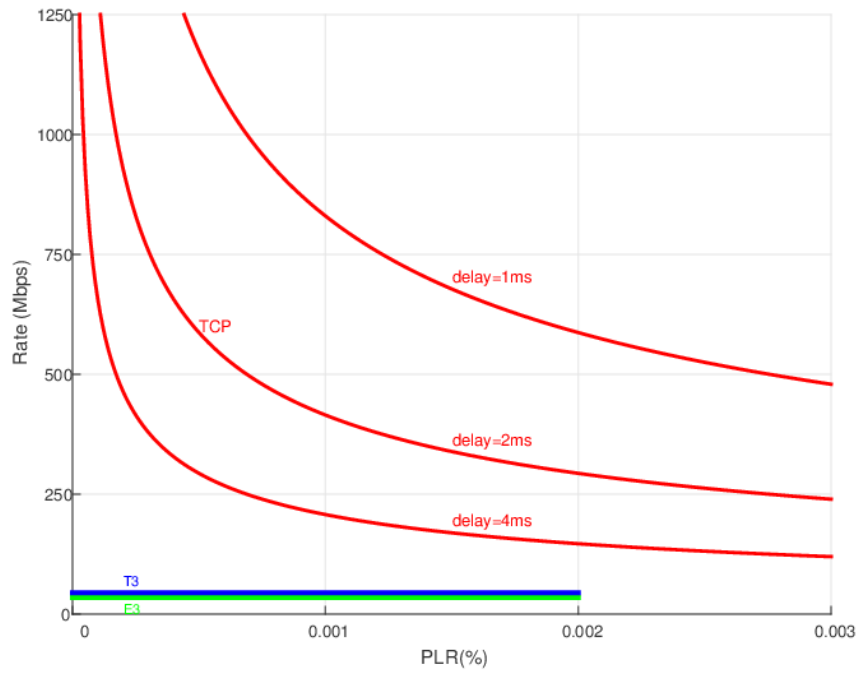
Figure 2: E1/T1 PWs vs. TCP for Segment Size 256B

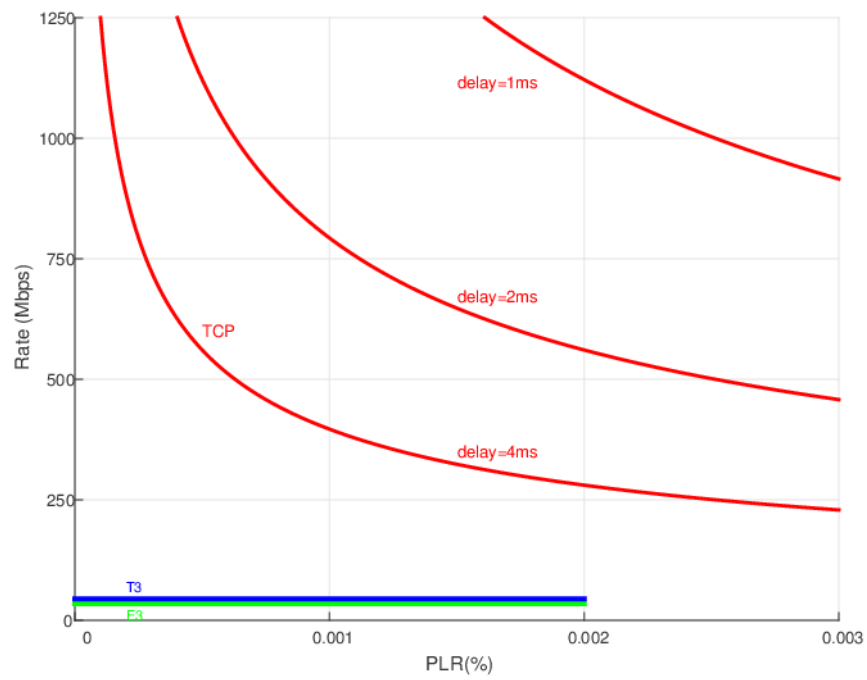Figure 3: E3/T3 PWs vs. TCP for Segment Size 536B

Figure 4: E3/T3 PWs vs. TCP for Segment Size 1024B

We can use the TCP rate equation to determine the precise conditions
under which a TDM PW consumes no more bandwidth than a TCP flow
between the same endpoints under identical conditions. Replacing the
round-trip delay with twice the one-way delay D, setting the
bandwidth to that of the TDM service BW, and the segment size to be
the TDM fragment (taking into account the PWE3 control word), we
obtain the following condition for a TDM PW:

$$D < \frac{4\ S}{BW\ f(p)}$$

where:

   D is the one-way delay,

   S is the TDM segment size (packet excluding overhead) in bytes,

   BW is the TDM service bandwidth in bits per second,

   $f(p) = \sqrt{2p/3} + 12\ \sqrt{3p/8}\ p\ (1+32p^2)$.

