



MEDIA LAB



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Response to the Commerce Commission on the TelstraClear Bitstream Access Draft Determination

20 May 2005

This submission is the product of a collaboration between

MediaLab Limited

and

The University of Waikato

Public Version

(Note: There is no restricted version)

Matt Brown
Research Analyst
MediaLab Limited
matt@medialab.co.nz
Ph +64 9 379 0462
Mob +64 27 561 1544

Table of Contents

Table of Contents.....	2
1 Executive Summary.....	3
2 Introduction.....	4
3 Measurement Approach.....	5
3.1 General Observations.....	5
3.2 Statistical Methodology.....	6
3.3 Response to Question 7 of the Draft Determination	6
3.3.1 Contention Ratio.....	7
3.3.2 One-way measurements.....	7
3.3.3 Hybrid Measurement Approach	7
3.3.4 Packet Size	8
3.3.5 Congestion	8
3.3.6 Throughput Tests	9
3.3.7 Layer Two Measurement.....	10
3.4 Response to Question 8 of the Draft Determination	10
4 Proposed Framework.....	12
4.1 Data Collection Methodology	12
4.1.1 Measurement Network Topology	12
4.1.2 Key Service Parameter Definitions.....	13
4.1.3 Key Service Parameter Sample Metrics.....	15
4.2 Statistical Framework	15
4.2.1 Experiment Design	15
4.2.2 Confidence Intervals.....	16
4.2.3 Analysis of Identical Services.....	17
4.2.4 Testing for Non-Identical Services	17
4.2.5 Time based requirements	17

1 Executive Summary

- 1) The Commerce Commission has released a Draft Determination regarding TelstraClear's access to a regulated bitstream service provided by Telecom under the Telecommunications Act 2001.
- 2) The Draft Determination specifies that the performance of the regulated bitstream is to be monitored to ensure that consistency of service is achieved with Telecom's retail Jetstream service.
- 3) MediaLab and the WAND group have expertise in network measurement and make this submission to assist the Commerce Commission in establishing an effective measurement regime.
- 4) MediaLab and the WAND group make the following recommendations
 - a) The measurements regime should utilise round-trip tests for the key parameters of Latency, Delay Variation (Jitter), Loss and Throughput
 - b) The measurement regime should be based on active measurement techniques.
 - c) That a range of packet sizes should be included in the measurement regime.
 - d) The measurement regime should include measurements of how the services handle congestion.
 - e) Throughput tests should be conducted using the TCP protocol and must be of sufficient duration to ensure that the full available bandwidth is measured.
 - f) All measurements should be conducted at the IP layer (OSI layer three)
 - g) The IETF standards should be used instead of the ITU standards when defining the key service parameters.
- 5) The final section of this submission presents a proposed measurement framework based on the recommendations made by MediaLab and the WAND group.

2 Introduction

- 6) On 21 April 2005 the Commerce Commission released a draft determination ('the Draft Determination') on an application made by TelstraClear Limited for a determination regarding access to two designated access services under Schedule 1 of the Telecommunications Act 2001 ('the Act').
- 7) The Draft Determination provides a set of performance standards to be met by the service granted to TelstraClear in the Draft Determination and guidelines on how these standards are to be measured.
- 8) Accompanying the Draft Determination is a list of questions on the proposed performance standards and an opportunity for interested parties to provide submissions to the Commerce Commission by the 6th of May 2005. This deadline was subsequently extended to 20th May 2005.
- 9) MediaLab is a limited liability company collaborating with businesses and tertiary institutions on ICT research projects. MediaLab employs staff and contractors to manage a wide range of independent network testing, measurement and modelling projects.
- 10) The WAND Network Research Group at The University of Waikato ('the WAND Group') is a world leader in the fields of active and passive measurement.
- 11) MediaLab and the WAND Network Research Group have collaborated on a number of network measurement projects over the past three years. MediaLab and the WAND Group are uniquely positioned to provide independent network measurement advice and services to the Commerce Commission.
- 12) This submission presents comments on the proposed performance standards and introduces a recommended approach. It is not the intention of this submission to comment on the other aspects of the Draft Determination. The remainder of this document is structured as follows.
 - a) Part 3 responds to questions 7 and 8 in the Draft Determination and comments on the general measurement approach detailed in the Draft Determination.
 - b) Part 4 details the test architecture and statistical framework recommended by MediaLab and the WAND Network Research Group.

