Cost Effective 1 Minute Network Data Collection: A New Paradigm

Bruce Hartley

Meteorological Service of New Zealand Ltd.

P.O.Box 722, Wellington, New Zealand.

Phone +64-4-4729379, Facsimile +64-4-4700772

bruce.hartley@metservice.com, www.metservice.com

13 January 2005

Abstract:

A new paradigm is emerging in the method, timeliness and frequency of data collection. New Internet Protocol (IP) based networks and communication technologies have made it possible for **cost effective 1 minute network data collection** from remote stations. Message intervals as frequent as once per minute are possible nationally and internationally at costs comparable to methods used for traditional 1 to 6 hour communications.

Messaging at 1 minute intervals can provide network and equipment operating benefits. Some statistical and derived data processing can be moved from stations to the central data server. Most IP network Data Link Layers include error detection and correction which alleviates the need for user level detection of message corruption. These benefits mean: Message sizes can be smaller thereby minimising communications costs; Frequency and complexity of remote equipment software updates are reduced in deference to single software updates on a local server; Processing overhead at stations is reduced thereby reducing their complexity, capital cost and maintenance.

One minute near real time data collection creates new opportunities. Weather forecast accuracy may be improved with more timely data or by triggering forecast reviews from appropriate changes in conditions. Fine detail weather forecasting and climatology research can be performed on a day-by-day basis or in hind sight after an event. Creation of commercial products with fine time resolution and near real time delivery are possible. In areas with cellular IP coverage rapid deployment stations are possible.

For AWS that do not fall within direct coverage of IP communication technologies alternative methods need to be considered when implementing new networks or upgrading existing ones.

This paper describes the design considerations and implementation steps taken during upgrade and expansion of MetService's network using 1 minute messaging and GPRS cellular IP data communications.

Key Phrases

- The world is undergoing a data revolution.
- High rate data can be provided at excellent value for money.
- People love watching the weather and now they can do it real time.

Acronyms

ATIS	Automatic Terminal Information Service
AWS	Automatic Weather Station
CDMA	Code Division Multiple Access - IP cellular network operated by Telecom NZ
GPRS	General Packet Radio Service - IP cellular network operated by Vodafone NZ
MetService	Meteorological Service of New Zealand Limited
NZ	New Zealand
PacNet	Packet Switch Network - X25 packet switch network operated by Telecom NZ
POTS	Plain Old Telephone System
SIM	Subscriber Identity Module - Small circuit module used in Vodafone modems and phones
TCP	Transmission Control Protocol - An IP network information transfer protocol
UDP	User Datagram Protocol - An IP network information transfer protocol
VPN	Virtual Private Network

Introduction

The author works for the MetService which is the official NZ public weather service provider (www.metservice.com). The author's area of expertise is automated surface observation and so the discussions and illustrations presented in this paper are focused towards this area, however the principles employed can readily be applied to other areas.

Changes in cellular communications technology that have been implemented in NZ by the two major telecommunications network operators has presented a relatively low cost cellular GPRS communications infrastructure and MetService is currently implementing network changes to take advantage of this. To begin with MetService has chosen to use the Vodafone NZ GPRS network however this does not exclude the CDMA network from being used in the near future, particularly where some AWS do not have GPRS coverage but do have CDMA coverage, and the use of both networks can provide diversity in the unlikely event that either network provider suffers a major network failure.

Network Features and Operating Considerations

The following are some of the features to be considered when deciding whether to invest in using a Cellular IP network.

- <u>Network coverage</u> - Areas with cellular coverage are also areas with high commercial opportunities for real time data. i.e. The areas where data is valuable as a comercial commodity is generally where there are higher population densities (where people live, congregate and travel), and for the same reason (higher population density) these are the areas where cellular network operators ensure good coverage e.g. cities and large towns, main highways, airports, marine ports, high density industry, popular leisure areas. Cellular networks use radio communications and due to the inherent limitations of the technology there will always be locations that do not have coverage. Some sites may need to be re-located nearby or use a radio modem or dedicated land-line link to a site where there is cellular coverage. Some sites may be completely outside coverage and therefore require an alternative means of communication. It should not be assumed that a new cellular network will provide a total replacement solution for an existing communications network.