One may view this condition as defining a "friendly" operating
envelope for a TDM PW, as a TDM PW that occupies no more bandwidth
than a TCP flow causes no more congestion than that TCP flow. Under
this condition, it is acceptable to place the TDM PW alongside
congestion-responsive traffic such as TCP. On the other hand, were
the TDM PW to consume significantly more bandwidth than a TCP flow,
it could contribute disproportionately to congestion, and its mixture
with congestion-responsive traffic might be inappropriate. Note that
we are sidestepping any debate over the validity of the TCP-
friendliness concept and merely saying that there can be no question
that a TDM PW is acceptable if it causes no more congestion than a
single TCP flow.

We derived this condition assuming steady-state conditions, and thus
two caveats are in order. First, the condition does not specify how
to treat a TDM PW that initially satisfies the condition, but is then
faced with a deteriorating network environment. In such cases, one
additionally needs to analyze the reaction times of the responsive
flows to congestion events. Second, the derivation assumed that the
TDM PW was competing with long-lived TCP flows, because under this
assumption it was straightforward to obtain a quantitative comparison
with something widely considered to offer a safe response to
congestion. Short-lived TCP flows may find themselves disadvantaged
as compared to a long-lived TDM PW satisfying the above condition.

We see in Figures 5 and 6 that TDM pseudowires carrying T1 or E1
native services satisfy the condition for all parameters of interest
for large packet sizes (e.g., S=512 bytes of TDM data). For the
SAToP default of 256 bytes, as long as the one-way delay is less than
10 milliseconds, the loss probability can exceed 0.3 or 0.6 percent.
For packets containing 128 or 64 bytes, the constraints are more
troublesome, but there are still parameter ranges where the TDM PW
consumes less than a TCP flow under similar conditions. Similarly,
Figures 7 and 8 demonstrate that E3 and T3 native services with the
SAToP default of 1024 bytes of TDM per packet satisfy the condition
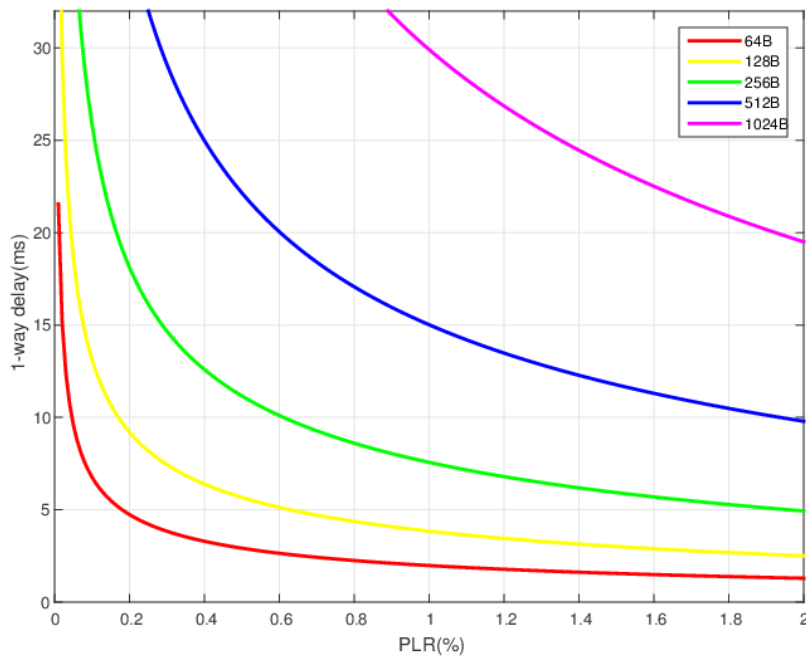for a broad spectrum of delays and PLRs.