3 Measurement Approach

3.1 General Observations

- 13) MediaLab and the WAND group believe that, in general, the proposed approach of using network measurement and statistical analysis to ensure that TelstraClear is receiving an appropriate regulated bitstream product is sound. However, care needs to be taken to ensure that the appropriate techniques are used.
- 14) There appears to be some confusion between the Draft Determination and the submissions of Telecom and TelstraClear regarding the exact nature and definition of how the two services should be compared.
- 15) TelstraClear has requested that the Commission require Telecom to provide service parameter equivalence between the regulated bitstream access and the retail Jetstream product.
- 16) Telecom and the Commission appear to have avoided using the term equivalence in favour of the term 'consistency'. In particular the Draft Determination requires Telecom to "monitor key service parameters to ensure consistency of service is achieved"¹.
- 17) MediaLab and the WAND group believe the two terms to have the following meanings:
 - a) 'equivalent' describes a state where there is no statistically significant numerical difference between the key parameters of the two services.
 - b) 'consistent' describes a state where although there is a statistically significant numerical difference between the key parameters of the two services, the difference is not practically significant in terms of overall the performance experienced by end-users of the services.
- 18) The definition of what constitutes a practically significant difference for each key parameter is likely to be a highly contentious issue. MediaLab and the WAND group consider that this definition should be agreed upon between Telecom, TelstraClear and the Commerce Commission. MediaLab and the WAND group are happy to provide further information on measurement techniques that could be helpful in the definition of this term if required.
- 19) MediaLab and the WAND group agree that the key parameters of each service are Latency, Delay Variation (Jitter), Packet Loss and Throughput, however our definitions are slightly different to those given in the Draft Determination. Our understanding of these definitions is described in part 4.1.2 of this submission.
- 20) Other factors relating to the services are referred to as design parameters. These include upstream and downstream speeds and the contention ratio. Design parameters are never directly measured.
- 21) As the two services differ at an architectural level (due to the layer two tunnel used to deliver the bitstream service) it is to be expected that there will be some statistically significant numerical difference between them. For this reason MediaLab and the WAND group believe it would be unreasonable to expect to be able to prove equivalence between the two services as defined by the above definitions.
- 22) MediaLab and the WAND group recommend that the key parameters of the two services be compared to determine if they are consistent (as defined above). This is the same definition that the Commission arrives at in paragraph 244 of the Draft Determination however we believe we have defined the term consistent more precisely.
- 23) While the primary focus of the measurement is obviously to determine the consistency of the regulated bitstream and retail Jetstream services, MediaLab and the WAND group recommend that the following goals also be taken in to account when designing the

¹ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, p. 50

measurement approach. We believe these are important considerations in ensuring the resulting measurements are useful and appropriate to the situation.

- a) The measurements should be representative of the performance experienced by end users and the real world applications that they use the access link for.
- b) The measurements should be easily repeatable by third parties such as access seekers or their clients.
- c) The measurement architecture should be as simple as possible to avoid errors introduced by complex design and unnecessary cost.

3.2 Statistical Methodology

- 24) The Draft Determination provides very little documentation regarding the statistical methodology that is proposed for the analysis of the key service parameters.
- 25) MediaLab and the WAND Group have some particular concerns that in some portions of the Draft Determination and earlier submissions a tendency has been shown to equate statistical confidence levels and the proportion of time that a parameter achieves a certain standard as the same thing. This is not correct.
- 26) The 'Indicated Rate' field of the Draft Determination implies that the Commerce Commission is intending that a standards based approach be used to establish consistency. However many important questions such as whether the proposed standards are to be met by both the regulated and retail services, and the statistical methods that would be used to compare a service to the standard are not described.
- 27) Additionally, the Draft Determination does not appear to address the fact that the creation of the requested regulated bitstream service for TelstraClear will result in the creation of a bitstream service with design parameters that may not be matched by any of the retail Jetstream services.
- 28) From a statistical perspective, the optimum method to determine consistency would be to calculate a confidence interval on the differences between each key parameter of the two services under test. The test of consistency defined in point 17) would then be applied to this confidence interval to determine if the key parameters of the two services are indeed consistent. While this method is optimal in determining consistency between services, it can only be used when the design parameters of the two services are equal.
- 29) As a secondary method MediaLab and the WAND group would accept a standards based approach, such as the Draft Definition appears to define, to determine consistency between the key parameters of two services that do not have identical design parameters. Such an approach would also utilise a confidence interval, calculated on a measured sample of the key parameter under investigation. If the relevant limit of the confidence interval met the specified standard the service could be said to be consistent. This is a significantly weaker test of consistency than the method proposed in point 28) as it does not deal with large differences between two services where both services happen to meet the standard.
- 30) Further detail on the methods for determining the consistency of two services can be found in part 4.2 of this submission.

3.3 Response to Question 7 of the Draft Determination

"Do the suggested criteria appropriately measure the key service parameters necessary to assess whether the network performance of the Bitstream service is consistent with the characteristics of the Bitstream used by Telecom to supply its Jetstream services?"²

- 31) No, we do not believe the suggested criteria appropriately measure the key service parameters for the following reasons.

² Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, p. 60

3.3.1 Contention Ratio

- 32) The contention ratio is included in the table found in Appendix A of the Draft Determination as a parameter to be measured. The contention ratio is a design parameter of the network and as such is not able to be directly measured. While the effects of the contention ratio would be observed in the performance of the key parameters of each service during times of congestion there is no way to infer or calculate the current contention ratio from this data.
- 33) The contention ratio cannot be measured alongside the key performance parameters of each service as it plays a role in determining what the performance of those parameters will be. A completely separate process would best carry out monitoring of the contention ratio.

3.3.2 One-way measurements

- 34) The testing architecture proposed by the Draft Determination specifies one-way tests for the key parameters of Latency, Delay Variation, Loss and Throughput. MediaLab and the WAND group do not believe one-way measurements are necessary in the scenario created by the Draft Determination.
- 35) The collection of one-way measurements requires a complex measurement infrastructure with synchronised clocks and high precision timing. The Draft Determination does not provide sufficient detail regarding the location of the measurement points or how clocks would be synchronised.
- 36) Obtaining high precision timing data, and the associated measurement infrastructure required for one-way measurements is much more costly than round-trip measurements. The Draft Determination does not provide justification for why one-way measurements are preferable over other measures of network performance.
- 37) MediaLab and the WAND Group consider that one-way measurements are not appropriate in the scenario presented by the Draft Determination for the following reasons which relate to the goals stated in point 23).
- a) Very few real world applications are affected by the one-way delay of a link.
 - b) One of TelstraClear's stated reasons for requesting defined performance characteristics is to allow identification of whether a fault reported on a customer's line is indeed a fault, or simply caused by the variance inherent in a best effort Internet grade service. The use of one-way measurements is not helpful in meeting this request. The required infrastructure for conducting one way measurements is unlikely to be available for use on the line of an arbitrary customer who reports a fault. This will most probably lead to customer diagnostics being performed using round trip time measurements.
 - c) TCP Throughput is unable to be calculated as a one way measurement as the sender requires the receipt of acknowledgement packets from the receiver of the data. The Draft Determination does not provide detail on how throughput is to be measured in a one-way context.

Recommendation 1:

MediaLab and the WAND Group recommend that the final measurement infrastructure utilise round trip tests for the key parameters of Latency, Jitter, Loss and Throughput. This reduces complexity and cost and is representative of the performance experienced by real world applications. Additionally the measurements match the tests that can be performed by an access seeker and their clients.

3.3.3 Hybrid Measurement Approach

- 38) Network measurement techniques may be grouped into two main approaches. The first approach, active measurement, introduces traffic into the network and measures the performance experienced by this traffic. This modifies the state of the network. The second approach, passive measurement, observes the traffic passing a point in the network. The state of the network is not modified in any way. A third approach can be created from a mix of the previous two approaches, we call this third approach 'hybrid measurement'.

- 39) The measurements described in the Draft Determination imply that a hybrid measurement approach is intended with traffic generated 'at the end-user's premises and measured at the Network Connection Point'³. While this definition is vague, the implication is that an active device generates test traffic at the end-user's premises to be sent over the network and captured by a passive device at the Network Connection Point.
- 40) MediaLab and the WAND Group believe that an approach using purely active measurement is more suited to this situation than an approach that includes passive measurements.
- a) Passive measurements, while potentially more accurate, are much more complex and costly to set up than the corresponding active measurements.
 - b) The extra complexity and cost of a passive measurement architecture is not justified in the scenario created by the Draft Determination as the required results can be achieved using active measurements.
 - c) As with one-way measurements, the use of passive measurement prevents an access seeker or its clients from being able to reproduce the measurements.
 - d) TCP Throughput can not be calculated using a purely passive approach.

Recommendation 2:

MediaLab and the WAND Group recommend that the final measurement infrastructure is based on active measurement techniques.

3.3.4 Packet Size

- 41) The testing methodology specified in the Draft Determination does not specify any packet sizes other than 64 byte packets used for the latency, jitter and packet loss tests.
- 42) Packet size can have a significant impact on how a packet is treated, especially in situations where congestion or rate limiting is in effect.
- 43) 64 byte packets are not representative of many of the common packet sizes used for data transfer on the Internet and will tend to have very different performance characteristics than larger packet sizes.
- 44) The addition of extra packet sizes to the test methodology does not add significant complexity.

Recommendation 3:

MediaLab and the WAND Group recommend that the Commission specify a range of packet sizes (e.g. 64byte, 200byte, 1472byte) to be measured in the final determination.

3.3.5 Congestion

- 45) TelstraClear's submissions to the Commerce Commission indicate a strong desire for the performance of the services to be evaluated using uncongested access links.
- 46) MediaLab and the WAND Group consider the performance of the service during times of congestion to be a key factor that must be considering when comparing the characteristics of two services. If this is not considered the possibility that one service degrades more gracefully than the other cannot be discounted.
- 47) There are two sources of congestion that must be considered when evaluating DSL based services such as Jetstream and the regulated bitstream.

³ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, p. 62

- a) User generated congestion on the access link.
 - b) Network congestion introduced by the contention ratio and the behaviour of other customers
- 48) The effect of congestion on the access link can be measured by generating a stream of background traffic (also known as cross traffic) at a specified rate designed to fill, or partially fill, the rate limit buffer on the RAN. As the cause of congestion is consistent across the two services it can be expected that the performance of the service characteristics will remain consistent (although they will be degraded).
- 49) Network congestion is best measured through long running tests that measure the performance of the link during both peak and off peak periods. In practice this means that it is likely that the measurements will need to run for a number of weeks before final results can be produced.
- 50) Given the focus on demonstrating consistency between the key service parameters of the two services concerns relating to specific performance standards not being met during times of congestion are unfounded.

Recommendation 4:

MediaLab and the WAND Group recommend that the Commerce Commission include measurement of performance degradation, as congestion is experienced, in the final determination.

3.3.6 Throughput Tests

- 51) Appendix A of the Draft Determination specifies three throughput tests as the key parameters relating to throughput. These three tests cover minimum upstream throughput, and minimum and average downstream throughput.
- 52) Throughput measurements are complex and the parameters of each test must be strictly specified if the test is to be reliably repeated. The Draft Determination does not provide information on many of the parameters that would be required to achieve this. In particular, the protocol that is used to measure throughput is not specified.
- 53) MediaLab and the WAND Group consider that TCP Throughput should be used to measure these characteristics. TCP is the dominant data transport protocol in use on the Internet. Use of TCP for the throughput measurements appears to be inline with the intent of Appendix A of the Draft Determination which intends to measure the “throughput rate for data traffic”⁴.
- 54) TCP Throughput is calculated as the quantity of data that was transmitted between an application running on two end points over a period of time. The quantity of data excludes TCP and IP overheads and data that had to be retransmitted due to packet loss. TCP Throughput is sometimes referred to as ‘Goodput’.
- 55) The Draft Determination specifies that throughput is to be measured every minute with five and sixty minute averages used to measure compliance with the standards. Given that both an upstream and a downstream throughput test must be conducted each minute this leads to the conclusion that the tests must be short in duration.
- 56) MediaLab and the WAND Group have serious concerns that short throughput tests may not accurately measure the full potential that a TCP session could reach given speed of the links under test and the unspecified Round Trip Time.
- a) Taking a theoretical 2Mbps/128kbps link with no loss and large router buffers, a TCP connection will take up to 60 seconds to complete the slow start phase and enter steady state.

⁴ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom’s fixed PDN service ‘Bitstream Access’*, 21 April 2005, p. 62

- b) Real world conditions such as congestion, rate limiting and packet loss are likely to extend the amount of time needed for a TCP connection to reach steady state.
 - c) To accurately measure TCP throughput the duration of the test must be long enough that the proportion of time spent in slow start does not impact the calculated throughput.
- 57) The highly asymmetric nature of the services under tests presents interesting challenges for TCP throughput measurements. In particular, the limiting factor is likely to be congestion on the upstream path caused by the acknowledgement packets needed to sustain a high rate data flow. Care must be taken in the measurements to ensure that the line under test is free of all other traffic (including other measurement traffic) during the time period that throughput testing is conducted.

Recommendation 5:

MediaLab and the WAND group recommend that TCP throughput is used as the basis for the throughput measurements and that a series of calibration tests be conducted on each service to determine the required test length to fully measure the available TCP throughput. This test length is likely to be 3 to 4 times longer than the length implied in the Draft Determination.

3.3.7 Layer Two Measurement

- 58) The Draft Determination specifies that the key service parameters should be measured on Telecom's underlying network at OSI layer two⁵. It is not clear whether this statement is referring to the PPP connection formed by each DSL connection or the ATM / Ethernet network that supports the entire DSL platform and the uplink to the access seeker. Additionally it is not specified how measurement of the specified IP packets, which exist at OSI layer three, is intended to be performed at OSI layer two.
- 59) The assumption that measuring the common layer two network supporting both the regulated bitstream and the retail Jetstream service will provide a true indication of consistency is false. In particular, it does not account for the performance effects on IP packets of having to traverse the layer two tunnel used to provide the regulated bitstream service. It also opens the measurements to the allegation that they are not detecting differences between the two services that are introduced at a higher layer than the measurements are conducted at.

Recommendation 6:

MediaLab and the WAND Group recommend that all measurements be conducted at layer three and in particular are conducted using the IP protocol. This is the primary protocol that is delivered to end-users and is the optimum layer at which to measure consistency between the key characteristics of the UBS and Jetstream services.

