Journal of Global Positioning Systems (2006)
Vol. 5, No. 1-2:127-134

Transmission of RTK Corrections and Measurements using Optimal Coding

William J Kellar, Miles P Moody

Cooperative Research Centre for Satellite Systems
Queensland University of Technology
Gardens Point, 2 George Street, Brisbane, Queensland, Australia, 4000

Abstract. The provision of data services by wireless telecommunications providers has spurred interest in using these data links to transmit differential GPS corrections. Technologies such as General Packet Radio Service (GPRS) are able to satisfy the bandwidth requirements for distributing corrections in a network RTK system, however many mobile service providers charge a per-byte cost for the quantity of data transmitted. This contributes significantly to the total cost of operation for a GPRS based RTK system. Previous research by the author has shown that a significant reduction in data volume requirements can be achieved through differential entropy coding of CA code differential GPS corrections. The technique has been expanded to include carrier phase and P code pseudorange information. The paper presents the design and implementation of a data compression scheme based on differential entropy coding. The bandwidth requirement of this data format is compared with that of several other popular RTK data protocols. A study of the information content of dual frequency RTK corrections was undertaken in order to estimate the maximum reduction in data size that may be achieved through the use of optimal codes. Several polynomial prediction strategies are compared. Results show that the volume of transmitted data for carrier phase and pseudorange measurements can be reduced by as much as 90 percent when compared to RTCM 2 messages, without any reduction in precision or decimation of the sample rate.

Keywords. Differential GPS, RTK, Data Compression, Wireless, GPRS

INTRODUCTION

The research presented in this paper demonstrates a low bandwidth technique for the transmission of Real Time Kinematic (RTK) Global Positioning System (GPS) corrections and measurements. The Predictive Entropy Coding of GPS (PECOG) coding scheme makes use of several entropy reduction techniques in order to reduce the average bandwidth of RTK corrections, whilst maintaining the accuracy and latency properties of the established RTK data protocols.

This paper discusses the motivation for investigating this data compression strategy as well as the design, implementation and testing of the coding scheme. Results include a study of the entropy characteristics of GPS observables and the compression results from testing the algorithm with realtime and stored RTK data.

RESEARCH MOTIVATION

There is currently an increasing interest in using wireless Internet connections for the purpose of disseminating Differential GPS corrections to users. Mobile technology such as GPRS provides data channel capacity in excess of 50kbps, compared with the limit of 9600bps for 2nd generation mobile (GSM) connections, however the provision of these services is changing from a ‘pay per time used’ basis to a ‘pay per byte’ basis (Peterzon 2004). This research aims to take advantage of the wider available bandwidth using a burst-mode variable length coding scheme, whilst minimizing the average volume of data transmitted and therefore the cost to users.

A goal of this work is to develop a data compression scheme for RTK data that exploits the time correlation of GPS measurements. Inspiration for the coding scheme comes from predictive techniques typically used in video compression such as the MPEG coding standard (Sikora 1997), although block coding is not used since the GPS corrections in the range domain are a scalar quantity per satellite.

As a secondary benefit, this research has applications for offline storage of GPS observations. Non-realtime compression of GPS data will allow a two-pass adaptive compression scheme to be implemented, leading to an additional reduction of entropy.

The existing RTCM 2, RTCM 3 and CMR data protocols that are discussed in this paper make use of an open loop model in which the reference station distributes correction data without feedback from the users. The use of the Virtual Reference Station is a technique for network based RTK solutions that requires users to send approximate location data to the reference network. The support for two-way data links for GPS receivers is already becoming popular. The proposed data format makes use of this duplex communications link whereby users can request certain information on an as-needed basis. The proposed data protocol introduces a client/server concept and is designed to be used over two way data links with strong error detection and persistent connections, such as TCP/IP.

The two-way property of the data link can be used to reduce the volume of transmitted data. The client is able to inform the server of signal interruptions and to request initialization data for differential encoding. This can be done implicitly if using a connection-based protocol such as TCP/IP for example, where initialisation data can be sent whenever a new connection is established with a user. This research assumes that the data link layer ensures that the differential data packets are delivered in order of transmission and that retransmission of missed packets is handled by lower level protocols. It is also assumed that both the client and server are made aware when connections are being established and broken.

