Consumer Internet Performance Monitoring

Richard Nelson, Tony McGregor

August 3, 2006

Contents

1 Introduction

This report has been prepared by the WAND network group for InternetNZ. It describes the feasibility and options for measuring consumer broadband services in New Zealand. Specifically: "InternetNZ wishes to provide a facility that will periodically test broadband service availability and quality dimensions such as speed, throughput, latency, jitter and other measures that may be of interest to New Zealand Internet consumers."

1.1 Aims of the Test Facility

The primary purpose of the facility is to provide a means for consumers to measure and understand the performance that they are receiving from their ISP. In order to do this, the facility needs to have valid measurements from a range of ISPs so that comparisons can be made.

A secondary purpose of the facility is to measure the overall performance of the Internet in New Zealand in such a way as to be able to publicly monitor improvements and inform regulators and other interested agencies.

Another possible role of the facility would be to increase the competitive pressure on ISPs to improve the performance and services that they offer to their customers.

1.2 The Report

There are two main sections to this report. Section 2 covers general issues that will need to be considered in terms of any system set up to achieve these aims. Section 3 covers the options of actively or passively performing the measurements. Subsequent to these main sections a brief description of WAND's relevant experience and capabilities is included in section 4. Finally a set of recommendations for a programme that would produce the desired facility is given in section 5.

The main assumption that this work is based on is that the objective is to achieve as much measurement as is possible for a limited budget. To this end, much of the report assumes that there will be some kind of software for users to download and install although options for other approaches, such as dedicated hardware are also included. There is also some emphasis on how to increase the rate of consumers opting in to use the facility.

2 General Issues

Regardless of what technical approach is taken, there are a number of issues that will be important. They include the nature of the metrics to be collected and the results produced, the need to maximise direct user benefit to encourage use, legal and IP issues and deployment requirements and constraints. These are discussed in the following sections.

2.1 Metrics Recorded

There are four types of information that a consumer broadband test tool could measure. These are:

- Performance metrics
- Service availability
- Customer usage
- Destination addresses or paths

Measured metrics may be of the "normal" network performance metrics such as available bandwidth and throughput, delay and jitter, loss and reordering, availability or they may be more user or application oriented metrics such as web page download time or estimates of Voice quality. The normal metrics are better defined and more easily and reliably measured so they provide a very good basis for comparison. The more application oriented metrics are closer to measuring the perceived performance by Internet users and are more easily understood by non-experts. However application orientated measurements are typically composite metrics making them less suitable for comparison between ISPs.

Service availability can cover the reliability of the connection, but also the availability of specific network services. Examples are blocking of port 25 to prevent spam forwarding and the availability of multicast or IPv6. Further performance metrics might need to be taken on a per port or service basis to discover whether certain services are being adversely controlled within the network. It would also be useful to be able to detect whether packets are being mangled within the network, examples include packeteer style manipulation of TCP windows to control throughput, MSS advertisement adjustments performed by firewalls and NAT. Other examples of services include the use of transparent proxys and the provision of static IP addresses.

Received error messages are another useful metric as they can show examples of service (un)availability and of network reliability.

Customer usage is not necessarily important in terms of the service received by individual customers but it is interesting in the sense of surveying the use customers are making of the network and the load being imposed on the ISP network. In particular the correlation of load and performance will be quite important particularly as different ISPs encounter peak load at different times of the day. Individual usage patterns could be of interest to some consumers as they may be unaware of some of the network activities of their computer. These may include applications such as spyware, peer-to-peer applications, or use by other members of the household.

Destination network address choice could be a standardised set or to measure to the destination addresses that the customer is actually using. If a standard set of destination measurement points is used, the selection is important. The points should be chosen to ensure measurement of different parts of the network and some should be representative of the destination networks that are most commonly used. Standardised destinations make comparisons between different customers connections more robust and enable the performance of different parts of the network. Measuring the customers actually used destinations provides a better representation of their actual performance. It is likely that there will be a number of commonly used sites (e.g. Trademe, Google, Windows update) that would provide good comparisons.

