ATTACHMENT B

TESTIMONY OF DOCTOR DAVID REED

Question 1c) How should congestion be defined in an ISP's network?

1. In CRTC Public Notice 2008-19, footnote 6, a definition of "network congestion" is given:

   > "Network congestion is broadly defined to mean a situation whereby the amount of traffic transiting the network may lead to a deterioration of service for some end users."

2. We refine this broad definition here to focus on the specific case of the Internet. The Internet itself is defined to be the "network of networks" that has resulted from voluntary interoperability among a wide variety of Autonomous Systems (AS) – networks that are not owned by each other, and which do not even have contractual obligations to each other in most cases. All it takes to be part of the Internet as an Autonomous System is to agree to participate in the core set of Internet datagram transport protocols according to the very simple ground rules of the Internet. The core ground rules were laid out in the original design begun in 1975 by Vinton Cerf and Robert Kahn. I participated in the original design of the Internet protocols, along with many other contributors, and that design has evolved through a process that is managed by rough consensus among all participants in the Internet worldwide.

3. The fundamental agreement among Autonomous Systems is that they collectively provide each host, that is each computer that is connected to any of the many Autonomous Systems, the ability to send and receive small messages called Internet datagrams to any of the other hosts on any Autonomous System in the Internet. I avoid defining a whole collection of technical terms by suggesting that you view these Internet datagrams as envelopes containing messages from one host to another on the Internet. The envelope is stamped on the outside with only four things: an address, a return address, a protocol identifier, and some marks that indicate how the message is handled as it is carried through the network. The content of each message is held "inside the envelope." This content is meaningful only to the sending and receiving hosts.

4. Each Autonomous System must agree to provide "best efforts" delivery of these Internet datagrams without reading or changing their contents – that is, a sender posts an envelope with its return address and a specified destination address, and it expects that the envelope will be routed through the network and delivered eventually to the specified address. When congestion becomes extreme in some AS, it is normal to discard messages. This is acceptable because the sender keeps a copy of each message. The sender resends that message in a new envelope until it is eventually acknowledged by the addressee.

5. Since the beginning of the Internet's design, its designers have focused on managing congestion that may arise in Autonomous Systems. From the beginning, it has been clear that the ultimate solution of the congestion problem requires that the senders causing the congestion must "slow down" their rate of sending and prioritize their traffic if need be. The network itself cannot eliminate congestion – solving the problem requires cooperation from the senders.

6. **Point 1**: Any operational definition of congestion in the Internet must begin with a definition of Queueing Delay.

7. In the Internet context, congestion manifests itself in routers or switches that forward Internet datagrams along the path between a particular source or destination. Each router or switch receives datagrams from incoming links and forwards them on an outgoing link that will move the datagram closer to delivery at its destination. Since the outgoing links may have capacity that is less than the rate of data arriving on incoming links, a queue of datagrams that must transit the outgoing link may accumulate until the incoming flow of datagrams subsides. This queue causes extra delay for each newly arriving datagram, as it waits for its turn to be sent on the outgoing link.

8. The extra delay is equal to the total size of the packets waiting in the queue, divided by the data rate of the link that must be used.

9. Congestion then occurs when the amount of data that must travel through a particular link out of a particular router exceeds the data rate of that link for a long enough period such that a queue builds up.

10. From the point of view of the hosts trying to communicate, the impact of congestion is seen as an increase in the end-to-end delay above the minimum time datagrams would take to be delivered when the network is completely empty of competing traffic. It is conventional to call the excess end-to-end delay, which results from queue build-up, Queueing Delay.

11. Queueing Delay builds up during bursts of traffic from one or more users, and then gradually goes away when the users applications slow down or go away. It is important to realize that when multiple users are communicating, not only does the available capacity get shared among multiple users, reducing individual shares, the real problem is that Queueing Delay accumulates, ultimately disrupting the network.