3.4 Response to Question 8 of the Draft Determination

"Do the parties agree that the ITU definitions for the parameters are the appropriate definitions to use as the basis for measurement of the key parameters?"⁶

- 60) MediaLab and the WAND group believe that the ITU definitions provide good guidelines for the definitions of the key service parameters to be measured, however they are not optimal for the purposes of establishing consistency between the key service parameters of the regulated bitstream and retail Jetstream services.

⁵ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, para 246 p. 50

⁶ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, p. 60

- 61) The ITU definitions are focused on low layer measurements and are more suited for measurement of high performance links such as backbone and core network elements than for measuring an access link service such as Jetstream or the regulated bitstream.
- 62) Use of the ITU definitions do not help the goals stated in point 23) of this submission as they use one-way tests and passive measurement which has been shown to be problematic in part 3.3.
- 63) As acknowledged by Telecom during the Commission's workshop⁷ there is a similar set of performance standards published by the IETF. These standards include RFC2330, RFC2681, RFC3393, RFC3128 and RFC2678. MediaLab and the WAND group do not believe that these IETF standards conflict directly with the ITU definitions. However the IETF definitions more closely reflect the intended measurements than the ITU definitions do. The application of the IETF standards to the proposed measurement framework is covered in more detail in the following section.

Recommendation 7:

MediaLab and the WAND Group recommend the use of the IETF standards for this network measurement scenario.

⁷ Workshop transcript, 11 February 2005, Dr Milner, p. 6

4 Proposed Framework

- 64) Based on the observations and recommendations regarding the Draft Determination in part 3 of this submission, MediaLab and the WAND group would like to propose the following measurement framework.
- 65) We believe that the proposed framework provides a strong foundation to support a robust analysis of consistency between the regulated bitstream and retail Jetstream services.
- 66) The two crucial components of this framework are the data collection and the statistical methodology. The data collection component includes the definitions of the key service parameters that we believe should be measured. The following sections provide further detail on these aspects of the proposed framework.
- 67) The proposed framework is based on techniques specified in the IETF RFC2330.
- a) In particular where the terms ‘singleton’ metric, ‘sample’ metric and ‘statistical’ metric are used, their use is consistent with the definitions found in section 11 of RFC2330.

4.1 Data Collection Methodology

4.1.1 Measurement Network Topology

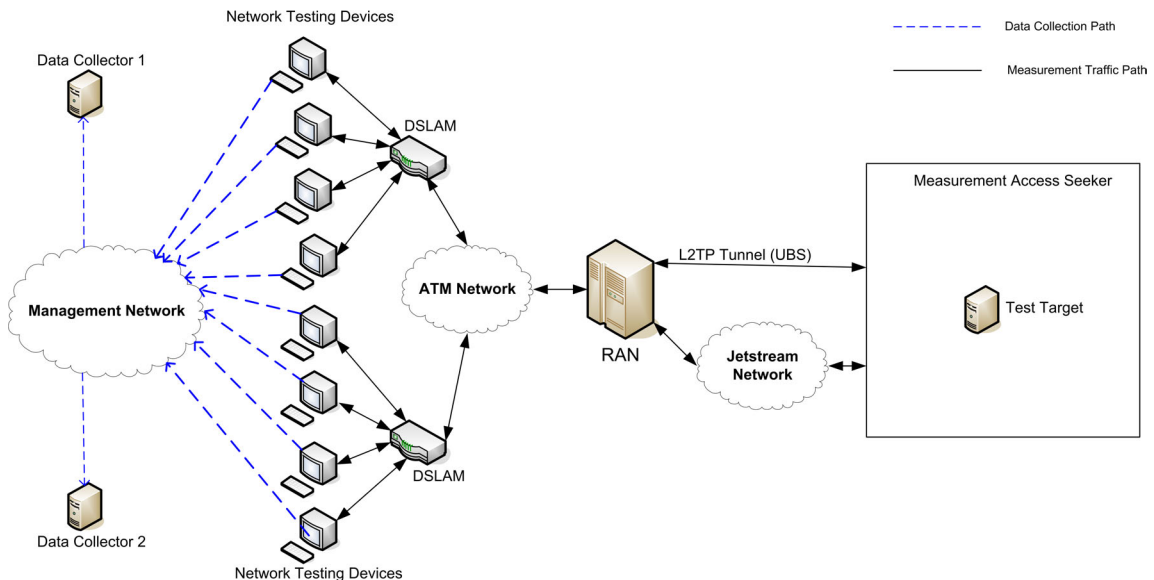


Figure 1: Proposed Measurement Network Topology

- 68) Figure 1 illustrates the desired topology of the measurement network that would be used to collect the performance results for each key service parameter. It is intended that all components shown in the diagram from the RAN leftwards may be present multiple times in the final network (i.e. Multiple DSLAMs and RANs would be used, but have been omitted from the diagram for simplicity).
- 69) The principle behind the proposed network is to recreate the environment that an access seeker is presented. However as this measurement access seeker has no clients other than the measurement devices, network performance can be carefully controlled to ensure that congestion caused by non measurement connections does not affect the results.
- 70) Tests of all key parameters would traverse the network from each network testing device, to the target and back to provide a round trip measure of performance.
- 71) The DSL connections connected to each network testing device would be randomly selected from the population of all possible DSL connections in the country as described in section 4.2.1.