There is clearly a very large variety of choice of what to measure. Measuring more things will provide a more complete picture of the overall service being offered by ISPs but will require a greater level of analysis and interpretation and may confuse users. Any choice of metrics, services or destinations to measure will need to be relevant to consumer use of the Internet and may need to adapt as consumers change their use.

2.1.1 Recommended Metrics

The most important metric to measure will be throughput. This is because that is the basis on which ISPs price and market their plans.

Standardised metrics such as delay, jitter and loss should be measured across a range of ports and protocols to enable checking of whether particular services are being throttled or otherwise controlled within the network.

Destination test point address choice is likely to be dictated by the choice of passive or active measurement system as described in section 3. In either case, careful design of the system should make the measurements useful and relevant.

Other measures are interesting, but not of primary importance. Application level measurements are specific only to the application they are measuring. Testing for availability of little used services is only relevant to a small number of consumers. These sorts of tests should be added to the system only if and when doing so is simple with an acceptable cost/benefit ratio.

2.2 User Benefits

The measurement system must provide benefits to justify implementing it. If the system relies in part or full on members of the public to choose to use it then it needs to provide obvious benefits to those users. The better the benefits, the more likely that a usefully large number of users will provide results.

Since the intent of the system will be to measure consumer broadband performance it makes sense to provide it as a basis for consumers to be able to see how the performance they are receiving compares with others. This also recognise that the users are generating the data and therefore have some right to see what the data is.

This implies that any monitoring system should provide an easily accessed and clearly comprehensible feedback of what service they are receiving. A very simple interface offering perhaps a peak throughput number or graph and a "traffic light" colouring indicating the overall health of the connection would be a possible initial screen. Further details would be available via some sort of drill down facility (a menu or icons).

Examples here are the common speedtest sites that are available. Although the usefulness of these can be questioned for a number of reasons, they provide a single number that is easily understood and can place users on a performance chart for easy comparison providing good feedback. A counter example is the Neti@home system which collects a lot of usage performance statistics but limits the feedback provided to simple geographical traceroute type tool, seemingly unrelated to the performance statistics. In particular there is no way for a user of the measurement tool to compare their connection performance with others.

It should be noted that for comparison and survey purposes, the facility needs a large number of users to be useful, but it also needs to be useful to attract a large number of users. Therefore making it immediately useful for monitoring individual connections is important.

2.3 Legal Issues

WAND does not have any formal legal expertise. We do know that collection of network performance data does have some legal considerations and this section is a collection of the issues we have encountered.

Some kinds of network data can provide a great deal of personal information and so careful consideration needs to be given to exactly what data should be collected and how the data collected is managed. In general a performance measurement system has no need to collect the user content of network packets but even recording the destination address from the packet header reveals information about a user. This can be anonymised but this process reduces the usefulness of the collected data.

Data privacy is important, both from a legal perspective and from maintaining the confidence of the users of the system. Publication of the data will have to provide sufficient anonymisation through aggregation and/or randomisation of network addresses so that individuals cannot be identified.

Matching collected statistics with the ISPs may be possible with a combination of the source address and BGP tables. However, it may also be interesting to match performance data with geographic locations and telephone exchanges. This means collecting addresses and telephone numbers. Also, a system may require the ability to contact users through email addresses. Personal data such as this should not be collected and stored unnecessarily, however not collecting some data will limit the usefulness of the data. Personal data needs to be stored securely and have appropriate access control.

Related to the data privacy issue is the choice of who will collect and store the data –in effect this is the choice of who will operate the system. Different organisations have different requirements –for example the University of Waikato (like other universities) has an established ethics approval process; WAND must work through this process. The organisation operating the system should be seen to be independent and have some public accountability to maximise public confidence in the facility.

It will be desirable to publish the data in various forms. The primary purpose is to provide users of the system a comparison with other users. Other reasons for publication may include to publicise the system and the results found, for submissions to regulatory bodies and for academic publications. Finally, there may be interest in offering some detailed data to ISPs if they are interested in aspects of their, and their competitors, performance. Clear policies should be set as to what uses the data will be put to and these should be publicly available.

The results have to be defensible in terms of their representation of the performance of various service providers. ISPs whose service is portrayed poorly by the results may be tempted to try to suppress them. It is therefore important to be specific and clear about what the results represent; it would be almost impossible to say that ISP X always offers more throughput than ISP Y but it may be possible to say that the throughput measured at ISP X by the tool had a certain value and that measured at ISP Y had another, different value.