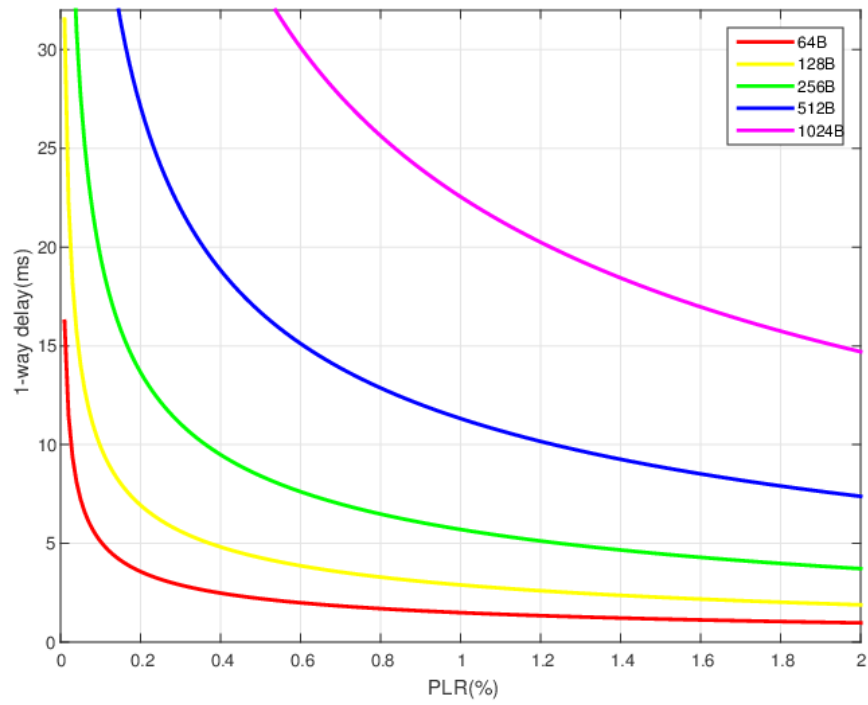Figure 5: TCP Compatibility Areas for T1 SAToP

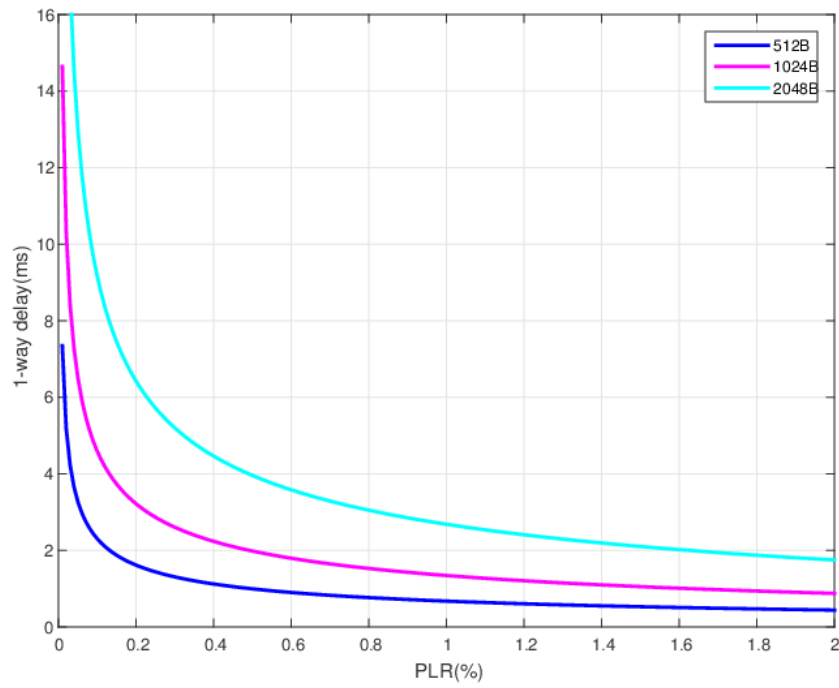Figure 6: TCP Compatibility Areas for E1 SAToP