PECOG PROTOCOL

In order to maintain compatibility with existing software and hardware, an additional application layer protocol was designed to sit between the RTCM and TCP protocols.

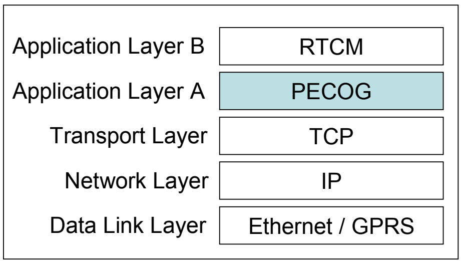

Figure 1 – Network Layers

The PECOG protocol consists of a collection of messages, known as frames, used for initialization and transmission of measurements and additional data such as measurement quality indicators and reference station properties. These data frames are known as the Initialisation Frame (I-Frame), Additional Data Frame (A-Frame) and Differential Frame (D-Frame). Separate frames are transmitted for L1 and L2 frequencies, with the L2 data being referenced to the last L1 message for that satellite. If L2 data for a given epoch is received by the encoder before the L1 data for that epoch, then it is delayed until the L1 data is received.

At the beginning of data transmission, the server sends an I-Frame for each satellite to initialize the predictor at the client. The I-Frames are analogous to the key frames used in many video compression techniques. The I-Frame contains several of the previous measurements received from the reference GPS receiver via the server, with the number depending on the order of the predictor used for that observable. The I-Frame is also sent whenever a new satellite becomes visible. The initialization information for only the new satellite is included. The PRNs of the satellites included in the frame are encoded as a 32 bit number, with each bit either true or false, depending on the presence data for a particular satellite in the frame. The fields following the Satellite Number Mask are transmitted in the order inherent to the mask, with the least significant bit corresponding to the first possible PRN.

The A-Frame contains the non-measurement data such as the satellite list, cycle slip count, data quality and multipath estimates. This frame is transmitted only at startup and when the contents change. The A-frame message does not make use of any specific entropy reduction techniques because the data contained within the frame does not change frequently enough to justify further optimisation.

The most frequently transmitted frame is the D-Frame. These frames are analogous to the ‘P-Picture’ frames used in MPEG-1 video coding in the sense that they are coded with reference to previously coded samples (Sikora 1997). This technique has been used in the past for CA code corrections (Hegarty 1992), and compression of RINEX data (Hatakana 1998).

The time series differences in this message use a high order prediction based on the rate of change from several of the previous measurements. Each D-Frame contains the Huffman coded difference between carrier phase or pseudorange measurements and the predicted values. Upon decoding a D-frame, the auxiliary data from the most recent A-frame is attached to the output message. The reference time for the D-Frame is transmitted only with the first frame for each measurement epoch. The selection of prediction method and Huffman codeword dictionary used for encoding the D-Frames is shown the results section of this paper.

Once the decoder at the client has been initialized, having received the appropriate I-Frames and A-Frames, the server transmits the quantised prediction residual for each epoch in the corresponding D-Frame. An example of the sequence of frames is shown in Figure 2.

Figure 2 – Example Frame Sequence

CODING SCHEME IMPLEMENTation

In order to implement the coding scheme, it was first necessary to select the most appropriate predictor for each of the required GPS observables.

A study was undertaken to measure the entropy of GPS carrier phase and pseudorange measurements and corrections in order to determine the optimum data compression ratio that may be achieved through entropy reduction coding.

Over 200 hours of dual frequency RTK data was collected from the QUT GPS reference station. The station consists of an Ashtech choke ring antenna and an Ashtech Micro-Z CGRS GPS receiver, synchronized to a 10MHz clock signal from a Symmetricom Cesium atomic clock.

The data was collected in RTCM SC-104 Version 2.3 format at 1 second intervals. The analysis was performed both on RTK measurements (RTCM 2.3 Messages 18/19) and RTK Corrections (RTCM 2.3 Messages 20/21). For the purpose of the entropy analysis, the RTK messages are considered to be a stream of synchronous, discrete-time, discrete-value symbols with the pseudorange quantized to 2cm intervals and the carrier phase quantized to 1/256 cycle intervals.