The specific legal issues that apply will depend on the details of the system chosen, particularly the kind of information that is collected. These issues can be managed reasonably but do require consideration during the design phase.

2.4 Measurement Reliability

In order for a measurement system to be useful, the results it reports must be credible. In normal scientific terms this means that the results should be statistically significant and repeatable. Network measurements are however inherently non-repeatable due to the rapidly changing environment, for example two delay measurements taken within the same second are unlikely to be exactly the same and may differ considerably. The focus for network measures is, therefor, on statistical validity.

A tool that can be downloaded and run by members of the public has very little statistical validity. The computer and network hardware is uncontrolled, the users are self-selected and the results may be considerable affected by other network traffic. Alternatives that involve controlling the hardware and network connection and network interfering traffic involve the expense of providing and managing dedicated equipment and connections. This kind of scenario can be very useful for an accurate comparison of ISP performance but it is removed from the idea of enabling consumers to measure the performance of their own connection.

This underlines the importance of being very clear on how the measured results are arrived at and being cautious about how they are interpreted.

2.5 International Service

There are two different international aspects that may affect a consumer broadband measurement system. The first is whether to measure connection performance for New Zealand based users to international sites. The other is whether to collect statistics for users and ISPs in other countries.

International performance is one area where ISPs may vary considerably. There is significant cost involved for the ISPs in their international transit and they can differentiate themselves through their international performance even where a common access network is used. A very high proportion of New Zealand Internet traffic is to international sites. It is therefore important to include international connection measurements. The choice of which countries and networks to measure to is also important. One obvious target is the tier 1 US based ISPs but there are sections of the New Zealand public who may be particularly interested in connection performance to other areas of the world (e.g. Asia) where ISPs have connection options with a wide range of performance characteristics.

It may be difficult to avoid having consumers in other countries at least attempt to download and try to use any software tool made available. It would be possible to discourage them by not providing any reporting of results if their address is not advertised by a known New Zealand ISP, but it is not clear that this is desirable. The ability to collect international consumer data has the potential to increase the amount of set-up configuration and vastly increase the number of users of the system, but it will provide a basis for comparing the performance received by New Zealand based consumers with that received by consumers in other countries. Interest in this type of system has already been expressed to WAND by an overseas based network architect. If the tool is testing performance to specific network addresses in New Zealand (e.g. pinging to servers at APE and WIX) then there may be problems, or at least the measurements may not be so useful if it is used overseas.

If the facility is based on dedicated hardware coverage will be more controllable, but probably more limited, even within New Zealand.

2.6 Deployment Issues

Without finalising the design of the system and especially if it is new and untried type of service it is very difficult to anticipate all the deployment issues that will arise.

If it works well and is of considerable interest it may scale to thousands or tens of thousands of users. If international interest develops it may have many more. Any server infrastructure needs to be on a separate domain so that it can be moved and scaled as appropriate.

Client software needs to be reliable, as simple to use as possible, and easily updatable. It is unlikely that the initial version will have all possible features and the extent to which it will be developed will be dependent on the success of the project. Clearly, minimising user difficulties and frustration will reduce the support requirements and increase the number of successful installations. Having configuration done through a web interface to the server will provide the ability to deploy interface changes without updating client software.

Time synchronisation will be important. Although measurements are likely to be round-trip, it is important to know accurately when the measurement was taken. If consumer computers are used, their time settings may be wrong by minutes, hours or days. It would be possible for an application installer to recommend or attempt to turn on an NTP service, but this may be later turned off or timezone settings may be wrong. A lightweight system calibrating the reported timings should be incorporated into any results reporting mechanism in this case.

Consumers, especially in New Zealand, often have data caps on their connection. It is therefore important that any measurement service that they are to use restricts the amount of network traffic it generates to a reasonable volume. For example if the service were restricted to generating 100MB of traffic per month, this would be $100x10^6$ /(30x24x60)= 2314 bytes per minute or 36 packets at 64 bytes/packet. This is not a lot of data to test, and report on, many different services to a large number of different network destination addresses. It excludes the higher rates required to do regular throughput testing. This is an argument for measuring the consumers actual traffic use rather than any standardised set of tests.