Figure 7: TCP Compatibility Areas for E3 SAToP
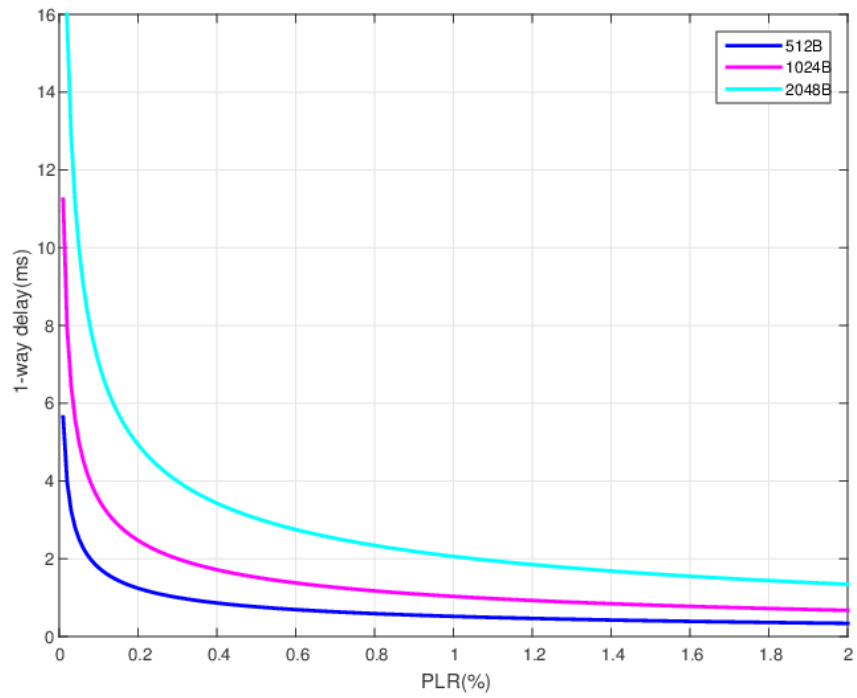
Figure 8: TCP Compatibility Areas for T3 SAToP

5. Conclusions

   The figures presented in the previous section demonstrate that TDM
   service quality degradation generally occurs before the TDM PW would
   consume more bandwidth than a comparable TCP flow. Thus, while TDM
   PWs are unable to respond to congestion in a TCP-like fashion, TDM
   PWs that are able to deliver acceptable TDM service do not contribute
   to congestion significantly more than a TCP flow.

   Combined with our earlier determination that Ethernet PWs
   automatically respond in a TCP-like fashion (see Section 3), our
   final conclusion is that PW-specific congestion-avoidance mechanisms
   are generally not required. This is true even for TDM PWs, assuming
   that the TDM management plane initiates service shutdown when service
   parameters are persistently below levels required by the relevant TDM
   standards. If the TDM service does not automatically shut down, a
   mechanism to block persistently unacceptable TDM pseudowires is
   required, or a transport circuit breaker [CIRCUIT-BREAKER] may be
   triggered as a last resort.

6. Security Considerations

   This document does not introduce any new congestion-specific
   mechanisms and thus does not introduce any new security
   considerations above those present for PWs in general.

7. Informative References

   [CIRCUIT-BREAKER]
              Fairhurst, G., "Network Transport Circuit Breakers", Work
              in Progress, draft-ietf-tsvwg-circuit-breaker-15, April
              2016.

   [G703]   ITU-T, "Physical/electrical characteristics of
              hierarchical digital interfaces", ITU Recommendation
              G.703, April 2016.

   [G775]   ITU-T, "Loss of Signal (LOS), Alarm Indication Signal
              (AIS) and Remote Defect Indication (RDI) defect detection
              and clearance criteria for PDH signals",
              ITU Recommendation G.775, October 1998.

   [G826]   ITU-T, "Error Performance Parameters and Objectives for
              International Constant Bit Rate Digital Paths at or above
              Primary Rate", ITU Recommendation G.826, December 2002.

   [P50App1]   ITU-T, "Telephone Transmission Quality, Telephone
               Installations, Local Line Networks: Appendix 1",
               ITU-T Recommendation P.50, February 1998.

   [P800]   ITU-T, "Methods for subjective determination of
               transmission quality", ITU Recommendation P.800, June
               1998.

   [P862]   ITU-T, "Perceptual evaluation of speech quality (PESQ): An
               objective method for end-to-end speech quality assessment
               of narrow-band telephone networks and speech codecs",
               ITU Recommendation P.826, February 2001.

   [PACKET-LOSS]
               Stein, J(Y). and I. Druker, "The Effect of Packet Loss on
               Voice Quality for TDM over Pseudowires", Work in
               Progress, draft-stein-pwe3-tdm-packetloss-01, December
               2003.

   [RFC2784]   Farinacci, D., Li, T., Hanks, S., Meyer, D., and P.
               Traina, "Generic Routing Encapsulation (GRE)", RFC 2784,
               DOI 10.17487/RFC2784, March 2000,
               <http://www.rfc-editor.org/info/rfc2784>.