12. Most routers and switches provide a capability for measuring Queueing Delay on each outbound link using the Simple Network Management Protocol (SNMP). In addition, an operator can use statistical sampling by injecting probe traffic onto network paths to sample actual congestion experience on the various routes in the network. Packet pair probes are a reasonable technique to sample apparent congestion. Such techniques can provide reliable measures of actual Queueing Delays with minimal overhead.

13. **Point 2**: Queueing Delay disrupts Internet use only when it exceeds a limit, the limit is application dependent, and the disruptive impact is a result of the combined effect of Queueing Delays in each AS through which a connection travels.

14. Since Internet applications span a wide spectrum, from use of the WWW pages in a browser, to transmitting isochronous media such as live voice and video conferencing, to variable rate media such as video recordings that can be streamed individually, to file downloads and email delivery, the impact of Queueing Delay between source and destination varies. Most applications tolerate end-to-end queuing delays that are less than 200 milliseconds quite well, and many will tolerate even longer delays. Highly interactive applications such as Voice over IP conversations and conferences and interactive videogames have more demanding requirements, and typically work well only when the end-to-end queuing delay is kept below 100 milliseconds.

15. Each queue that builds up along a path introduces Queueing Delay, and the total delay end-to-end is the sum of the individual Queueing Delays. Since traffic for many destinations shares any particular queue, congestion typically arises from unpredictable and highly time-varying interactions among users.

16. **Point 3**: The normal traffic in the Internet is very bursty, which causes Queueing Delay even when average traffic demand is below the full capacity of any bottleneck link. Thus, individual links in Internet AS's must be operated at an average capacity significntly below their peak capacity, if congestion is to be avoided. Therefore, traffic loads on links or between AS's are a relatively poor indicator of disruptive congestion. A more proper measure of congestion would be average Queueing Delay divided by bitrate.

17. Bursty traffic is a fundamental property of the Internet. It is quite different from many other familiar networks, such as the voice network and video distribution networks. Under such conditions, Queueing Delay is a very non-linear function of traffic load. In addition, the burstiness of Internet traffic on bottleneck links between any two hosts on the Internet does not benefit much, if at all, from statistical smoothing.

18. As the **average** traffic volume entering a bottleneck link increases close to the capacity of the outgoing link, the **average** queue length becomes **infinite** (this is a fundamental result from mathematical queueing theory, e.g. Little's Lemma). In many situations, even when the average traffic is less than half of the total capacity on a bottleneck link, the Queueing Delay will be unacceptably long for any but the least demanding applications. Ultimately, the variance of the traffic through the link is as important a cause of Queueing Delay as is the total average traffic.

19. **Point 4**: Control of congestion within the Internet depends on preventing the buildup of Queueing Delay, and temporary reduction of inflow into queues that form. That requires responsiveness to *both* the total traffic through any one link and the variability of that traffic from the average load. Since the total traffic originates from chance correlation of diverse and widespread sources, managing traffic requires quick decisions and feedback

to each source and destination, so that source and destination can respond in an application dependent manner to reduce traffic.

20. When any particular queue starts to build up to an unacceptable level, the only way to deal with that problem is to cause some or all of the hosts sending data through the link to slow down or to stop sending Internet datagrams that must pass through that link. Also, some of the traffic can be rerouted through a less congested path to each individual destination. The sources must be informed quickly, since there may be a long stream of datagrams already in the network that will cause the queue to build even more. To be effective in eliminating congestion, the slowdown of traffic into a congested link must persist for some time, in order to allow the queue to drain fully. After that slowdown the traffic rate can be increased again, gradually, until a queue starts to build.

21. The sources of traffic causing congestion may or may not be on the same ISP, and in fact may be quite far from the congested link. If there is congestion along the path from the point of congestion to the source of the traffic, the signal to "slow down" takes longer, and this exacerbates the buildup of congestion while the congested link waits for slowdown.

22. So to summarize the points made in answer to this question, congestion in an ISP is best defined (1) in terms of average Queueing Delay (in seconds) caused by the properties of all competing traffic sharing common links, and (2) by how quickly a congested link can signal enough of the hosts whose traffic causes the congestion to reduce their sending rate for a period of time needed to drain the queue of its current data and all of the data already "in flight" from the source host.