If the system is successful then it has the potential to become a target for attack. This will encompass the normal issues of running publicly accessible servers but also include issues of ensuring the submitted measurements are genuine. Encryption of results to stop tampering is possible but will it also be necessary to guard against people modifying the measurement client to make results look better or worse than reality. There will also be the possibility of ISPs blocking access by their users to any server receiving the results. To counter this, it might be possible to design a peer to peer submission system, but a simpler alternative might be to design the public interface such that ISPs with few or no submitted results will look similar to ISPs with poor results.

2.7 Results

To help with design it will be useful to speculate on what sorts of results could be expected and what features of the results are important to display. Fortunately, other measurement efforts provide a base of knowledge of the NZ Internet environment.

As a starting point, it can be expected that most traffic will be bound for the US. Most ISPs will have direct connections to the US and so differences will mostly be seen in queueing effects such as jitter and loss. Traffic to other international destinations will mostly go through the US, but there will be delay differences for ISPs that have direct connections to Australia and possibly through Australia to Asia. There will be differences in delay and throughput performance for access to some New Zealand content depending on ISPs peering arrangements; measuring how much traffic goes to locations that show these kinds of differences may be of interest.

Most broadband connections use Telecom's ADSL services so there will be great similarities in the access network performance with differences for TelstraClear's cable network and WISPs. The performance of Telecom's ADSL network is clearly of major interest, particularly how it varies as the load varies. As unbundling evolves more competitive networks may arrive and the relative performance of these will be of interest too.

Looking for attempts to restrict specific services is important, these may show up as decreased throughput or increased delay or jitter. It would be of interest to see if this occurs on 3G mobile networks but some thought would have to go into how to target users of these networks.

2.8 Intellectual Property

There are two kinds of intellectual property associated with this facility, the software and the results data. The results data will be subject to privacy concerns and a publication policy as described in section 2.3.

There are two main parts to the software, the client and the server. In a world of spyware and other forms of malware it would be good to publish the source code to the client to enable technically literate users to understand what it is doing and increase confidence in the use of the facility. This would be in line with the general practice within WAND and other InternetNZ sponsored developments. It does mean that any development work should be based on suitably licensed libraries if required.

The server software will not be distributed to consumers as such. It is also not clear that consumers gain any particular benefit from having it available. Further, the server may contain techniques to attempt to detect submission of bogus results which would benefit from not being public knowledge. It is therefore suggested that the server code be kept unpublished but consideration could be made to sharing if with any group attempting a similar project in other countries if they request it.

3 Technical Approaches

Network measurement systems need packets to time and count. There are two possible sources for these packets, the measurement equipment itself or the normal network traffic of the users. These two approaches are termed active measurement (the measurement system actively injects packets into the network) and passive measurement (the measurement system simply observes what is already there).

In either system, it is common to provide dedicated hardware to do the actual measurement. This ensures that measurement differences reflect differences in the network, not in the measurement system from place to place. In a consumer broadband system any dedicated hardware would have to be very cheap and, although the accuracy of individual measurements may be improved, it may be preferable not to use it at all in order to keep the cost reasonable obtain as many measurement sources as possible.

This section describes how to do active and passive measurement across a wide range of consumers and describes the advantages and disadvantages of both.

3.1 Active Measurement

An active measurement system consists of a set of machines performing network tests. The most common such test is a simple ping which can measure delay jitter and loss. Other common tests include traceroute, TCP throughput and OWAMP. Active systems are not limited to simple tests though and can be setup to test any specific application and any metric. Tests may be run between the devices themselves as a mesh, to a set of test servers installed for the purpose or to any well known network service.

Because active measurement systems only deal with the traffic that they produce they do not need much computational power or storage and can be very cheap and so it is feasible to install reasonably large numbers of them if suitable locations can be found. It is feasible to install active measurement probes on consumer adsl/wireless routers or 486 class embedded devices, see: `http://www.ripe.net/ripe/policies/proposals/2005-10.html` and `http://wiki.ntop.org/mediawiki/index.php/CBM`

Alternatively, an active measurement probe running as a background process would consume very little resource on a consumer computer.