- <u>Network availability</u> - The quality of service provided by the network operator should be investigated. This will include the network availability (up-time), mean time between failure and mean time to repair. The Vodafone NZ GPRS network has maintained an availability of better than 99.9% over the 12 months to December 2004

- <u>Network Support</u> - The level of support to be provided by the network operator should be considered. This will include demarcation points, the fault reporting process (help desk), and the expected time to respond. Support should also include notification by the network operator of scheduled maintenance well in advance, faults that have been reported and are under action, and when faluts are cleared.

- <u>Network Security</u> - Some network providers can implement a VPN, normally for a routine management cost. This allows secure data transfer from the AWS to the central data processor, and fixed remote modem IP addressing which makes communication with a remote modem easier.

- <u>Message transfer protocol</u> - With IP communications messages are broken into packets and then commuicated over a physical medium (e.g. an ethernet network, the internet) to a pre-defined destination. There is a protocol layer between IP and the user application that controls the way that messages are packetised, sent, received and handshaked. Different communications modems may implement one or several of the available protocols. For the NZ cellular modems the two main protocols are UDP and TCP.

  <u>UDP</u> messages are split into packets of several hundred characters in size (512 characters for Vodafone NZ) and then the packets are individually sent like a letter or an email to the destination. After the packet is sent the sender does not know how long the packet takes to reach the destination or in fact whether the packet even reaches the destination. Each packet can travel by a different route so multiple packets from the same message or even different messages can arrive in a different order at the receiving system. Due to the relatively large time difference between messages (typically 1 minute) compared to the time between packets from the same message (typically milliseconds) it is very unlikely that packets from different messages will become mixed. The receiving system software is therefore required to sort and reconstitute messages that have been split before they are passed to the message decoding process. This problem can be avoided by keeping individual message sizes below the packet size. UDP packet overheads are less than TCP.

  <u>TCP</u> messages are split in a similar way to UDP however this protocol includes the re-ordering and stitching process at the receiver as well as packet resend requests when they are corrupted or lost. The process is a like a two-way radio conversation in that the sender knows whether the message gets through from the responses sent back from the receiving system and the information can be resent if required. As long as there is some connectivity within the protocol timeouts, the message is more likely to get through, and if it does not the sender is notified. TCP packet overheads are higher than UDP, and the characters transmitted per message will be higher, particularly if there are retries.