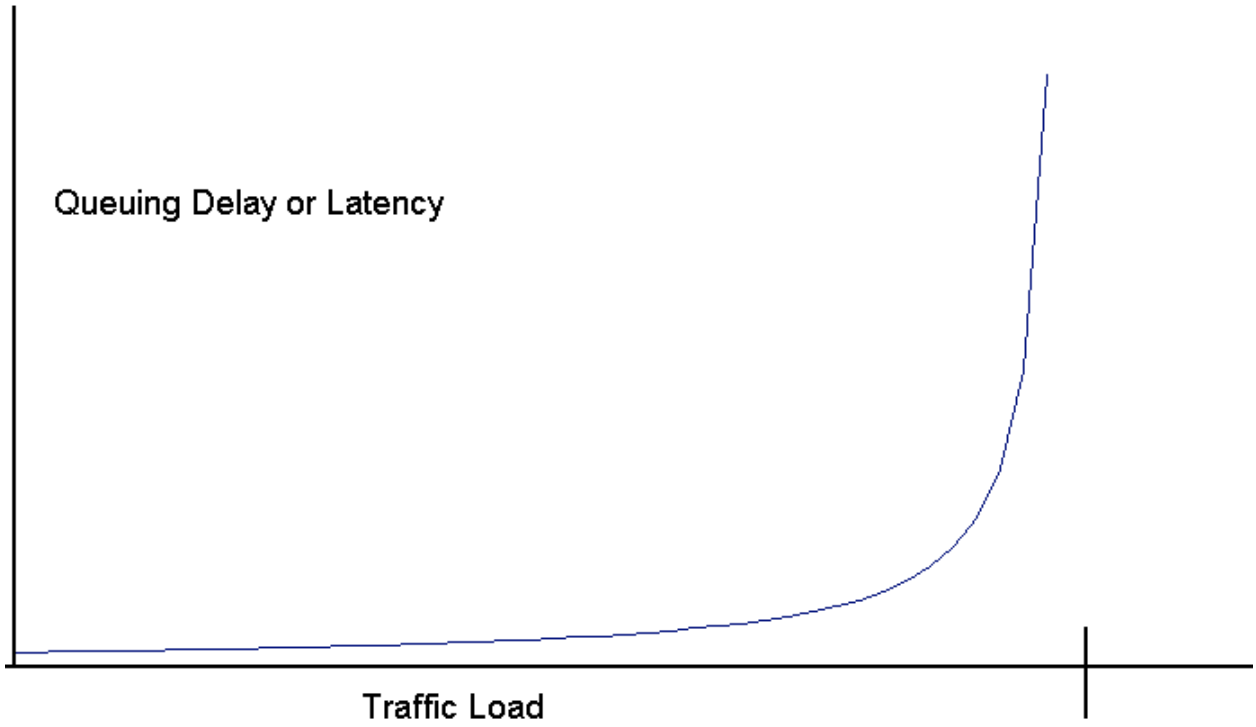
23. One should note that a very simple way to avoid complex congestion management is to make sure that the capacity of individual links is significantly larger than the peak average traffic of all users. Clearly building in too much overcapacity is costly, but attempting to operate links at nearly full capacity will ensure that unacceptable congestion is constant.

24. In their response to the Commission's interrogatories, Bell stated that

> Bell Wireline and Bell Aliant Atlantic measure latency and dropped packets (referred to as "cells" in the case of an ATM network) as well as the level of utilization of links in the network. Although both measurements are important, it is the level of utilization of links that is used as the primary criteria to determine where the congestion is occurring. Such an approach is more efficient since utilization level measurements are more consistently and readily available than latency measurements at the level of detail that is required to make provisioning decisions. The approach is also reliable given the observed close relationship between the measured utilization levels and the latency (the higher the utilization in a link, the higher the latency).