The quantized symbol stream was input into the predictive encoder shown in Figure 3.

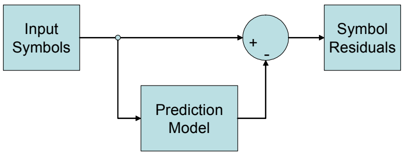

Figure 3 – Predictive Differential Encoder

Five prediction models were used in addition to the non-predictive case (null predictor). These were the differential (N=1), linear (N=2), quadratic (N=3), cubic (N=4) and quartic (N=5). Each predictor used the weighted sum of the previous N epochs as the prediction for the correction or measurement for a given satellite. In the case of the null predictor, the prediction is always zero, resulting in the input symbol being encoded without any change.

The weighting coefficients for the predictors were derived before the analysis was conducted, although it may be possible to gain a further decrease in entropy by adaptive calculation of the predictor coefficients.

Each satellite was considered separately for the generation of symbol residuals, but the outputs were collated after all satellites had been processed. The L1 and L2 frequencies were processed separately for the entropy analysis, in addition to the difference between the measurements of the two frequencies.

A frequency tally was used to calculate the entropy of each symbol residual. The average entropy of the signal was calculated using the following:

$$H = -\sum_{s} P_s \times Log_2\left(P_s\right) \quad (1)$$

where: H is the average signal entropy in bits/symbol

P_s is the probability of symbol s

The entropy of the predicted symbol residuals was used to select the most appropriate predictor for the encoder. The predictor selection is shown later in this paper.

The symbol residuals are compressed using Huffman coding. The implementation of the Huffman component required four different Huffman trees with different structures – one for each of the different GPS observables transmitted. The Huffman trees are generated using the algorithm described in Lynch (1985) using the symbol probabilities identified in the Results section of this paper.

Memory and computational requirements imposed a limit on the range of the Huffman encoded symbols. The input range of the correction data is +/-2^{31} and therefore the number of possible input symbols for the encoder is 2^{32} for each observable. It would not be practical to build a Huffman tree of size 2^{32} elements, even if gigabytes of RAM could be affordably placed on a GPS receiver. Hankamer (1979) presented a technique by which the memory requirements for Huffman coding can be significantly reduced. The symbols outside a finite range are grouped and the collective probability summed to calculate the probability for an "out of range" symbol. When a symbol outside the range must be transmitted, the "out of range" symbol is transmitted, followed by the non-coded version of the symbol.

Upon decoding, the encoded Huffman bit stream is read into a buffer and each of the four Huffman decoders removes only as many bits from the buffer as is necessary to produce a decoded symbol. The unused portion of the buffer is then converted to either an absolute quantity (in the case where an out-of-range symbol is received) or the next Huffman coded symbol.

The implementation of the coding scheme revealed several issues that needed to be considered for the successful operation of a system based on these techniques.

The use of an N-order predictor causes an initialization delay at the server such that transmission of corrections may not begin until at least N corrections have been received by the server from the reference station. This would present an N-second delay if the server and client were to be started synchronously, however this would not be the case for a continuously operating reference station, and would be an infrequent occurrence in the case of a mobile or temporary base type system.

The differential coding requires the exclusive use of integer mathematics for the predictive and residual calculations in order to avoid rounding issues caused by limited precision for floating point numbers. Integer maths also allows a minor speed increase when compared with that of floating point operations.

Careful consideration was given to the synchronization between L1 and L2 measurements, since the L2 data is transmitted relative to that of L1. The compressed L2 corrections cannot therefore be generated until input messages have been received for the L1 frequency for the given satellite. Consideration must be given to the situation where tracking of a satellite by the reference station may be lost on L1 but not on L2.

TESTING OF CODING SCHEME

The implemented compression scheme was tested in two modes known as ‘offline’ and ‘online’ tests. The first test was the offline mode in which previously stored RTCM 2.3 data was compressed using the PECOG method, then decompressed and the resulting decompressed messages compared to the input (see Figure 4). A large set of data collected over a period of a week was processed along with several smaller data sets of 48 hours each.

