5.19

COMPRESSION OF NEXRAD (WSR-88D) RADAR DATA USING BURROWS-WHEELER ALGORITHM[1]

Steven D. Smith
Radar Operations Center, National Weather Service, Norman, OK.

Kevin Kelleher
National Severe Storms Laboratory, Norman, OK.

S. Lakshmivarahan
School of Computer Science, University of Oklahoma, Norman, OK.

1. INTRODUCTION

The National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD) recently completed installation of 158 WSR-88D Doppler weather radars (NEXRAD) throughout the U.S. These radars provide high spatial and temporal resolution radar data. The Cooperative Radar Acquisition Field Test (CRAFT) project was established to begin addressing the need for real-time access of the data by an ever-growing user community. The principle goal of CRAFT is the development of a national prototype demonstrating the real-time compression, delivery, and Internet-based transmission of WSR-88D Archive Level II data from multiple radars with a view toward nationwide implementation [2].

Presently radar data are archived on-site using 8 mm tape. Once a tape is filled, it is sent to the National Climatic Data Center (NCDC) for permanent archival. The main disadvantages of this technology are the unreliability of the tape drives, the expense of maintaining the on-site equipment, and the reliance upon NWS technicians to change the tapes and mail them to NCDC. A secondary goal of CRAFT is to eliminate the need for in situ archival of radar data. In essence, radar data collection would shift from on-site to a designated off-site centralized repository, NCDC.

A typical radar generates data at rates up to 45 Kbytes/s or over 1 terabyte/year. In order to reduce the bandwidth requirements necessary to transport these data off-site (and the associated communications costs) and reduce storage requirements for long-term archival at the central repository, compression of the data is absolutely necessary.

There are many compression algorithms available (e.g., dictionary-based, probability-based)[3]. Without knowledge of the structure of the data, it is nearly impossible to objectively determine which algorithm(s) work well on radar data. In CRAFT, the freeware compression algorithm known as *bzip2* [4] is being used in the operational prototype. *bzip2* is a member of the class of hybrid algorithms based on the Burrows-Wheeler transform [1]. Is this one of the better algorithms for use in compressing radar data? If it is one of the better algorithms, why does it perform so well? How does the performance of *bzip2* compare to the popular UNIX *compress* algorithm, a dictionary-based scheme?

Properties of radar data collected at different radar sites, at different times of year, under different types of weather conditions, and using different radar scanning strategies were examined to answer these questions [5]. Our goal is to recommend a compression algorithm or algorithms to meet the dual requirements of high compression ratio (CR) and low space/time complexity. Although not a requirement, it is desirable to choose an existing (i.e., off-the-shelf) algorithm which is freely available with a limited copyright agreement. Almost equally important is the algorithm should be easy to integrate into the NEXRAD radar network. The popular off-the-shelf freeware program *bzip2* satisfies these criteria.

2. CHARACTERIZATION OF RADAR DATA

---

[1] This research was funded, in part, by a grant from the National Oceanic and Atmospheric Administration's Environmental Systems Data and Information Management (ESDIM) program.

Like the weather, weather radar data has spatial and temporal variation. To help assess the performance of compression algorithms applied to radar data, data from different radars, at different times of the year, for different types of weather, and collected using different scan strategies were examined. The following sections describe the various weather event examined.

2.1 Spring Low-Level Jet Event

There are two data sets representing the nocturnal low-level jet event which occurred in central Oklahoma on 05/01/92. One set was collected at the NEXRAD Radar Operations Center (call letters KOUN), the other at the National Weather Service's radar located at Twin Lakes (call letters KTLX). The KOUN data set consists of 10 volume scans collected in VCP 31 starting at 04:34Z. The KTLX data set consists of 10 volume scans collected in VCP 32 starting at 02:30Z. Because of the close spatial and temporal proximity of the two data sets, this allowed us to compare the data characteristics of the same event using different scan strategies and radar operating characteristics.