25. There is a relation between utilization rate and Queueing Delay, but it is a problematic relationship: if you look at an M/M/1 queueing model, the latency, L in seconds/bit, is L

= 1/(mu-lambda), where mu is the link rate (bits/sec) and lambda is the arrival rate (bits/sec). So the relation isn't proportional or linear. As the traffic (arrival rate) grows closer to the link capacity (link rate), the latency goes to infinity, with the curve looking like this:



So the difference between acceptable delay and unreasonable congestion occurs in a very small range of traffic load, which creates a very large change in Queuing Delay or congestion. Looking only at "percentage of full capacity" distorts what users perceive.

26. I conclude that ISP claims of congestion are not adequately backed up, at least not with data made public. It is important that regulators be able to assess all claims of network congestion. The best way to do so is to assess a metric of average Queueing Delay. Average Queueing Delay should be measured as a function of time of day between 100 random high-speed internet customers and the various internal and external switching centers, with Queueing Delay being the difference between total delay and the minimum delay achieved on that path on an unloaded network. This measurement would illustrate or disprove claims of unmanageable congestion.

27. The mention by Bell Wireline and Bell Aliant Atlantic of ATM statistics is of some concern regarding proper measurement and network management in an Internet context. ATM measurement statistics are even less directly related to experienced congestion than are link capacity measurements. Internet congestion effects on user experience cannot be characterized by underlying ATM measurements of traffic.

Question 2a) What technologies could be employed by ISPs to manage Internet traffic?

28. The primary way that Internet traffic load is managed is as follows. Each connection between a source host and destination host is responsible for monitoring changes in end-to-end Queueing Delay, and when that delay increases, to reduce the rate at which data is sent until the end-to-end Queueing Delay will come back to near zero. The primary example, and by far the most common case, is used in the standard end-to-end protocol called TCP (the Internet Transmission Control Protocol, where "control" means controlling the rate at which data is transmitted from the source host to respond to the ability of the network to transmit data and the destination to process it). This mechanism is an "end-to-end" control system that operates at the edge of the network, outside of the ISP's control.

29. An ISP's primary responsibility with regard to congestion management in the Internet is to provide the proper and clear signals to the sources of traffic, so that the sources can reduce their traffic rate by whatever application specific means may be appropriate. We'll discuss those signals shortly. In addition, the ISP can restrict/reduce data rates on any individual link within, into, or out of its network, and alter the routing of traffic through its portion of the Internet to balance the traffic among multiple paths that can be used between any particular source host and destination host. Finally, the ISP can prioritize traffic on any particular link, or among particular routes according to well-defined standard priority classes, which have been defined by the IETF, carried in the IP header TOS field (this last is not widely implemented at this point in time).

30. The default (and most common) way for a congested queue to signal congestion is to begin discarding incoming Internet datagrams ("**queue drops**" – packets in the queue are dropped after the queue gets to be a certain length). This technique, though counterintuitive, was chosen as the default for sound reasons, perhaps most importantly because it helps mitigate the damage of "uncontrolled" sources. But it's important to realize that the primary reason for dropping data is to signal the source to slow down, which works as follows. When the destination host detects that an expected packet is missing, it stops acknowledging receipt of data to the source. When the source host stops seeing acknowledgement from the destination, it starts resending data from the point where the drop occurred at a slower rate (typically half of the rate it was sending). The effect at the point of congestion is that a large percentage of high rate traffic from all sources is discarded, and the same traffic then is resent at a much lower rate.