- Message transfer parameters- Cellular IP networks are normally relatively high speed (tens to hundreds of kilobits per second) and the time to transmit simple data messages is typically less than a second. Latency (the average time taken for message to travel from an AWS to the central data processor) is typically a few seconds. Jitter (the variation in the latency) is typically hundreds of milli-seconds. It is worth noting that when a VPN is employed the latency for the first packet in a burst of packets is increased by several seconds due to the overhead in setting up the virtual "tunnel".
- Billing regiem (volume plans) - Message communications costs are usually based on volume with minimal connect time and/or SIM management charges. Network providers normally provide a selection of data communication plans and the plan that best fits you network will need to be determined (see Network volume usage). Available plans may include: Intermittent transmission or "telemetry" (suitable for event reporting); Moderate usage 10s of MegaBytes per month (suitable for single stations); 100s of MegaBytes per month (suitable for AWS networks). Some network operators my support special negotiated volume rate, plans based on your special requirements, but take note of any special conditions like minimum volume per transmission, or minimum volume per event and time period. Typical monthly charges can be as follows (in NZ currency):
  - Network management fee - per network - hundreds of dollars
  - Virtual private network management fee - per network - hundreds of dollars
  - Data volume plan - per network - hundreds of dollars for 100s of MegaBytes
  - Excess usage over the volume plan - per network - dollars for MegaByte
  - Telemetry plan - per station - tens of dollars for kiloBytes
  - SIM management fee - per station - few dollars
- Network volume usage - The calculation to determine the volume of characters transmitted from each remote station per time interval will include:
  - Message size (including spaces and line terminators);
  - Network overheads per message (e.g. 32 characters per message for network transport);
  - AWS minimum volume per transmission per time interval (billing requirement) i.e. if a message is sent then a minimum volume charge will be made for the billing time interval starting from the send time.
  - Any user initiated communications with remote stations (normally small or 0);
- Network setup costs - To support cellular IP messaging the central data processor will need to interface to the Internet or the network provider servers. Whether a VPN is implemented or not, the communications pathway to the central data processor may at some point use the public Internet. This may provide an access point for outsiders to try and connect into or disrupt the integrity of the central data processor. Security measures must be put in place and these should include at least setup of the Internet gateway (switch) to accept only messages from valid IP source addresses and ports, and firewall software on a separate processor that only allows in messages with valid formats. The costs to be considered will include:
  - Network connection (hardware/firmware) - hundreds to thousands of dollars
  - Virtual private network setup - hundreds of dollars
  - Network switch setup - staff or contractor time, and documentation
  - Network firewall software - staff or contractor time, and documentation
- Central data processing - New software will be required on the central data processor to receive and process messages, and to store and provide access to the new data for the creation of products. Where a network is being progressively upgraded the central data processor will need to support the existing and new message delivery and processing systems in parallel. This will require a software switch that defines, for each AWS, how messages are received and processed. For the new communications system the following software modules will be required:
  - Message ingest, decoding, validation and control (software switch to turn the station on/off)
  - Station data element control (software switches to turn each data parameter acceptance on/off)
  - Data storage for online real time product creation and long term archival
  - Processing to create longer period data values i.e. averages, maxima, minima, accumulated totals. This processing will need to allow for missing data values e.g. Allow 1 missing minute for 2 minute derived values; 2 missing minutes for 10 minute derived values; 6 missing minutes for 1 hour derived values; 6 sequential and 60 non-sequential missing minutes for 24 hour derived values.
  - Create existing message formats and input these to their existing processes
  - Create new products and delivery mechanisms
  - Automated monitoring and alarms (see Monitoring)
- Monitoring - Although the communications network provider normally performs their own monitoring it is also useful to have a separate full network traffic monitor at the central data processor ingest server. To monitor the

correct operation of each station requires information to be extracted at various stages in the message ingest process, and due to the large volumes of high rate data involved this process must be automated. It may be useful to modularise this process so that statistics and fault information is placed in files that are read by a separate program which raises warnings and alarms when appropriate. Software implementation could be as follows:

- Server - Volume monitoring of the main pipe for network wide faults and billing.
- Firewall - Volume monitoring per modem IP number for station communications faults (stream volume too high or low) and billing.
- Message ingest - Message statistics per AWS for missing station and missing/corrupted messages.
- Message decoding - Extraction of AWS house keeping data (e.g. power supply, equipment temperature) and sensor status (alarms, missing sensor, missing data, data quality indicators).
- Alerting - Read all status files, check for fault conditions, raise warnings and alarms (on screen, emails, cellular text, pagers), and produce routine performance summaries.

- Network Tools - Tools to allow technicians to perform remote diagnostics and maintenance should be coinsidered.
  - PING - An IP network tool that can be used to establish that a remote modem is operating and is contactable (the network provider must support the PING protocol and the modem should use fixed IP addressing).
  - Some form of station remote access, particularly for more complex stations that incorporate aviation sensors.
- AWS upgrade costs - To upgrade existing AWS the following will need to be considered:
  - Cellular IP modem and cables - several hundred dollars
  - SIM activation (per modem) - 0 to tens of dollars
  - AWS software upgrade - staff or contractor time, and documentation
  - AWS software and hardware installation - staff or contractor time, and documentation
- GPRS modem setup and control - A GPRS modem must be pre-programmed with the appropriate cellular IP network provider access information and the message destination IP address and port number. Once the modem modem is powered on and the initial connection is established with the IP network then the connection (or context) is always available (providind the link is physically available) i.e. there in no further need for modem control or dialling whenever a message is sent.

  The required network operator information is: ***Internet or VPN APN*** (e.g. Internet); ***Username*** if required (e.g. aname); ***Password*** if required (e.g. thepassword); ***message source IP addressing*** (this will be a fixed address range for modems on a VPN, and a range for dynamic modem IP addressing. e.g. 123.45.67.890/123).