   [RFC2914]   Floyd, S., "Congestion Control Principles", BCP 41,
               RFC 2914, DOI 10.17487/RFC2914, September 2000,
               <http://www.rfc-editor.org/info/rfc2914>.

   [RFC3031]   Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
               Label Switching Architecture", RFC 3031,
               DOI 10.17487/RFC3031, January 2001,
               <http://www.rfc-editor.org/info/rfc3031>.

   [RFC3931]   Lau, J., Ed., Townsley, M., Ed., and I. Goyret, Ed.,
               "Layer Two Tunneling Protocol - Version 3 (L2TPv3)",
               RFC 3931, DOI 10.17487/RFC3931, March 2005,
               <http://www.rfc-editor.org/info/rfc3931>.

   [RFC3985]   Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation
               Edge-to-Edge (PWE3) Architecture", RFC 3985,
               DOI 10.17487/RFC3985, March 2005,
               <http://www.rfc-editor.org/info/rfc3985>.

   [RFC4023]   Worster, T., Rekhter, Y., and E. Rosen, Ed.,
               "Encapsulating MPLS in IP or Generic Routing Encapsulation
               (GRE)", RFC 4023, DOI 10.17487/RFC4023, March 2005,
               <http://www.rfc-editor.org/info/rfc4023>.

   [RFC4553] Vainshtein, A., Ed. and YJ. Stein, Ed., "Structure-
              Agnostic Time Division Multiplexing (TDM) over Packet
              (SAToP)", RFC 4553, DOI 10.17487/RFC4553, June 2006,
              <http://www.rfc-editor.org/info/rfc4553>.

   [RFC5033] Floyd, S. and M. Allman, "Specifying New Congestion
              Control Algorithms", BCP 133, RFC 5033,
              DOI 10.17487/RFC5033, August 2007,
              <http://www.rfc-editor.org/info/rfc5033>.

   [RFC5086] Vainshtein, A., Ed., Sasson, I., Metz, E., Frost, T., and
              P. Pate, "Structure-Aware Time Division Multiplexed (TDM)
              Circuit Emulation Service over Packet Switched Network
              (CESoPSN)", RFC 5086, DOI 10.17487/RFC5086, December 2007,
              <http://www.rfc-editor.org/info/rfc5086>.

   [RFC5087] Stein, Y(J)., Shashoua, R., Insler, R., and M. Anavi,
              "Time Division Multiplexing over IP (TDMoIP)", RFC 5087,
              DOI 10.17487/RFC5087, December 2007,
              <http://www.rfc-editor.org/info/rfc5087>.

   [RFC5348] Floyd, S., Handley, M., Padhye, J., and J. Widmer, "TCP
              Friendly Rate Control (TFRC): Protocol Specification",
              RFC 5348, DOI 10.17487/RFC5348, September 2008,
              <http://www.rfc-editor.org/info/rfc5348>.

Appendix A. Loss Probabilities for TDM PWs

   ITU-T Recommendation G.826 [G826] specifies limits on the Errored
   Second Ratio (ESR) and the Severely Errored Second Ratio (SESR). For
   our purposes, we will simplify the definitions and understand an
   Errored Second (ES) to be a second of time during which a TDM bit
   error occurred or a defect indication was detected. A Severely
   Errored Second (SES) is an ES second during which the Bit Error Rate
   (BER) exceeded one in one thousand (10^{-3}). Note that if the error
   condition AIS was detected according to the criteria of ITU-T
   Recommendation G.775 [G775], an SES was considered to have occurred.
   The respective ratios are the fraction of ES or SES to the total
   number of seconds in the measurement interval.

   All TDM signals run at 8000 frames per second (higher rate TDM
   signals have longer frames). So, assuming an integer number of TDM
   frames per TDM PW packet, the number of packets per second is given
   by packets per second = 8000 / (frames per packet). Prevalent cases
   are 1, 2, 4, and 8 frames per packet, translating to 8000, 4000,
   2000, and 1000 packets per second, respectively.