31. In addition to this default mechanism, there are a wide range of standards that allow ISPs to manage and prioritize Internet traffic, including diffserv, ECN, RED, flow-based routing, and traffic smoothing:

    a. **Diffserv** (differentiated service labels) is a labeling or marking of packets that allows the endpoints to specify which packets or flows can tolerate delays or reduced priority for capacity. This allows users to fine tune their needs so that ISPs can be “fair” in a way that is informed by the application needs. Essentially,

the endpoints are saying, “if there is congestion, these packets should be treated as more timely or less timely than others when one must decide which ones to drop or mark to signal congestion” (similar to First Class, Second Class, Air, or Ground in the postal system).

b. **ECN (Early Congestion Notification)** is a standard method for marking envelopes that pass through congested regions of the network, so that the endpoints can decide to slow down their traffic without discarding the traffic.

c. **RED (Random Early Drops)** is a standard method for AS's to signal congested conditions by randomly discarding packets before a queue builds up, which has the effect of signalling the endpoints to slow down as in the default method. It provides a more granular control that smoothes end-to-end delivery.

d. **Flow-based routing** refers to rerouting flows in order to rebalance load when alternative paths are available to the destination. This usually is appropriate for more persistent or stable levels of congestion.

e. **Traffic smoothing** (also sometimes called **packet grooming**) refers to spacing out datagrams arriving close together so that they are less bursty due to short-term timing variation, smoothing variance-caused congestion. This involves low-pass filtering of rate. These are rather broad terms and almost slang in nature, being descriptive of a wide class of vendor-supplied techniques that limit the peak rates of bursty packet flows at an ingress point. They can reduce congestion as measured by Queueing Delays experienced by competing flows. Neither is specifically IETF-approved, since neither is a well-defined term, but neither are they viewed by anyone at IETF as problematic when applied to all packets arriving on a physical link – in other words, blind to source, destination, content, and application.

32. Queue Drops (packet drops), ECN, RED are ways to signal to endpoints that congestion is imminent which causes them to reduce their usage quickly. These are the quick ways to reduce congestion.

33. All of these methods have been developed by the Internet research community, analyzed, simulated, and are standards available for use today for ISP congestion. There is no order of preference among these options specified by any group or consensus. The engineers at an operator are expected to use good engineering judgement. These are accepted traffic management approaches that should be used before other more interventionist approaches are adopted.

34. In addition, some companies have gone beyond the standards and are delivering systems that attempt to prioritize traffic by inspecting packet content, and trying to "guess" the type of traffic. These are called **Deep Packet Inspection** techniques. They do not comply with IETF standards, and

35. One well known one (apparently used by Comcast in the US) was using Deep Packet Inspection to inject TCP RST messages to the end hosts indicating that the other host was terminating the connection (misleading each end host, causing the connection to drop).

36. Neither Deep Packet Inspection nor RST Injection are standards, and are not acceptable behavior by Autonomous Systems in the Internet, for a simple reason: they each violate the expectation that the contents of the envelopes are untouched inside and between Autonomous Systems, and delivered by best efforts. TCP RST injection is problematic because it is supposed to mean "the other end of the connection has failed" dropping packets/datagrams when there is no actual congestion.

37. The only recorded IETF discussion I am aware of that discusses RST Injection is a paper by a respected Internet expert, Sally Floyd, which strongly rejects the notion that using RST's for congestion control is a good design. [Sally Floyd," Inappropriate TCP Resets Considered Harmful," Internet RFC 3360 (Aug. 2002) <http://www.ietf.org/rfc/rfc3360.txt?number=3360>].I observe that in a number of documents filed in this proceeding, parties have identified DPI's interest in the "application header. For example, The Companies (CRTC)4Dec08-8 PN 2008-19 Abridged, at p. 15 states "the term "Deep Packet Inspection" refers to its ability to look "deep" into a packet, examining application headers for unique signatures, in order to classify application traffic.

38. See also the report from Professor Mark Coates for Union Consommateurs, filed February 22, 2009, at paras. 12-13.

39. Applications can have headers, along with other things. But in the Internet architecture, the transport protocol is not capable of knowing what they are, because the architecture states and requires that the packets being transported from source host to destination host carry bits, and the meaning of those bits, is defined *entirely* by an agreement between the source and destination host.

40. The characterization of the "application header" as distinct from other parts of the payload is suspect, implying that it is acceptable to impute a meaning to bits in the content part of packets by intermediaries that have no reason to know what that meaning is. It is required that end hosts are the only ones who definitively know what bits mean for a very important and fundamental technical reason, which was laid out in the beginning of the Internet: so that unanticipated new protocols can be deployed without requiring any changes in the underlying packet delivery infrastructure.