Active measurement has a number of inherent advantages:

- Since the traffic measured is generated by the measurement system itself there are no privacy concerns regarding the contents of packets.

- The traffic can be precisely controlled so it is clear which parts of the network are being measured.
- The traffic is standardised across all measurement points so direct comparisons of the results can be made.
- Measurements can be made at controlled regular intervals giving coverage of the full day.

It is important with active measurement systems that the generated traffic does not interfere with the network being measured. This is particularly true when measuring the current generation of consumer broadband connections with low speeds and low data caps. Normal active measurement involving regular pings and less frequent traceroute tests generates a very small amount of traffic - a few hundred bits/per second or a few tens of MB per month. However throughput tests generate much more traffic and can fully utilise the link while running and testing a very large number of different services and traffic types would also generate a lot of traffic. These sorts of tests could be run only infrequently, probably randomly from each test point. If enough test points are available for each ISP then the aggregate measurements would provide a good indication of what service they offered.

Another possibility would be to offer a control panel such that individual users could choose to manually run a test. The test would interfere with other traffic on the link but the user would be aware of it happening. This option is similar to the common web based speed tests, but could be more sophisticated with options such as “test bittorrent throughput”, “test VoIP jitter”, “test web throughput to the USA” etc.

3.1.1 Dedicated measurement infrastructure.

In order to measure the performance of an ISP’s network it is useful to remove the effect of a loaded last mile connection. This implies that a separate dedicated connection should be used. This is an option that can be used to provide a baseline for an ISP. The disadvantage is the cost of these connections, the need to keep their purpose secret and the need to measure a enough connections to be able to correct for differences in the performance of the copper loop or variations in the loading of specific DSLAMs and back haul links.

3.1.2 Risk

The technical risk of establishing a system of this type is very low. Active systems have been around for a while and are reasonably mature. The main problem is that the measurements will have some interaction with individuals normal Internet use.

3.2 Passive Measurement

Passive measurement is the process of capturing actual users data and processing it to derive the performance metrics for the network. Additional processing is required because the traffic is normally application oriented and not specifically designed to measure network performance, e.g. an active system measures round trip delay by sending a packet and waiting for the response, but a passive system has to dissect a TCP stream to match packets in each direction and then calculate the RTT.

Dedicated passive measurement equipment, such as that made by Endace systems (www.endace.com), is very expensive and the installation requires careful planning. However this equipment is designed to reliably capture and accurately timestamp every packet on high speed optical links. It is unnecessary for the traffic speeds and volumes that are possible on individual consumer connections. In the consumer scenario, simple capture using standard network interfaces is sufficient. Very little of the CPU or memory resources of a modern computer are required to monitor the traffic of todays consumer broadband connections which are comparatively slow.

Measuring user generated traffic is both an advantage and a disadvantage. The advantage is that measurements are precisely of the types of traffic and network paths that the users are most interested in. The disadvantage is that there is no consistency or control over what is being measured so comparisons can be difficult. For example if many users are trying particular services a passive system will record detailed measurements of the performance, but conversely, a new service with no users cannot be measured. Also, there will be lots of measurements from the busiest periods of the day but it may be difficult to establish a baseline measurement from the quiet times.

Passive measurement only uses the connection bandwidth for report results so it is possible to monitor a large number of different metrics limited only by the actual use made of the network.

Another advantage of passive measurement is that it can provide detailed information about what use is being made of the network.

Disadvantages of passive measurement also include the fact that it is much more privacy sensitive. Data content obviously must be removed, but even then users may not want details of which web or other servers they have used to be recorded. There is a balance between maintaining user privacy and recording useful information.

An example distributed passive system is the Neti@home project http://neti.gatech.edu/. This does not appear to have been hugely successful, in part because there is little incentive for users to be involved and possibly because it is largely staffed by post-grad students and so may not have had much continuity. Nevertheless, it has produced published results showing that a system of this kind is practical.

3.2.1 Measurement from Service Providers.

As an alternative to measuring from consumer ends, it is also possible to measure performance from the service provider end of a connection. This clearly requires the co-operation of the service provider and may be restricted by their hosting arrangements.