  The required receiving system information is, (web facing server) information is: ***destination IP address*** (e.g. 123.45.67.890); ***destination IP port number*** (e.g. 123456).
- Backup System - At first glance it would seem beneficial to have a hot backup system in place e.g. POTS toll calls to a backup computer system or POTS dial up to Internet Service Provider and then email or UDP, etc. The decision whether to implement such a backup system is based on the cost to risk beniefit. The cellular IP network reliability and network operator track record will need to be considered compared to the cost of installing a dual but hardly ever used communications system. An alternative method is to build diversity into the network by strategically spreading AWS between physically different cellular networks (in NZ the options would be Vodafone GPRS and Telecom CDMA).
- Options where there is No Cellular Coverage - There will be areas where there will be no cellular IP coverage even when using high gain antenna so alternatives will be required. These can be catagorised as follows:
  - Extend the existing netork coverage by: Talking to the network operator about cell site pattern enhancement or mini-cell sites; Use a radio modem link to the nearest coverage point (extends coverage by up to 50 km).
  - Alternatives for 1 minute interval messages: On-line hard wire or wireless Internet access; Third party Internet access (tap into an existing local network); Leased lines and dedicated on-line modems.
  - Alternatives for greater than 1 minute interval messages: Retain existing communications (if not being phased out); Dial-up internet access with Email, UDP, TCP, etc.; Dial up POTS; Satellite; HF packet radio.

Whether to Upgrade or Not

Upgrading any network can involve significant infrastructure costs which need to be justified before the project to upgrade proceeds and the new method needs to satisfy some or all of the following: alleviate threats; save money; generate new revenue. The pros (benefits) and cons of upgrading the existing MetService network to cellular IP communications were considered during the early planning phases and the following points were considered:

Pros (Benefits)

- <u>Limited life of PacNet</u> - The communications system used by most of the MetService network of AWS ia a dial up POTS to a PacNet gateway. The PacNet service provider in NZ has signalled that this network has a limited life in the order of only a few years. This posed a threat to MetService's data collection operations and so alternative communications methods had to be found. The GPRS network alleviates this threat at all sites where there is coverage.
- <u>Reduction in communications costs</u> - The monthly cost of 1 minute message delivery vis GPRS is about half the cost of a POTS circuit rental plus message volume costs of PacNet so it will be possible to reduce communications costs at most sites.
- <u>Reduction in AWS complexity</u> - First, as the GPRS modem operation is simpler from the AWS point of view, the AWS can be a simpler in design and therefore lower cost and easier to operate i.e closer to a plug-and-play type system. Second, more frequent messaging, and ideally real time (once per minute), means a small and simple data set can be implemented at the AWS with post processing to produce more complicated data sets and codes being performed at the central data collection centre. This has the flow on effect that it is easier to implement and manage software upgrades for coding changes, for fixing bugs, and to produce new data products because these are performed at the central data collection centre.
- <u>Improvements in forecast accuracy and timeliness</u> - The availability of real time data to forecasters and numerical models may make it possible to produce better trending, event timimg and provide data closer to the time of forecaset issue, therefore improving forecast accuracy and timeliness. Higher time resolution data may provide a better understanding during post forecast analyses thereby paving the way to improve future forecasts.
- <u>New real time data products</u> - Real time data communication to a central data collection centre means data products can be created and delivered to multiple clients in real time. The Internet can be used for real time data delivery and therefore there are minimal limitations on client geographical location i.e. The Internet is n ormally accessible from most places using either dial-up, dedicated wire of fibre optic, cellular and wireless, or satellite. Some typical applications of real time data collection and delivery include:
  - Aviation - Air Traffic Control and ATIS, pilot briefing
  - Coast guard (voice to speech radio braodcast of conditions)
  - Port operations (current conditions and alarms)
  - Television (up-to-the-minute weather conditions)
  - Road weather (current conditions and alarms)
  - Horticulture (current conditions and alarms)
  - Consumer weather (up-to-the-minute weather conditions via phone, cell phone, internet)
- <u>New archived data products</u> - Archived real time data provides more useful event timeliness information that can be used for insurance, accident and forensic investigations.
- <u>New portable rapid deployment AWS</u> - In places where there is cellular IP coverage it becomes very cost effective (in terms of installation time and infrastructure setup) to perform short term installations for items such as: Research projects; Recreational events (land, marine and aviation sports events); Environmental accidents (chemical spills, fires); Search and rescue; Structural construction projects.