41. Deep Packet Inspection and RST Injection are sometimes characterized as "traffic shaping". This is a general term that is not well-defined in the Internet engineering community. However, many traffic shaping initiatives share the following characteristics that are clearly a problem:

- they selectively or arbitrarily restrict traffic associated with certain sources, destinations, content types, applications or protocols,

- they are based on information not provided by the customer host or application for use in shaping (in other words, they look inside the content, rather than at the envelope), and

- they do not follow the standard mechanisms for signalling congestion.

I would characterize traffic shaping initiatives that share these characteristics as "traffic interference".

42. Traffic *management* is justifiable in the presence of congestion on the network (the fact of congestion borne out by excessive Queuing Delays predicated on acceptable oversubscription ratios). Traffic *interference* – invasive practices that interfere with end user traffic – is not necessary for ISPs to manage network congestion. Traffic interference should be permissible only transparently, as a last resort, where finely tailored to target congestion, and where implemented in a manner that minimally impairs the user experience.

Question 2b) What developments are under way with respect to traffic protocols and or application changes, which could assist in addressing network congestion?

43. Congestion control techniques can only work well if they are standardized across the entire Internet, for the following reasons. Reduction of congestion requires that there be a predictable response by the source host and applications to well understood "signals" of congestion.

44. New techniques are introduced carefully, typically orchestrated in the Internet Engineering Task Force, which is a collection of engineers and researchers who resolve these issues Internet-wide, independent of the vendors and operators, but taking their needs seriously. Today's standard congestion control techniques involve mechanisms for detecting and notifying endpoints of congestion – the province of the message-switching elements of the Internet – and mechanisms to translate detections into action.

45. Responsibility for indicating priority and slowing down traffic is part of the standard end-to-end protocols, in particular TCP. TCP responds to such notification by rapidly slowing down its transmission. All file transfers, including BitTorrent, use TCP, so when congestion is detected, the senders slow down.

46. Active research is focusing less on new congestion control methods, and more on achieving better notions of "fairness". Fairness refers to how the degree of slowdown is allocated among distinct end-to-end flows on the network.

47. To achieve "fairness" requires a good definition of what it means to be fair among different kinds of applications and over different time frames. For example, should one

try to achieve fairness during every single second of operation, or fairness averaged over an hour, a day, or a month? Since users needs are bursty, this is not an easy problem.

48. Similarly, fairness in use of a particular congested link may not be appropriate, since applications and web pages are accessing the network over a diverse set of links that vary over time. If it is the only time a particular user is using a particular link that month, one could argue a kind of fairness that gives that user a very high share of that link for a few minutes compared to the users who use it 24 hours per day, every day of the month.

49. One area of active investigation into traffic management improvements is the use of flow labelling by endpoints to signal the likely resource requirements of individual end-to-end flows. Techniques such as RSVP have not been widely deployed, by network operators, but new approaches such as the TIA-1039 standard are being explored to extend the expressiveness of labelling that can be used by ISPs to decide how to manage multiple competing traffic flows when congestion arises.

Question 2c) What are the specific capabilities offered by the technical solutions identified in (a) and (b) above?

50. Restricting flow rates on links into the ISP from hosts or other networks: reducing the total traffic load relative to capability of the network to carry it.

51. Rerouting flows to balance across alternate paths (traffic engineering): better use of internal resources. Also, if an alternate path through another AS is available, reducing load on this ISP vs. other AS's.

52. Queue Drops (to control the default end-to-end congestion): signaling congestion to the endpoints, ameliorating effect of unresponsive traffic sources.

53. Diffserv: allowing endpoints to indicate the desired prioritization of data flows, thus facilitating a choice of packets to delay.