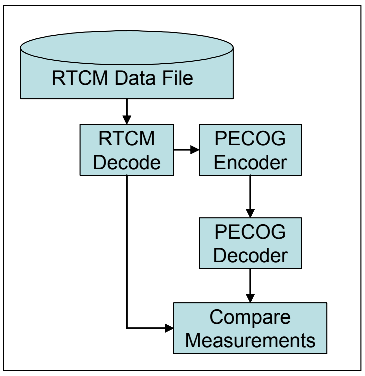

Figure 4 – Offline Test Setup

The second test involved the realtime transmission of the PECOG encoded messages over a mobile TCP/IP data link, conversion back to RTCM format messages and sending the reassembled messages to an unmodified RTK capable GPS receiver (see Figure 5).

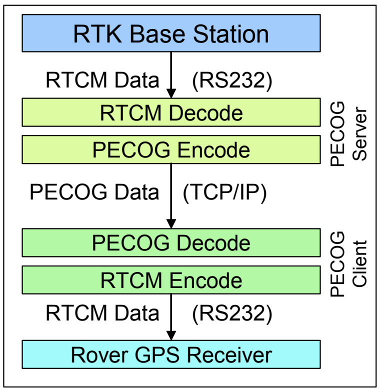

Figure 5 – Realtime Test Setup

The online positioning test was performed using dual frequency RTK using the QUT GPS reference station and an Ashtech Z-Surveyor rover receiver and a choke ring antenna over a short baseline of approximately 2 meters. The data link used for the test was a Voxson GPRS modem.

entropy results – RTK measurements

The entropy measurements for several data sets were compared in order to select the most appropriate predictors for implementing the coding scheme.

The first data set used in the entropy study consists of Uncorrected Pseudorange Measurements and Uncorrected Carrier Phase Measurements (RTCM message types 18 and 19). The results are presented in Figures 6 and 7.

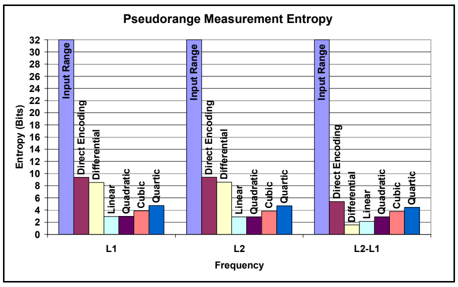

Figure 6 – Entropy of Pseudorange Measurements

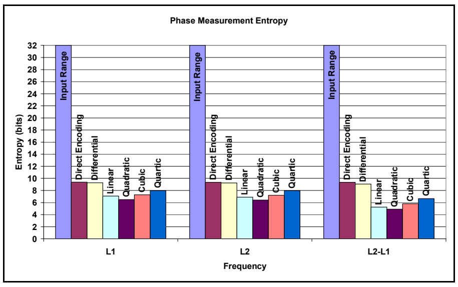

Figure 7 – Entropy of Carrier Phase Measurements

It can be seen that the linear and quadratic predictors produce the lowest entropy values for L1 and L2 frequencies for range measurements. The lowest entropy for the pseudorange difference between L1 and L2 is produced by the differential predictor.

For carrier phase measurements, the quadratic predictor shows the lowest entropy for both the L1 and L2 frequencies, as well as for the difference between the measurements on the two frequencies.

It should be noted that the entropy reaches a minimum and then starts to increase again as the predictor order increases. This increase in entropy is partially due to an increase in the weighted sum of the quantization error and measurement noise, as more data points are used in the calculation.

entropy results – RTK CORRECTIONS

The second group of data consists of pseudorange corrections and carrier phase corrections (RTCM 2.3 Message Types 20 and 21). The results are shown in Figures 8 and 9.