Cons

- No remote on-line dial in" access for UDP - The ability to remotely access remote stations is more difficult and requires additional commuinications protocol software. The ability to remotely call into AWS to get current data is however not really a loss because the delivery of real time data means this facility is not required. The ability to remotely call into AWS for technical access to perform detailed diagnostics and control is however a loss. The loss of this feature is however balanced by the availability of diagnostics information in real time, and that as the AWS equiment and operation is simplified there is less need for remote control. For those sites where remote technical acces is still required an option is to retain the POTS connection for maintenance purposes.
- <u>Higher volume "pipe" required for central data collection</u> - To handle the increased volume of message traffic at the central data processor more bandwith (a bigger "pipe") may be required. This must be considered when evaluating the per AWS operating costs. The size of the "pipe" may need to be reviewed periodically as the num,ber of stations changes.
- <u>Security of data</u> - When the public Internet is used for message communication then the data is insecure (can be listened to by others). Due to the perishability of the data with age (i.e. the data has less value as it time increases) this may not be an issue, however where it is an issue the use of a VPN of encryption of messages can be employed. A VPN will incur a network provider routine management cost and data encryption will require additional AWS and central data processor software complexity.

- ❑ New message formats - The requirements of a real time data stream are not the same as for much less frequent message reporting. As the new data stream costs are normally based on volumes of traffic it is more cost effective to minimise the size of messages (down to any billing thresholds) and this can be done by sending a base set of 1 minute data that can be further processed at the central data processor. This is not traditionally the way data has been communicated so a new messages code will most likely need to be defined.
- ❑ Increased complexity of central data processing software - The central data processing software will need to perform additional tasks compared to traditional hourly or less frequent data collection. This will include: decoding the likely new message code; performing data quality control; deriving additional time period data values; creating tradition message codes.
- ❑ Coverage not always available - Not all AWS may have coverage so alternative communications methods (dial up Internet, radio link, satellite) will still be required.

MetService Implementation

MetService's new low cost mSTAR system that sends CSD messages once a minute using cellular GPRS communications. mSTAR (Standard or Pro version) supports the measurement of:

- ❑ Wind Speed
- ❑ Wind Direction
- ❑ Air Temperature
- ❑ Relative humidity
- ❑ Pressure
- ❑ Precipitation
- ❑ Solar Radiation
- ❑ Soil Moisture and Earth Temperature
- ❑ 3 Additional Temperaures (any location)
- ❑ System battery voltage and cabinet temperature

The Harvest SPE GPRS modem with internal antenna (pointing down) is located in the top half of the equipment cabinet and on the very left side.

GPRS Modem

MetService uses a modem manufactured by company in NZ called Harvest Electronics (www.harvest.com). The modem is called a "GSM/GPRS Serial Port Extender" or Harvest SPE for short.

- ❑ SIM Card - The modem requires a SIM card to be installed that has been programmed by the local GPRS network provider with GPRS communications enabled and the modem IP number.
- ❑ Power consumption - The modem operates from a dc power supply voltage of 6 to 30 volts DC. This allows the modem to be powered directly from backup batteries or solar power supplies. After power up and establishment of a context the modem consumes 29mA during no message transfer and 90 mA during a message transmission.
- ❑ Antenna options - There are 3 standard modem antenna options. For good signal coverage an indoor antenna may be used; Where coverage is marginal or an external antenna must be used a vandal resistant wire rope with ground plane antenna may be used; Where coverage is even lower a high gain yagi antenna can be used.
- ❑ RS232 Connectivity - The modem is fully configurable/controllable using an extended Hayes AT command set. The modem is configured before installation so that at power on it automatically establishes a context with the

GPRS network so that all messages sent to the modem serial port are automatically forwarded to the central data processor without any control or handshaking from the AWS.