54. RED: early signaling of congestion, facilitating more rapid and stable responses by the source.

55. ECN: early signaling of congestion, facilitating more rapid and stable responses by the sources, working better with time-sensitive flows than RED.

56. Packet grooming: eliminating unnecessary bursts of traffic that would cause unnecessary short-term congestion.

57. Deep Packet Inspection (non-standard, invasive): claimed benefit is that by reading content of packets, can infer proper priorities. Other benefits claimed have to do with eliminating unwanted traffic such as "spam", viruses, and "copyright infringing" content. Actual benefit depends on reliability of inference of content type and application requirements.

58. RST Injection (non-standard): claimed benefit: management of aggressive sources that will not respond to ordinary standard mechanisms. Likely disruptive to many standard applications, and depends on the reliability of Deep Packet Inspection.

Question 2d) With reference to questions (a) and (c) above, how effective would these solutions be in addressing network congestion in the ISP networks?

59. All of the standard solutions already provide very effective tools that can dramatically change congestion. In particular, traffic management, diffserv and ECN, which are not widely deployed today, can have significant effects if hosts make use of them as designed.

60. In addition, adding more link capacity whereever frequently congested links are discovered so that the peak "average loads" are well under the point where queueing delay starts to build up toward unacceptable and disruptive service degradations as experienced by users. The point where this occurs is best determined by measures of queueing delay, rather than a set percentage of capacity.

61. Deep Packet Inspection for prioritization has no obvious benefit, and being non-standard, have not had widespread review or evaluation. The basis by which DPI "understands" which flows are more important is unreliable, non-standard, and invasive, depending on guessing characteristics of application data, or violation of end user privacy.

62. Unreliable and non-standard: The Internet Protocol, IP, is defined as an “envelope” that contains data whose meaning is an agreement between the hosts on both ends of the conversation. A device that inspects the application data “inside the envelope” is making a guess about how the bits are to be interpreted. For example, if the data seems to contain a sequence of numbers that look like a sound encoded in one of the standard telephony encodings, that packet may or may not need timely delivery. It could be a transmission of an archive recording of a historical podcast, or a timely conference call among emergency responders, or even a stream of compressed textual or image data. The *content bits* are only meaningful to the end systems, which have encoders and decoders that they agree to use for that flow. Guessing the content wrongly can result in mishandling – suppressing important flows, and promoting unimportant ones. For this reason, IP is designed so that the end hosts define the response to congestion signals, as the primary way that congestion is managed. As a secondary technique, IP includes mechanisms for flow labeling, such as RSVP and the proposed TIA 1039.

63. Invasive: It has also been part of the design of the Internet that end user traffic need not be exposed to the operator for normal operations, because of privacy and “common carriage” concerns that are well understood in the communications industry. More and more traffic is being encrypted on an end-to-end basis because of concerns that traffic may pass through intermediaries that cannot be trusted – such as switching centers in intermediate countries unfriendly to the communicating parties or commercial intermediaries who may be tempted to exploit private information to extract passwords,

marketing data, or competitive intelligence. Such traffic may be fully encrypted using standards like SSL and TLS. In such cases it is very clear that an argument that traffic should be open to inspection, merely for the minor goal of optimizing congestion, should be avoided if there are alternatives, such as labeling of flows.

Question 2e) Also with reference to questions (a) and (c) above, what impact could the implementation of technical solutions have on the Internet Engineering Task Force Standards upon which the operation of the Internet is based? Could these solutions create interoperability challenges for application developers?

64. As noted, most of the solutions are IETF standards. Applying these would be compatible with improving the operation of the Internet to the extent that the IETF has exercised sound engineering judgement heretofore. To the extent that problems with them are discovered, the data and issues can be shared, leading to improved functioning of the Internet as a whole when adopted in other ISPs and Autonomous Systems within the Internet.

65. The non-standard solutions based on DPI and RST mentioned above have negative impacts. If each ISP implements unpredictable and secret congestion management techniques, application developers will not be able to design applications that work equally well on all parts of the Internet, and diagnosing problems seen by users will become much more difficult or impossible.