2.2 Severe Storm Squall Line Event

The data for this squall line event was collected at Melbourne, Florida (call letters KMLB) on 03/13/93. The data set consists of 10 volume scans of VCP 11 starting at 09:26Z. The squall line was large and intense producing much severe weather, including tornadoes, damaging winds, and flash flooding. Events of this intensity over such a large geographic area are rare and are believed to present challenges to compression algorithms owing to the presence of large horizontal gradients in the moment data.

2.3 Heavy Stratiform Precipitation Event

The data for this precipitation event was collected at Houston/Galveston, Texas (call letters KHGX) on 10/18/94. The data set consists of 10 volume scans of VCP 21 starting at 20:04Z. Northwest of the Houston/Galveston area is a weak line of storms preceded by a large area of stratiform rain. Heavy rains were reported with these storms. This event was selected to represent the typical heavy stratiform precipitation event. Such events are common throughout the U.S. in all seasons of the year.

2.4 Hurricane Event

The data for this hurricane event was collected at Eglin Air Force Base, Florida (call letters KEVX) on 08/03/95. The data set consists of 10 volume scans in VCP 11 starting at 13:48Z. This is hurricane Erin that struck South Florida. At least 4 weak (F0) tornadoes were reported around the time of data collection along with strong winds and very heavy rain. This event was selected because of widespread radar return.

2.5 Light Stratiform Precipitation Event

The data for this stratiform precipitation event was collected at Sacramento, California (call letters KDAX) on 02/19/96. The data set consists of 10 volume scans in VCP 21 starting at 08:24Z. There is nothing remarkable about this event and it is believed to be fairly representative of light stratiform precipitation events in general.

2.6 Winter Clear-Air Event

The data for this wintertime clear air event was collected at Amarillo, Texas (call letters KAMA) between 01/27/99 and 01/28/99. The data set consists of 9 volume scans of VCP 31 staring at 23:53Z. This data set was chosen to represent the typical wintertime clear-air event. In the winter months with high atmospheric pressure the atmosphere is generally dry and quite stable. For such an event we do not expect significant radar return both in intensity and areal coverage.

2.7 Spring Clear-Air Event

The final data set examined is a spring clear-air event that occurred at Aberdeen, South Dakota (call letters KABR) on 05/29/96. The data set consists of 9 volume scans of VCP 32 starting at 04:06Z. Clear-air returns were observed to several thousands to over 10K feet above ground level owing to vertical mixing.

3. IMPROVING RADAR DATA COMPRESSION

Based on results the cases examined *bzip2* appears to be a good choice for compressing radar data [6]. It generally outperforms UNIX *compress* in terms of CR by nearly a 2:1 margin. The question remains: "Can we improve upon the *bzip2* compression?" We investigated the use of data transformations and logical compression techniques in a data preprocessor to *bzip2* as a means to improve the overall compression performance. The following data preprocessing schemes were found to improve overall data compression if applied in conjunction with *bzip2*.

3.1 CTM Header/Trailer

The 16 bytes of Channel Terminal Manager (CTM) protocol header/trailer encapsulating the radial message is now obsolete yet the header/trailer continues to be carried with the radial message. Users of the data routinely discard this information prior to processing the radial message. We recommend this header/trailer be zeroed-out (vice discarding) prior to compression. Zeroing out the header/trailer avoids post-processing the data after decompression.

3.2 Padding Radial Message

In a single radial message, there are a maximum of 460 measurements of reflectivity and 920 measurements of velocity and width. The actual number of measurements is determined by the range at which the radar beam intersects the 70 Kft height ceiling. At 19.5 degrees elevation, the height ceiling occurs at a range of 70 km (bin number 70 for reflectivity measurements, bin number 280 for velocity and width measurements).

Although the radial format supports variable length moment fields through the use of pointer assignments in the metadata (and consequently variable length radial message), the radial message is fixed in length at 2432 bytes. When the number of measurements for any moment is less than the maximum allowed, the radial message data occupies less than the allotted 2432 bytes. Moment data beyond the 70 Kft height ceiling represents unuseable data or “garbage”. It has been observed that this “garbage” is often data from a prior radial message and this data adversely affects the performance of the compression algorithm.