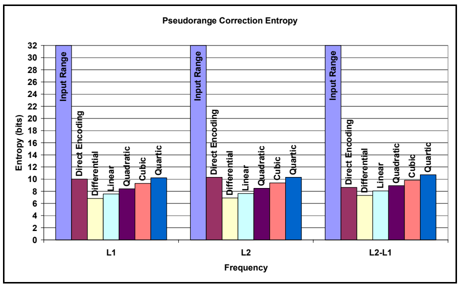

Figure 8 – Entropy of Pseudorange Corrections

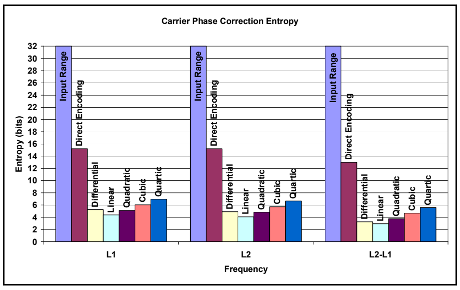

Figure 9 – Entropy of Carrier Phase Corrections

The differential predictor produced the lowest entropy for the pseudorange correction messages although the entropy values were significantly larger than those achieved by the first group of measurement data.

The carrier phase corrections achieved the lowest entropy when using the linear predictor. The entropy values were smaller than those of the first group of measurement data.

The best prediction models for the sample data were chosen to be those with the minimum entropy and are shown in Table 1.

Table 1

Quantity	Frequency	Best Predictor	Entropy
Range Measurement	L1 L2 L2-L1	Linear Linear Differential	2.929 2.895 1.578
Phase Measurement	L1 L2 L2-L1	Quadratic Quadratic Quadratic	6.497 6.422 4.896
Range Correction	L1 L2 L2-L1	Differential Differential Differential	6.842 6.910 7.334
Phase Correction	L1 L2 L2-L1	Linear Linear Linear	4.378 4.079 2.939

Whilst the quadratic predictor showed the lowest entropy for the L2 range measurements, the linear and quadratic entropy results were within 0.6% of each other. The linear predictor was therefore chosen over the quadratic predictor in order to reduce the initialization time.

It can be seen from the data that, with the exception of the L2-L1 pseudorange, the minima for the correction data sets are consistently one predictor order lower than the minima for the measurement data sets. This is because the range difference over time due to the relative motion between the reference receiver and each satellite can be considered linear over the sample period and is cancelled by the differencing in the correction messages.

Project requirements meant that the transmission must include both carrier phase and pseudorange information for both L1 and L2 frequencies. Observations could have been transmitted either as absolute measurements or as corrections. The L2 data could have been transmitted as a single quantity, or referenced to the L1 data.

The smallest average message size per satellite is achieved through the transmission of measurements using a linear predictor for the L1 range measurements, differential encoding for the offset L2 range and a quadratic predictor for both the L1 carrier phase measurement and offset L2 carrier phase. This produced an average combined size for all four fields of 15.90 bits per epoch per satellite.

The large training data set used to estimate the statistical properties of the GPS observations consists of dual frequency RTK measurements collected over a duration of one week.

Outlier symbols which appear fewer than 5 times in the training data set are excluded from the calculations, as are symbols outside the range -10^4 to $+10^4$ caused by lengthy signal interruptions. The width of the Huffman coded band is set to 3 standard deviations either side of the mean and the limits are shown in the following table.

Table 2

	Sample Mean	Standard Deviation	Coded Region
L1 Pseudorange	2.8208	2.1720	[-4..+10]
L2-L1 Pseudorange	-0.0004	0.7356	[-3..+3]
L1 Phase	-0.0110	27.5031	[-83..+83]
L2-L1 Phase	0.0029	6.3285	[-19..+19]

The sample means and standard deviations of the residuals are used to calculate the symbol probabilities used in the construction of the Huffman trees.

IMPLEMENTATION TEST RESULTS

The data compression results for both the offline and online tests both converged to similar values. The large data set for the offline test achieved an overall compression ratio that provided a bandwidth saving of 90.81% when compared with the re-assembled RTCM 2.3 messages.

The PECOG message sizes were also found to be significantly smaller than the calculated sizes for the equivalent RTCM 3.0 and CMR messages (Chen 2005, Talbot 1996).

The PECOG format produced an average data output rate of 290 bits per second for dual frequency carrier phase and pseudorange observations. This is an improvement on the 600 bps for the SNUR-2000 v2.2 format described by Kim et. al. (2004) without the added data latency or decimation of the pseudorange sample rate, although the SNUR-2000 format may be more appropriate for certain applications where the data link is less robust than TCP/IP.