Message Codes

Before upgrading: METAR, SYNOP, 01HRx, ENGxx, DDBT After upgrading: CSD, DDBT

- METAR, SYNOP - Designed many years ago and managed by ICAO and WMO. These codes were primarily designed for hourly or less frequent manual observations within the meteorological and aviation communities.
- DDBT - **D**isplay **D**ata **B**inary **T**ransfer (MetService NZ 1990). This code was primarily designed for passing data from AWS to on site display and ATIS systems. The communications was by hard wire or radio circuit with no routine communications costs so optimising message size was not required. A typical message size for an aviation AWS with a full compliment of sensors is 500 to 600 characters. The code was designed to be tranmitted once every minute but includes pre-processed information e.g. 2 minute and 10 minute wind averages, gusts, lulls, counter-clockwise and clockwise direction variability to allow display on simple terminals.
- 01HRx - **01 H**o**ur** reported Data (MetService 1992), DLYCL - **D**ai**ly Cl**imatological Data (MetService 1992), ENG1x - **Eng**ineering Data (MetService 1999). These codes enable the passing data and engineering information not catered for in the METAR and SYNOP codes and normally included in the remarks sections.
- CSD - **C**omma **S**eparated **D**ata (MetService 2004). This code was designed with opimal message size in mind but still using simple ASCII characters for ease of interpretation and maintenance. A typical message size for an aviation AWS with a full compliment of sensors and 1 minute messaging is 200 to 250 characters. The code includes the minimum data set required to generate all post processed data values i.e. post processing to generate longer time interval data values is performed by the receiving system. The code was primarily designed for 1 minute message sending of 1 minute data values, but the code has the capability of being used for other message communication intervals such as 2, 5 or 10 minutes by including additional data items that can be used by the receiving system to calculate all traditional data values including continuous climatological records.
- BUFR, CREX - These WMO codes were primarily designed for the international exchange of observational data. They were considered but were not implemented due to: Table descriptor overhead is high for small and frequent messages; Additional processing is required at the site; Not easily read by humans.
- Message Volumes - For Vodafone NZ GPRS UDP 1 minute interval messaging the optimum message size is determined from the minimum volume per time interval of 5 kB per 20 minutes as follows...

  Optimum message size = = (5000 / 20) - 32 = 218 characters (32 is the network provider message overhead)

  There is no saving for sending messages of less than 218 characters. Messages larger than this will use up part of a bulk volume plan that could be used for other AWS so the CSD message code is the most cost effective.

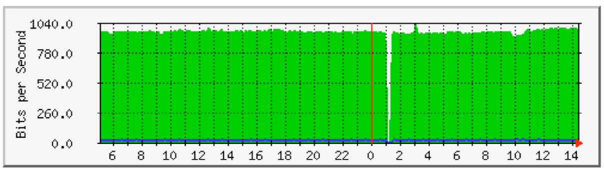
  The minimum billed volume for a month = 5 kB x 3 x 24 x 30 / 1000 MB = 10.800 MB

Monitoring and Network Tools

The whole "pipe" is monitored using network software (a scheduled network maintenance can be seen at 1am is in the Daily graph and the expansion of the number of stations can be seen in the yearly graph).

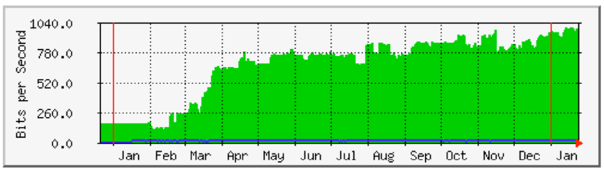
Pipe Volumes Green = Incoming traffic Blue = Outgoing traffic (bits per second)

`Daily' Graph (5 Minute Average)



Max In:1032.0 b/s (0.8%) Average In:960.0 b/s (0.8%) Current In:992.0 b/s (0.8%)
Max Out:40.0 b/s (0.0%) Average Out:24.0 b/s (0.0%) Current Out:24.0 b/s (0.0%)