We recommend zeroing-out the data i.e, set the data to the below SNR threshold indicator, vice supporting variable-length messages. The zeroed-out data compresses well owing to Run-Length Encoding embedded in *bzip2*, and no special post-processing is required after decompression.

We have found that application of the preprocessing techniques described above improves overall compression up to 19%, with average compression over all cases examined being 7%. Considering low overhead cost and the fact no special processing is required during data decompression, preprocessing of the data appears worthwhile. We also found compression time decreases when preprocessing is applied so the overhead cost associated with preprocessing the radar data is more than compensated for by the reduction in compression time.

3.3 Metadata Transformation and Compression

Each metadata element can be assigned membership to one of the following 4 groups or categories: 1) data which does not vary during a volume scan, 2) data which does not vary within an elevation scan, 3) data which varies in a predefined manner from radial message to radial message, and 4) data which can and often does vary from radial message to radial message. To reduce the redundancy in the metadata, we have devised the following logical compression technique: The full radial header is carried with the radial message at beginning of volume, elevation, whenever a value considered a constant changes, or whenever a value changes other than in a predefined manner. Otherwise an abbreviated radial header is carried with the radial message.

The abbreviated radial header is approximately 1/3 the size of the full radial header and consists of message type, message time, radial generation time, radial azimuth and elevation angle. After decompression, the full radial header is easily reconstructed from the abbreviated header as long as the data from the most recent full radial header are available to the postprocessor. Unlike the first two methods discussed, implementation of this method requires state information be carried with the radial message.

The following table presents the percent gain in CR using *bzip2* in conjunction with zeroing out the CTM header/trailer, padding the radial message, and transforming/compressing the radial message metadata. In all cases there is an improvement in CR.

4. Summary and Recommendations

In this paper, compression of WSR-88D radar data in CRAFT was discussed. At the present time, the freeware compression program *bzip2* is being used in the operational prototype to provide data compression. Although BZIP2 offers very good performance, ways to improve the overall performance were examined in this paper. Preprocessing radial messages by zeroing out the CTM header and trailer, padding the messages with 0's, and transforming/compressing the message metadata offered incremental improvement in overall compression using *bzip2.* For the 8 data sets examined, improvement ranged from as little as 5% to nearly 30%.

Case	***bzip2***	***bzip2*/PP**	**% Gain**
KAMA	25.4	32.6	29
KDAX	11.0	12.7	15
KABR	8.8	9.2	5
KTLX	8.0	8.4	5

KHGX	8.1	9.3	14
KOUN	10.6	13.5	27
KMLB	8.0	8.9	11
KEVX	6.4	7.1	10

Table 3.1 *bzip2* compression without and with data preprocessing (denoted by PP).

Acknowledgments

The authors would like to thank Mark Bradford, Doug Kennedy, Sai-Siu Lam, and Feng Wu for helping make this paper possible.

5. References

[1] Burrows, M. and D. J. Wheeler, 1994: A Block-sorting Lossless Data Compression Algorithm. SRC Research Report, May 10, 1994.

[2] Droegemeier, K. K., J. J. Levit, K. Kelleher, T. D. Crum, S. A. Delgreco, L. Miller, D. W. Fulker and H. Edmon, Project CRAFT: A Test Bed for Demonstrating the Real Time Acquisition and Archival of WSR-88D Base (Level II) Data. Draft for submission to Bulletin of the American Meteorological Society, Oct. 2000.

[3] Hankerson, D., G. A. Harris and P. D. Johnson, 1998: Introduction to Information Theory and Data Compression. CRC Press, Boca Raton, Florida.

[4] Seward, J., "How to Use bzip2". Available from bzip2 and libbzip2 official home page found at URL http://sources.redhat.com/bzip2.

[5] Smith, S. D., K. Kelleher and S. Lakshmivarahan, 2001: Compression of NEXRAD (WSR-88D) Radar Data Using Burrows-Wheeler Algorithm. NSSL Report.