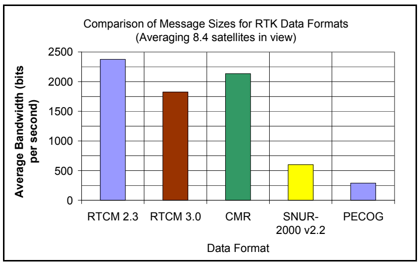

Figure 10 – Comparison of RTK Data Format Sizes

The D-Frames made up most of the compressed data transmitted, with A-frames and I-frames making up only 1.11% of the data. The average size of the D-Frames was 8.98 bytes, which included an average framing overhead of 3.25 bytes plus an average of 5.73 bytes for the variable length data. This equates to an average of 5.46 bits of variable length information per satellite per observable, assuming an average of 8.4 satellites per epoch. A complete set of measurements for L1 and L2, carrier phase and pseudorange data averaged 21.85 bits of variable length data, which is slightly larger than the calculated entropy value of 15.90 bits from the earlier section of this paper. The minor increase can be attributed to the padding bits added to the variable length data for each observable in order to align with a whole byte boundary.

The algorithm was found to be robust under these conditions, however several issues were revealed. Only the first measurement for a satellite is transmitted in the case where the same satellite appears more than once within a message for a given epoch. This is an uncommon event which occurs when the reference receiver changes from tracking the C/A code to tracking the P code. The effect of this only lasts for a single epoch.

The offline test also revealed a problem when the predictions approached the maximum allowable value for a signed 32 bit integer. This occurred when a satellite resumed tracking after a long break due to integer cycle slips or when the L2 carrier phase data was significantly different from that of the L1 data due to the integer ambiguity. This problem was remedied by forcing retransmission of an I-Frame for that particular satellite in such an instance.

The online test produced similar compression results to the offline test with a data bandwidth saving of 90.5%. The 2DRMS position error was 5.8mm and 7.7mm for the PECOG correction data sets, compared to the 2DRMS error of 7.4mm and 8.6mm achieved by the same receiver using the data sets that used native RTCM corrections.

The mean age of the RTCM corrections was 2.10 seconds, compared to 1.54 seconds for the PECOG corrections.

Figures 11 and 12 show the fixed-ambiguity position errors relative to the mean of the combined data for uncompressed RTCM and PECOG corrections respectively.

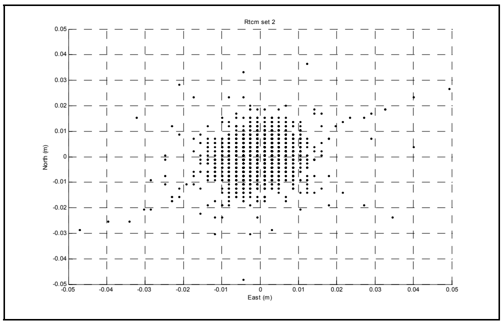

Figure 11 – RTCM Realtime 2D Position Error

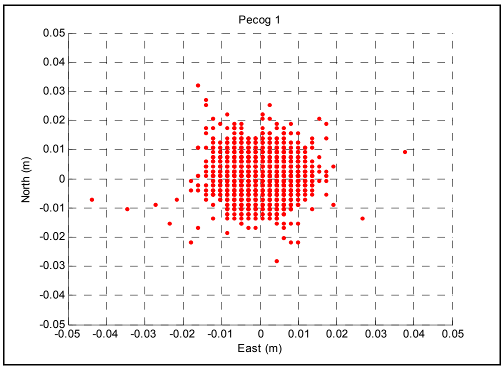

Figure 12 – PECOG Realtime 2D Position Error

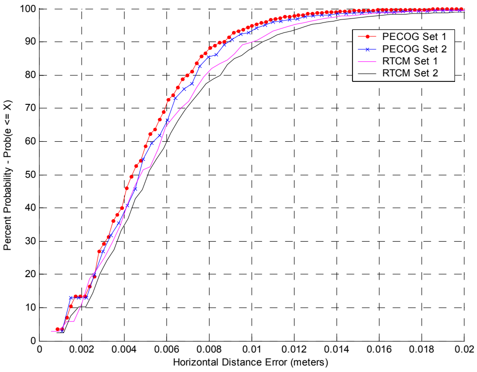

Figure 13 – 2D Position Error verses Probability

Figure 13 shows a marginal improvement in the horizontal position error for the PECOG data sets when compared with the RTCM corrected data. This is most likely due to the reduction in the average correction age when using the PECOG corrections.