The advantage of this approach is that it would be possible to monitor the performance of all users of that service from a single point. The disadvantage is that it is only a single service that is being monitored. Although multiple service providers could be used, for some network applications there is no single central service provider that could be used (i.e. peer to peer services).

The simplest way to do this would be to use web server log analysis. Some service providers already do some performance comparisons (e.g. Trademe see: http://www.trademe.co.nz/Community/SurveyResults.aspx). More detailed results could be achieved using passive capture and analysis providing metrics such as RTT and loss.

This kind of testing does cannot easily provide individual measurements for individual consumers but if the opportunity is found it would be a useful basis for comparing ISPs using broad survey of their customers.

3.2.2 Measurement Summary and Comparison

The results reporting for a passive measurement system will require a lot of careful design work in order to produce meaningful comparisons as users are all doing different things on the Internet, however this is definitely achievable. A system is required that selects comparable measurements (e.g RTT to Trademe) and provides summaries based on that. Some metrics may be more broadly comparable (e.g. peak download speed) and some may be quite specific.

Due to the wide range of end node systems/configurations it will be important that the system can present some idea of the range of measurement results. Also there should be some calculation of the minimum number of results that are required for statistical validity.

This need for selection of measurements to be compared will increase both the amount of development work required and the processing overhead on the server

Data reduction through processing will be desirable due to the potential for a large volume of data. This processing can be distributed over the measurement points. Thus results are likely to be aggregated by destination networks (also providing a small amount on anonymisation of the results). Flow based averages and/or time binning might also be used.

3.2.3 Risk

A distributed passive system will have some technical risk due to the development work required to build a server that selects appropriate results for comparison. However, we have a clear understanding of the requirements and the actual risk is not high. There is also risk presented by the need to manage the private nature of the data that will be collected. Mainly this increases the risk of consumers not accepting the system. This risk could be mitigated by including consumer selectable privacy settings within the tool.

3.2.4 Hybrid Passive-Active Measurement

It is possible to combine active and passive measurement. In particular a measurement point doing passive measurement could actively connect to a limited set of standard points to provide standard comparison measurements and a baseline measurement at all times.

Some sites will be connected to by most users and will provided a degree of commonality without explicitly generating traffic. There are, for example, a range of network services that modern operating systems connect to automatically and regularly for activities such as software updates and network time synchronisation. Also, any computer running such a passive measurement collector will regularly connect to a server to report the results. Many users use similar sites such as Google, Yahoo and Trademe.

Only with the experience of measuring a group of consumers will it be possible to tell how useful specific active probes to a passive measurement system will be.

3.3 Comparison of approaches

	Active Measurement	**Passive Measurement**
Traffic added	Low, but increases for throughput tests or large numbers of tests	Very low -only measurement reporting
User Privacy	Very good	Controlled by capture software
Metrics	Wide range: any that can be test code can be written for	Wide range: any that can be derived from captured traffic
Test choice	Completely controllable	Entirely set by network user
Service performance tests	Individually setup	Automatically tests all services used
Service availability tests	Individually setup	Only if the user attempts to use the service
Application tests	Can add as extra tests	Can add as extra analysis
Network destination addresses	Controlled and standardised	Set by network user
Network usage data	No	Yes
Measurement reliability	Very Good	Requires careful interpretation
Relevance to consumers	Good	Very good
Results Comparison	Direct comparison possible	Requires careful selection

4 WAND

The WAND Network research group is based in the Computer Science Department of the University of Waikato. Although InternetNZ are familiar with WAND through our hosting of NZNOG'04 and '05 and commissioning this report, this section presents a few details of some of our major measurement projects to clarify our experience. WAND is network measurement group with international standing and would be very interested in direct participation in developing and establishing a consumer broadband testing facility.

4.1 Active Measurement

The active component of the US National Laboratory for Applied Research (NLANR) Network Analysis Infrastructure (NAI) is their Active Measurement Project (AMP). (`http://watt.nlanr.net/`). The project is lead by Tony McGregor who is a member of the Waikato University WAND group.

The AMP project has two components, a deployed network of measurement machines (Approximately 150 monitors focused on the US advanced research and Education Networks) and about 30 internationally on national advanced research networks. It is the largest project of its kind.