- 72) The size of the sample is intended to be sufficiently large that it encompasses connections across the full range of devices such as DSLAMS that exist in Telecom's network. However it is not proposed to weaken the sample by choosing a representative group of connections as specified in the Draft Determination⁸.
- 73) Out of band management connections (indicated by dashed blue lines in the figure) are required to ensure that the only traffic placed on the links under test is measurement traffic.

4.1.2 Key Service Parameter Definitions

- 74) The sample rates and intervals described in the definitions below are intended as indicative measures only and are subject to confirmation by the pilot study recommended in paragraph 92).

Latency, Delay Variation (Jitter) and Packet Loss

- 75) The three key parameters of Latency, Delay Variation and Packet Loss are described together as the singleton metrics of Delay Variation and Packet Loss are calculated from the sample Latency metric.
- 76) In the following paragraphs reference to a packet of Type-P refers to an N-byte-ICMP packet where the value of N is selected from the set of packet sizes under test.
- 77) Latency measurements are conducted using the Type-P-Round-trip-Delay method defined in section 2 of RFC2681.
- a) One singleton metric is collected for each Type-P packet under test every 5 seconds using random additive sampling⁹ to distribute the measurement packets.
 - b) A 60-second interval is used to construct a sample metric for each of the Type-P-Round-trip-Delay measurements under test.
- 78) Delay Variation measurements are collected using a Type-P-Round-trip-ipdv metric, which is derived from the Type-P-One-way-ipdv metric, defined in section 2.1 of RFC3393. The particular selection function that is used in the calculation of a singleton Type-P-Round-trip-ipdv metric selects the singleton metrics with the maximum and minimum observed round trip delays from each Type-P-Round-trip-Delay sample metric.
- a) One singleton metric for each Type-P packet under test is calculated for each Type-P-Round-trip-Delay sample metric. This gives one singleton Type-P-Round-trip-ipdv metric every 60 seconds.
 - b) A 300-second interval is used to construct a sample metric for each of the Type-P-Round-trip-ipdv measurements under test.
- 79) Packet Loss is calculated as the ratio of packets for which an ICMP echo reply packet was not received to the total number of ICMP echo request packets that were sent in each Type-P-Round-trip-Delay sample metric. Response packets that take longer than 20 seconds to arrive at the network testing device are recorded as a loss.
- a) One singleton metric for each Type-P packet under test is calculated for each Type-P-Round-trip-Delay sample metric. This gives one singleton Type-P-Round-trip-Loss-percentage metric every 60 seconds.
 - b) A 300-second interval is used to construct a sample metric for each of the Type-P-Round-trip-Loss-percentage measurements under test.

⁸ Commerce Commission, *Draft Determination on the application for determination for access to and interconnection with Telecom's fixed PDN service 'Bitstream Access'*, 21 April 2005, para 249 p. 50

⁹ RFC2330 Framework for IP Performance Metrics. M. Mathis, J. Mahdavi G. Almes. May 1998.

TCP Throughput

- 80) TCP Throughput is calculated using the guidelines specified in RFC3148. Each test is conducted between the test target (referred to as the server for throughput tests) and a network testing device (referred to as the client for throughput tests).
- 81) Downstream throughput tests involve the server sending the largest sized TCP packets permitted by the path MTU to the client at the fastest rate allowed by the TCP protocol. All received data is discarded by the client after the appropriate TCP acknowledgement packets have been sent to the server.
- 82) Upstream throughput tests are conducted via the same procedure, except data is sent from the client to the server.
- 83) During a throughput test the only traffic permitted on the access link is packets related to the TCP session being used to test throughput.
- 84) TCP Throughput tests are performed using a standardised TCP reference stack.
- 85) A singleton TCP Throughput metric is collected for upstream and downstream throughput every 600 seconds.
- 86) A 3600-second interval is used to construct a sample metric for the upstream and downstream throughput measurements.