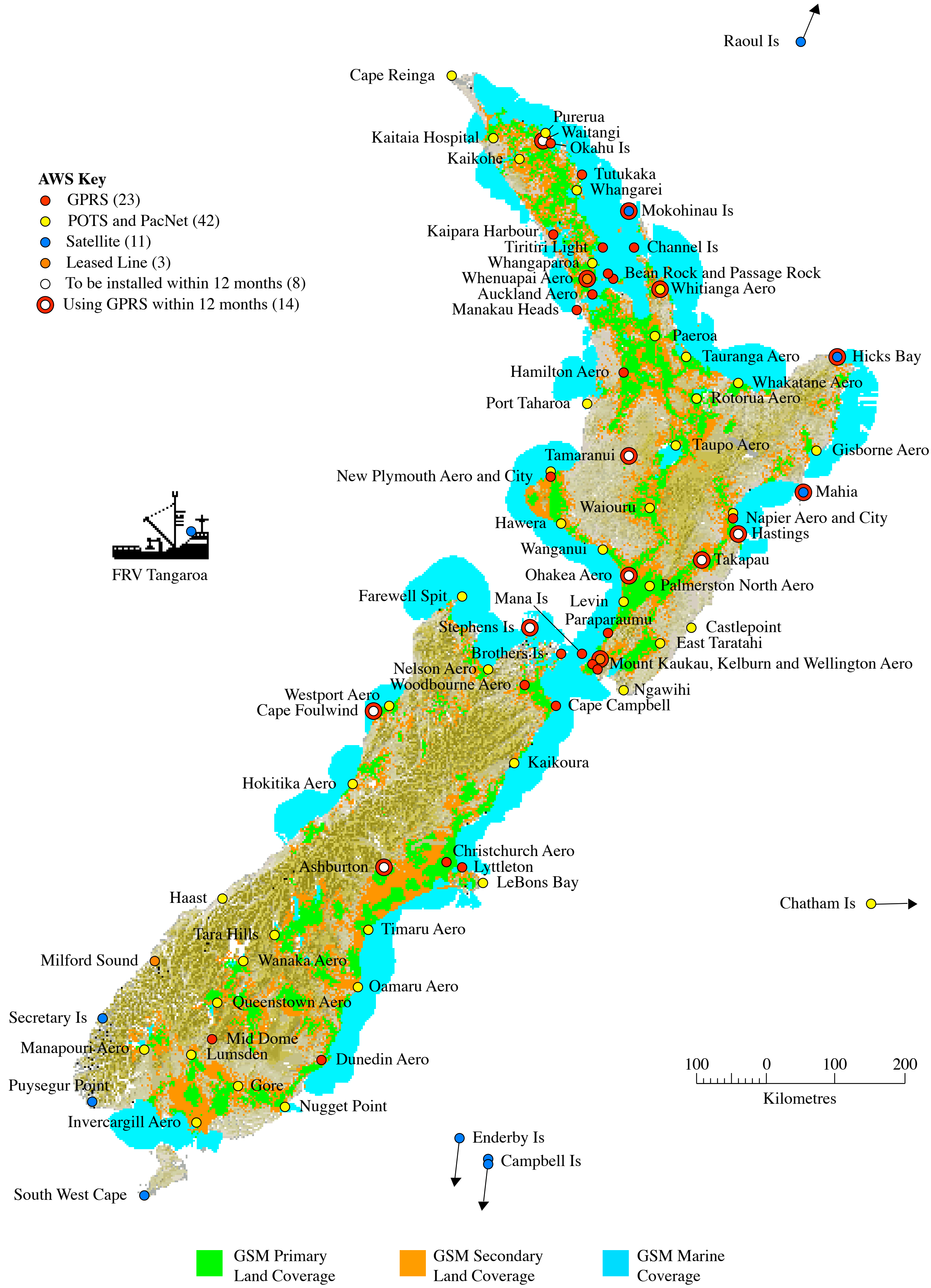
`Yearly' Graph (1 Day Average)



Max In:1008.0 b/s (0.8%) Average In:672.0 b/s (0.5%) Current In:976.0 b/s (0.8%)
Max Out:24.0 b/s (0.0%) Average Out:24.0 b/s (0.0%) Current Out:24.0 b/s (0.0%)

Ingest software counts characters from each modem IP address, and counts messages from each AWS and produces hourly and daily message counts. Message housekeeping and sensor statistics are extracted and placed in files and a separate program called "Fault Monitor" analyses all monitor files and produces email alarms when appropriate.

The Vodafone VPN supports PING. For aviation sites with ceilometers, present weather and/or visibility sensors, the POTS is retained for remote maintenance. The POTS is not used for any message sending.



MetService New Zealand AWS Nework
and Vodafone GPRS Coverage

January 2005