CONCLUSIONS

The work presented in this paper has shown that predictive entropy coding of GPS RTK messages can achieve significant bandwidth savings for applications with persistent and reliable data links such as TCP/IP. The coding scheme achieved these savings whilst maintaining positioning accuracy comparable with the established RTCM 2.3 standard messages. The lower bandwidth requirements will enable RTK operations over low-rate data links such as trunked radio and lead to significant cost savings for RTK operations over 'pay per byte' data links such as GPRS.

The successful online realtime test verified that the PECOG protocol can be used for dual frequency RTK.

A secondary benefit of the reduced data packet sizes is a reduction in the latency of corrections caused by the GPRS packet scheduling. This may explain the incremental improvement in the positioning error. Further research will investigate what is the best correction packet size for fastest transmission over the GPRS link.

Further work will investigate an adaptive filter for the RTK measurements to further reduce the entropy of the prediction residuals. Such a technique will include dynamic recalculation of predictor coefficients and the Huffman codebook, and the subsequent realtime distribution of these to users.

ACKNOWLEDGMENTS

This work was carried out in the Cooperative Research Centre for Satellite Systems with financial support from the Commonwealth of Australia through the Cooperative Research Centre program, the Australian Research Council, the Queensland State Government and the Queensland University of Technology.

The authors would also like to thank GPS Online and their staff for their financial and in-kind support of this research.

REFERENCES

Chen X. et al. (2005) ***On the Influence of Different Network Correction Formats on Network RTK Performance***, Proceedings of ION GNSS2005, Long Beach, California, September.

Hankamer M. (1979) ***A Modified Huffman Procedure with Reduced Memory Requirement***, IEEE Transactions on Communications, Vol COM-27, No. 6, pp930-932.

Hatakana Y. (2000) ***Compact RINEX Format***, Geographical Survey Institute, Japan

Hegarty C.J. (1992) ***RTCM SC-104 Data Compression***, Proceedings of ION GPS '92, Albuquerque, New Mexico. September, pp501-502.

Kim J. et. al. (2004) ***Budget Analysis and Realization of Precise RTK-GPS Via Compact Correction Messages Applied to Commercial GPS Receivers***, ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, California, 21-24 September.

Lynch T.J. (1985) ***Data Compression Techniques and Applications***, Van Nostrand Reinhold, New York, USA, pp55-56.

Peterzon M. (2005) ***Distribution of GPS-data via Internet***, Masters Thesis, Uppasala University, Gavle, Sweden, pp29

RTCM Special Committee No. 104 (2001) ***RTCM Recommended Standards for Differential GNSS (Global Navigation Satellite Systems) Service***, V2.3, RTCM Paper 136-2001/SC104-STD, August 20, 2001.

RTCM Special Committee No. 104 (2004) ***RTCM Recommended Standards for Differential GNSS (Global***

Navigation Satellite Systems) Service, V3.0, RTCM Paper 30-2004/SC104-STD, 10th February.

Sikora T. (1997) ***MPEG Digital Video-Coding Standards***, Signal Processing Magazine, Volume 14, Issue 5, IEEE, September, pp90.

Sklar B. (1988) ***Digital Communications Fundamentals and Applications***, The Aerospace Corporation, El Segundo, California, Prentice Hall, New Jersey, pp629.

Talbot N. (1996) ***Compact Data Transmission Standard for High-Precision GPS***, Proceedings of ION GPS '96,Kansas City, Missouri, The Institute of Navigation, Washington, D.C., Sept. 17-20, pp861-871.

Williams R.N. (1991) ***Adaptive Data Compression***, Kluwer Academic Publishers, Boston, USA.