Service Availability

- 87) Service availability is measured as the connectivity between a network testing device and the test target. This is expressed using the Type-P-Interval-Temporal-Connectivity singleton metric, which is derived from the metric described in section 6 of RFC2678. This singleton metric is calculated based on each sample metric for Type-P-Round-trip-Delay. For each sample metric where at least one non-loss singleton Type-P-Round-trip-Delay metric was recorded Type-P-Interval-Temporal-Connectivity is inferred. In other words if the Packet Loss ratio is less than 100% for a Type-P-Round-trip-Delay sample metric Type-P-Interval-Temporal-Connectivity is established.
- 88) One singleton Type-P metric is produced for each Type-P-Round-trip-Delay sample metric. This gives one singleton Type-P-Interval-Temporal-Connectivity metric every 60 seconds.
- 89) A rolling 30 day interval is used to calculate the sample metric for each Type-P-Interval-Temporal-Connectivity measurement under test.

4.1.3 Key Service Parameter Sample Metrics

90) Table 1 summarises the sample metrics that are collected for each of the key service parameters defined in part 4.1.2 and for each of the recommended packet types.

Parameter	Metric Name	Sample Rate	Sample Period
Latency (also known as Packet Transfer Delay)	64-byte-ICMP-Round-trip-Delay	5 seconds	60 seconds
	200-byte-ICMP-Round-trip-Delay	5 seconds	60 seconds
	1472-byte-ICMP-Round-trip-Delay	5 seconds	60 seconds
Jitter (also known as Delay Variation)	64-byte-ICMP-Round-trip-ipdv	1 minute	5 minutes
	200-byte-ICMP-Round-trip-ipdv	1 minute	5 minutes
	1472-byte-ICMP-Round-trip-ipdv	1 minute	5 minutes
Packet Loss	64-byte-ICMP-Round-trip-loss	1 minute	5 minutes
	200-byte-ICMP-Round-trip-loss	1 minute	5 minutes
	1472-byte-ICMP-Round-trip-loss	1 minute	5 minutes
TCP Throughput	TCP-downstream-throughput	10 minutes	60 minutes
	TCP-upstream-throughput	10 minutes	60 minutes
Service Availability	64-byte-ICMP-Interval-Temporal-Connectivity	1 minute	30 days
	200-byte-ICMP-Interval-Temporal-Connectivity	1 minute	30 days
	1472-byte-ICMP-Interval-Temporal-Connectivity	1 minute	30 days

Table 1 : Key Service Parameter Sample Metrics

4.2 Statistical Framework

4.2.1 Experiment Design

91) The aim of this experiment is to compare the regulated bitstream and the retail Jetstream services and test if there are differences in the performance of their key service parameters that are of practical significance. These key service parameters include latency, jitter, loss, and TCP throughput. Test data will be collected from a number of test subjects over a number of days. The test subjects will comprise a random sample of DSL enabled phone lines.

92) Before a full scale comparative test can begin, a small scale pilot study needs to be performed to gain an understanding of the performance characteristics of the two services. In particular, this pilot study will provide information on the performance patterns and the significant variables or factors that are contributing to the overall variance observed within each service. The primary outcome of the pilot study is to estimate the variance components that exist in the network performance measurements of the DSL lines.

93) The required sample size for the study cannot be determined until the variance components have been measured. Higher variance will result in the need for a larger sample to achieve the desired confidence level.

94) The pilot study would require the collection of at least twenty five individual measurements per day for each key service parameter across four lines for each service. These measurements would be collected for three days.

95) Both the fixed-effects and the mixed-effects modelling approaches have been considered as options for representing the data. Fixed effects are parameters associated with an entire population or with certain repeatable levels of experiment factors. Random effects are associated with individual experimental units drawn at random from a population. In this

situation, phone lines, days and any interactions between them are treated as random effects and a mixed-effects model is used to represent the data. The advantage of this approach is that it avoids the problem of only modelling the specific sample of phones lines and days used in the experiment when the main interest is the populations of lines and testing days from which the sample is drawn from. The fixed-effects model treats all effects as being fixed and would not provide a useful representation of the data.

- 96) The variance of a particular observation is given by the sum of the variance components $\sigma_i^2 + \sigma_d^2 + \sigma_{i,d}^2 + \sigma_r^2$. These variance components can be estimated using the outcome of the pilot study. Based on these variance estimates, a sufficient sample size for further testing can be calculated. As noted in paragraph 93) the sample size will in some part depend on the magnitude of differences during the pilot study. The sample size will define the required number of lines, the number of days and the number of measurements per line per day that data must be collected for. The first priority in selecting the sample size is to minimise the number of lines required, followed by the number of days per lines and then the number of measurements per day.