   For both E1 and T1 TDM circuits, G.826 allows an ESR of 4% (0.04),
   and an SESR of 0.2% (0.002). For E3 and T3, the ESR must be no more
   than 7.5% (0.075), while the SESR is unchanged. Focusing on E1
   circuits, the ESR of 4% translates (assuming the worst case of
   isolated exactly periodic packet loss) to a packet loss event no more
   than every 25 seconds. However, once a packet is lost, another
   packet lost in the same second doesn't change the ESR, although it
   may contribute to the ES becoming an SES. Thus for 1, 2, 4, and 8
   frames per packet, the maximum allowed packet loss probability is
   0.0005%, 0.001%, 0.002%, and 0.004% respectively.

   These extremely low allowed packet loss probabilities are only for
   the worst case scenario. With tail-drop buffers, when packet loss is
   above 0.001%, it is likely that loss bursts will occur. If the lost
   packets are sufficiently close together (we ignore the precise
   details here), then the permitted packet loss ratio increases by the
   appropriate factor, without G.826 being cognizant of any change.
   Hence, the worst-case analysis is expected to be extremely
   pessimistic for real networks. Next, we will consider the opposite
   extreme and assume that all packet loss events are in periodic loss
   bursts. In order to minimize the ESR, we will assume that the burst
   lasts no more than one second, and so we can afford to lose in each
   burst no more than the number of packets transmitted in one second.
   As long as such one-second bursts do not exceed four percent of the
   time, we still maintain the allowable ESR. Hence, the maximum

permissible packet loss ratio is 4%. Of course, this estimate is
extremely optimistic, and furthermore does not take into
consideration the SESR criteria.

As previously explained, an SES is declared whenever AIS is detected.
There is a major difference between structure-aware and structure-
agnostic transport in this regards. When a packet is lost, SAToP
outputs an "all-ones" pattern to the TDM circuit, which is
interpreted as AIS according to G.775 [G775]. For E1 circuits, G.775
specifies that AIS is detected when four consecutive TDM frames have
no more than 2 alternations. This means that if a PW packet or
consecutive packets containing at least four frames are lost, and
four or more frames of "all-ones" output to the TDM circuit, an SES
will be declared. Thus burst packet loss, or packets containing a
large number of TDM frames, lead SAToP to cause high SESR, which is
20 times more restricted than ESR. On the other hand, since
structure-aware transport regenerates the correct frame alignment
pattern, even when the corresponding packet has been lost, packet
loss will not cause declaration of SES. This is the main reason that
SAToP is much more vulnerable to packet loss than the structure-aware
methods.

For realistic networks, the maximum allowed packet loss for SAToP
will be intermediate between the extremely pessimistic estimates and
the extremely optimistic ones. In order to numerically gauge the
situation, we have modeled the network as a four-state Markov model,
(corresponding to a successfully received packet, a packet received
within a loss burst, a packet lost within a burst, and a packet lost
when not within a burst). This model is an extension of the widely
used Gilbert model. We set the transition probabilities in order to
roughly correspond to anecdotal evidence, namely low background
isolated packet loss, and infrequent bursts wherein most packets are
lost. Such simulation shows that up to 0.5% average packet loss may
occur and the recovered TDM still conforms to the G.826 ESR and SESR
criteria.

Appendix B. Effect of Packet Loss on Voice Quality for Structure-Aware
             TDM PWs

Packet loss in voice traffic causes audio artifacts such as choppy,
annoying, or even unintelligible speech. The precise effect of
packet loss on voice quality has been the subject of detailed study
in the Voice over IP (VoIP) community, but VoIP results are not
directly applicable to TDM PWs. This is because VoIP packets
typically contain over 10 milliseconds of the speech signal, while
multichannel TDM packets may contain only a single sample, or perhaps
a very small number of samples.

The effect of packet loss on TDM PWs has been previously reported
[PACKET-LOSS]. In that study, it was assumed that each packet
carried a single sample of each TDM timeslot (although the extension
to multiple samples is relatively straightforward and does not
drastically change the results). Four sample replacement algorithms
were compared, differing in the value used to replace the lost
sample:

1. Replacing every lost sample by a preselected constant (e.g., zero
    or "AIS" insertion).

2. Replacing a lost sample by the previous sample.

3. Replacing a lost sample by linear interpolation between the
    previous and following samples.

4. Replacing the lost sample by STatistically Enhanced INterpolation
    (STEIN).

Only the first method is applicable to SAToP transport, as structure
awareness is required in order to identify the individual voice
channels. For structure-aware transport, the loss of a packet is
typically identified by the receipt of the following packet, and thus
the following sample is usually available. The last algorithm posits
the Linear-Predictive Coding (LPC) speech generation model and
derives lost samples based on available samples both before and after
each lost sample.