The other component of the AMP project is open software for active measurement. This software is only available as a beta release but has been deployed in a number of measurement efforts including other Nations' Advanced Networks (e.g. Korea) and within ISP and public networks. Much of the software for the AMP system has been developed by members of the WAND group working at Waikato. There are several smaller AMP systems in New Zealand run by the WAND group.

4.2 Passive Measurement

The measurement work within the WAND group developed as a need to obtain accurate inputs for network simulation. This led to the DAG series of passive capture interface boards. Today these boards have been spun of into the company Endace Systems (`http://www.endace.com`) which is one of New Zealand's fastest growing companies. The DAG boards continue to have world leading performance and unique capabilities for network measurement.

Libtrace (`http://research.wand.net.nz/software/libtrace.php`) is a more general purpose cross-platform library, designed for both packet capture and trace analysis. It supports pcap interfaces as well as Endace DAG cards and traces. Also a unique feature is a real-time remote-capture interface with support for multiple simultaneous users. Libtrace extends pcap type functionality with packet parsing features and the distributed software provides a number of useful traffic analysis utilities. Libtrace is used internationally by network researchers. Libtrace is developed under Linux but supports a range of UNIX style operating systems and the latest version has been ported to MS Windows an supports WinPcap captures.

WDCap (`http://research.wand.net.nz/software/wdcap.php`) is used by WAND for its high performance capture projects at the University of Waikato and elsewhere. The current version is optimised for reliable capture and storage of extremely long traces (months to years). It supports remote capture execution, adaptive removal of user data and key based address anonymisation via the CryptoPan.

In addition we have a limite prototype passive measurement tool aimed at ISP customers that demonstrates some of the requirements of a consumer broadband test facility. This tool is based on the Libtrace library and records statistics per destination and aggregates them per /24 of Internet address space. Results are reported every 5 minutes. The tool is not yet suitable for widespread use and only reports a limited number of statistics. Also, so far there are only a small number of users intermittently submitting data so there is no basis for constructing aggregate performance plots.

5 Recommendations

For any network, the ideal situation is to perform all forms of testing possible as each adds different information enabling a more complete picture of the network to be constructed. Unfortunately this is generally not possible due to budget and other resource constraints.

For the consumer broadband test facility it would be possible to get a good understanding of ISP service performance using either active or passive capture. In either case, WAND has software that can be used as a basis for developing the facility. Further active testing could be performed on dedicated hardware connections or using consumers connection and computers.

A dedicated active test infrastructure would be the most accurate way to perform performance testing and could be supplemented with a tool to allow consumers to do comparison testing of their own infrastructure - this however would require a large number of dedicated connections and would therefore have significant ongoing testing.

WAND recommends that a passive approach be pursued if the budget needs to be minimised. The key reasons for this are:

- Throughput is a key metric due to the way Internet access is sold. Active testing of throughput requires significant data volumes and is inaccurate if the connection is being shared with other traffic.
- Passive collection enables accurate monitoring of all services that are used. This makes it adaptive as new applications become available and enables monitoring of a wide range of services with negligible increase in traffic.
- If run on a real users connection, an active system would have considerable interaction with customers normal network use. A passive system simply measures that use.

A passive system will require a greater amount of initial development work, particularly at the server. For this reason the risk is somewhat higher than a purely active system and the cost of establishing the facility will be higher than a software only active system.

It may desirable to add a limited number of active tests as experience with system develops to test the availability of key services and measure latency to a small set of common network addresses.ma

Whatever technical approach is eventually chosen, it will be very important that the system is easy to use and provides benefits such as clear and useful information to individual consumers so that wide uptake is encouraged.

The legal and practical issues of deploying this system are important but do not appear to present obstacles that would prevent deployment or restrict the utility of the facility.

5.1 Programme

A consumer broadband test facility such as this is a new development and since it is critical that it be easy to use and provide good results feedback a reasonable development effort is required. It also is important that a useful number of lead users are found in order to establish sufficient data that when it is promoted to the general public new users find the results quickly useful. The programme should therefore include a phase of limited release and testing after the initial development and before full public release and promotion.