4.2.2 Confidence Intervals

- 97) Once the full data collection period has been completed, analysis of the performance data collected for each key parameter sample metric described in Table 1 will be performed and the results expressed using confidence intervals.
- 98) A confidence interval for a parameter is an estimate of the range in which an unknown population value will lie. This estimate is based on a random sample from the population. For this measurement project the unknown population parameters of interest are the difference between a particular set of key service performance parameters such as the difference in mean latency between the regulated bitstream service and the retail Jetstream service. The confidence interval contains a set of plausible or believable values for the difference
- 99) Confidence intervals can be expressed in two forms. Suppose that the parameter of interest is the difference in mean latency for a particular size of packet between the regulated bitstream service and the retail Jetstream service, the following confidence intervals can be constructed:
- a) $a < \text{difference in mean latency} < b$ [two-sided confidence interval]
 - b) $a < \text{difference in mean latency}$ [one-sided confidence interval].
- 100) The two-sided form of confidence interval is more commonly encountered and may be paraphrased as “the difference in mean latency between the regulated bitstream and the retail Jetstream service is between a units and b units”. The one-sided form may be paraphrased to say “the difference in mean latency between the regulated bitstream and retail Jetstream services is at least a units”. The quantities a and b are respectively called the lower limit and the upper limit of the confidence interval.
- 101) A level of uncertainty may be defined with a confidence level between 0% and 100%. For a $(100 \times (1 - \alpha))\%$ confidence interval, we could say that $(100 \times (1 - \alpha))\%$ of all confidence intervals formed by repeating the experiment will include the population parameter. In other words, there is only a $\alpha\%$ chance that the sample is so extreme that the interval calculated will not cover the true population parameter of interest. There is a trade off between the amount of confidence in an interval and its width. Increasing the level of confidence will result in wider intervals and will require more data to generate intervals of usable length. Ninety-five percent has been found to be a convenient level for conducting scientific research and is used almost universally. High levels of confidence are almost always preferred as intervals of lesser confidence would lead to too many misstatements. MediaLab and the WAND group would aim for a ninety-nine percent confidence level in the statistical analysis however achieving this level is contingent on the results of the pilot study. If large variability (such as between individual phone lines) is shown in the pilot data then we may choose to use a ninety-five percent confidence interval to ensure that a high level of consistency is still achieved between the services.

4.2.3 Analysis of Identical Services

- 102) Establishing the consistency of two services is easiest when the services have identical design parameters such as upstream and downstream speed.
- 103) In such a scenario, the analysis is conducted using the calculated differences between the measured values for each key service parameter. Each resulting difference value is an estimate of the true difference between the two services for that parameter. The question of interest is whether 0 is a plausible value for the true unknown population difference. If 0 lies in the $(100 \times (1-\alpha))\%$ confidence interval, it can be concluded that there is no significant differences between the key service parameter under analysis given the level of confidence. In other words the experiment is unable to demonstrate a difference between the performance of the two services using this parameter. On the other hand if 0 does not lie in the $(100 \times (1-\alpha))\%$ confidence interval, then the differences can be considered statistically significant and the test results suggest a real numerical difference exists between the two services.
- 104) If a statistically significant difference is discovered in the previous step, the final step in the analysis process is to determine if the size of the difference is large enough to be practically significant. If the difference is not practically significant then the services can be declared consistent as defined in point 17)b).
- 105) The definition of what level of difference is classified as practically significant is the most difficult part of this analysis. It is suggested that this is an area where the figure is best agree between Telecom, TelstraClear and the Commerce Commission. The statistical analysis is able to work with whatever figure is defined.

4.2.4 Testing for Non-Identical Services

- 106) The method described in part 4.2.3, for analysing identical services can not be used to establish the consistency of two services with difference design parameters. For example the method may not be used to compare a connection with a 2Mbps downstream rate with a connection that has an unlimited downstream rate.
- 107) In such a situation a standards based approach must be used to determine the consistency of the two services. Once the standard has been set, both services are compared to it using a confidence interval. If the performance of both services meets the standard they are said to be consistent.
- 108) Again, the definition of any standards to be used in establishing the consistency of services with non-identical design parameters is best agreed upon between Telecom, TelstraClear and the Commerce Commission.
- 109) This standards based approach is a significantly weaker test of consistency than the approach proposed in part 4.2.3 as it allows to services with significantly different performance characteristics to be declared consistent if they both meet the standard. This risk can be somewhat mitigated by defining tight standards, however there are still no absolute guarantees that two services shown to be consistent via the standards based method do not have large variations in performance.

4.2.5 Time based requirements

- 110) The indicated rate field of Appendix A in the Draft Determination frequently requires a key service parameter to achieve a specified standard for >95% of the time. This type of definition is only relevant in the case of comparing services with non-identical design parameters.
- 111) It is important to note that the requirement to meet a standard 95% of the time is not equivalent to a 95% confidence interval. The correct statistical method to calculate the compliance of a key service parameter to a time based standard would be to classify each measurement as 1 or 0 depending on whether it met or failed the standard. A 99% confidence interval would then be generated on the distribution of 1s and 0s. Compliance with the standard for 95% of the time is achieved if the lower limit of the confidence interval is 0.95 or greater.