The four algorithms were compared in a controlled experiment in which
speech data was selected from English and American English subsets of
the ITU-T P.50 Appendix 1 corpus [P50App1] and consisted of 16
speakers, eight male and eight female. Each speaker spoke either
three or four sentences, for a total of between seven and 15 seconds.
The selected files were filtered to telephony quality using modified
IRS filtering and down-sampled to 8 kHz. Packet loss of 0, 0.25,
0.5, 0.75, 1, 2, 3, 4, and 5 percent were simulated using a uniform
random number generator (bursty packet loss was also simulated but is
not reported here). For each file, the four methods of lost sample
replacement were applied and the Mean Opinion Score (MOS) was
estimated using PESQ [P862]. Figure 9 depicts the PESQ-derived MOS
for each of the four replacement methods for packet drop
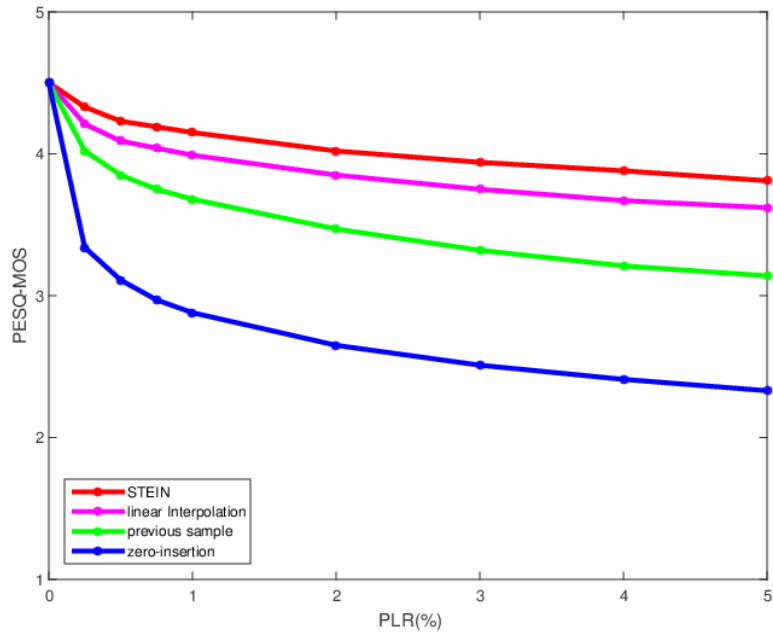probabilities up to 5%.

Figure 9: PESQ-Derived MOS as a Function of Packet-Drop Probability

For all cases, the MOS resulting from the use of zero insertion is
less than that obtained by replacing with the previous sample, which
in turn is less than that of linear interpolation, which is slightly
less than that obtained by statistical interpolation.

Unlike the artifacts that speech compression methods may produce when
subject to buffer loss, packet loss here effectively produces
additive white impulse noise. The subjective impression is that of
static noise on AM radio stations or crackling on old phonograph
records. For a given PESQ-derived MOS, this type of degradation is
more acceptable to listeners than choppiness or tones common in VoIP.

If MOS>4 (full toll quality) is required, then the following packet
drop probabilities are allowable:

    zero insertion - 0.05%

    previous sample - 0.25%

    linear interpolation - 0.75%

    STEIN - 2%

If MOS>3.75 (barely perceptible quality degradation) is acceptable,
then the following packet drop probabilities are allowable:

    zero insertion - 0.1%

    previous sample - 0.75%

    linear interpolation - 3%

    STEIN - 6.5%

If MOS>3.5 (cell phone quality) is tolerable, then the following
packet drop probabilities are allowable:

    zero insertion - 0.4%

    previous sample - 2%

    linear interpolation - 8%

    STEIN - 14%

Authors' Addresses

   Yaakov (Jonathan) Stein
   RAD Data Communications
   24 Raoul Wallenberg St., Bldg C
   Tel Aviv 69719
   Israel

   Phone: +972 (0)3 645-5389
   Email: yaakov_s@rad.com

   David L. Black
   EMC Corporation
   176 South St.
   Hopkinton, MA 69719
   United States

   Phone: +1 (508) 293-7953
   Email: david.black@emc.com

   Bob Briscoe
   BT

   Email: ietf@bobbriscoe.net
   URI: http://bobbriscoe.net/