10.8 COMPRESSION AND RELAY MANAGEMENT SYSTEM IN GRIDDED FX-NET, SYSTEM ARCHITECTURE AND USE OF WAVELET COMPRESSION

*Jebb Stewart, Sher Schranz, Ning Wang, Evan Polster
NOAA, Earth System Research Laboratory
Boulder, Colorado
In collaboration with the Cooperative Institute for Research in the Atmosphere (CIRA)
Colorado State University, Fort Collins, Colorado

1. INTRODUCTION

The Compression and Relay Management System (CRMS) is the enabling technology for the Gridded FX-Net system, a meteorological forecaster workstation being evaluated for operational use by fire weather prediction forecasters and operational control center forecasters.(Refer to Paper, S. Schranz '06 AMS IIPS 6.8 – Gridded FX-Net Prototype Project for NIFC) The Gridded FX-Net system is an extended version of the National Weather Services' Advanced Weather Interactive Processing System (AWIPS), Bullock (1994.) What sets this system apart from AWIPS is the capability of serving multiple, remote, D2D clients. The CRMS developed for this project provides an intelligent 'push' system using existing technologies such as Local Data Manager (LDM) and Wavelet Compression.

The LDM server was a key addition to the CRMS; however, modifications were made to all components of AWIPS. These changes were made to take advantage of the data handling, product generation and display manipulation capabilities of AWIPS.

Wavelet-based data compression techniques to compress both satellite images and model grids are used in the CRMS to enhance data delivery to remote, Gridded FX-Net clients. These techniques have demonstrated the ability to significantly reduce the meteorological products file size, and therefore the transmission time of the products. To apply the wavelet data compression to CRMS, several adaptations and optimizations were made in order to meet the requirements of the operational environment.

2. BACKGROUND

As the Internet continues to evolve and data sets continue to grow in size, there exists a need to distribute large raw and imagery data sets quickly and effectively. The original idea for Gridded FX-Net was built upon technology from FX-Net, Schranz (2005), and AWIPS. The goal was to maximize throughput on limited bandwidth networks and create Internet distribution of NOAAPORT data to remote D-2D clients. Many different technologies exist for compression and file transfer, but lack a central management system. CRMS was created to provide a management system for the entire process, from data ingest, through data compression and distribution.

During the design of CRMS, special consideration was made to make the system highly adaptable and configurable to take advantage of any compression technology or transfer technology. During initial development two different types of compression were used. A specially designed wavelet compression technique is used for satellite imagery and gridded data sets. With this wavelet compression, CRMS can gain higher compression ratio's and maintain data loss within National Weather Service (NWS) specifications. For other data types, including radar, metar, and other data sets standard to NOAAPORT, CRMS uses GZIP compression.

3. COMPRESSION RELAY MANAGEMENT SYSTEM

CRMS is composed of 4 parts described below; data ingest server, data discovery service, compression manager, and transfer service.

3.1 Data Ingest Server.

The data ingest server for CRMS uses existing AWIPS technology, Bullock (1994.) The data server is built using the AWIPS data ingest server software, which collects data from the NOAAPORT feed over the Satellite Broadcast Network (SBN).

3.2 Data Discovery Service

The data discovery service is responsible for finding new data files as they are stored to disk by the data ingest server and notify the compression manager. One way of discovering new files is by using the notification server technology from AWIPS. CRMS registers for notifications for data with the notification server on the data ingest server. As new data is ingested, a notification is sent out and received by the data discovery service. The data discovery service then notifies the compression manager of new data available. For other data sets to which the notification server may not apply, the data discovery service includes a directory watcher script that continually monitors directories specified by configuration for new files. As files are added to the directory the watcher then notifies

Corresponding author address:
Jebb Stewart, NOAA/ESRL/GSD,
Boulder, CO 80305
Phone:303-497-6724;
email: jebb.q.stewart@noaa.gov

the compression manager that new data is available.

3.3 Compression Manager

The compression manager is responsible for compressing files when new data is available. When the compression manager is notified of new data via the data discovery service, the data are analyzed, sorted and added to the queue. If the data set is a grid consisting of many different variables, each variable is added to the queue as its own instance. The compression manager then monitors the queue, checking data inventory and file updates to confirm data is available and complete. When the data set is complete, it is compressed and stored to disk. The transfer service is then notified that the compressed file is ready for transfer. The compressed files stored to disk are made accessible for other web based applications such as in a pull based system.

The compression manager is highly configurable allowing users to adjust compression parameters for particular data sources, types, and by variable. If no configuration is found for a specific type, defaults are used.

The CRMS system was built in a modular way and can be modified to handle most compression schemes. Currently the compression manager utilizes two forms of compression, wavelet and GZIP, and we continually analyze the best compression to use for different data types.

3.4 Transfer Service

The transfer service of CRMS is responsible for transferring the data to remote clients. After a file is compressed by the Compression Manager, the file is handed off to the transfer service. The transfer service makes the correct system calls to transfer the file. Once the file is transferred, the item is removed from the queue, completing the process.

For our prototype, we use the Local Data Manager (LDM) software from UNIDATA, UNIDATA (2005.) LDM has a product queue, which maintains copies of transferred data, allowing clients to retrieve past data if they lose Internet connectivity. LDM is a well known and widely used service throughout the meteorological community. LDM allows Gridded FX-Net clients to select data they wish to receive. By using an LDM link, a client can be configured too retrieve only certain data types, such as grids, or to receive all data processed. This is important to the principle of Gridded FX-Net as users may have only low bandwidth options available and not want to receive the entire set of processed data.

CRMS is not limited to using LDM transfer technology. Because of the adaptable design of the CRMS, with minimal modification, this service can be configured to use any type of transfer technology.

4. WAVELET COMPRESSION

The driving force allowing gridded FX-Net to maximize throughput on limited bandwidth is the use of a specially designed wavelet compression technique developed by Forecast Systems Laboratory (FSL), Wang (2002.) The wavelet compression technique originally developed for FX-Net, Madine (2002), allows us to compress gridded data sets and satellite imagery with a high compression ratio and minimal error in reconstructed files. Results of a recent study on compression technologies for weather data show specially designed wavelet compression techniques perform better than other compression technologies like JPEG2000 and GRIB1, Steffen (2003.) The wavelet compression technique developed by FSL can be used as either a lossless or lossy compression. Lossless compression achieves data size reductions without information loss. Lossy compression reduces data size with controlled fidelity loss. Fidelity of this wavelet compression can be controlled by specific error-sensitive parameters. Using these parameters, wavelet compression can be configured to compress gridded data within specific precision requirements.

The current implementation of wavelet compression can be adjusted to compress grids at a certain compression ratio, by maximum error tolerance, or by average error tolerance. By using a compression ratio parameter, the user is guaranteed a certain file size regardless of fidelity. By using either maximum error tolerance or average error tolerance as the compression parameter, users are guaranteed precision but not file size.

4.1 Wavelet Compression for Satellite Imagery

Wavelet compression is used to achieve high compression ratio's for satellite imagery. The criteria for the resulting files is subjective. The error must be visually unnoticeable and the reconstructed images must be “meteorologically useful.” Listed in **Table 1** are the compression ratios wavelet is able to achieve with different satellite imagery with criteria mentioned above. Examples of imagery will be available during presentation.

Imagery Type	***Compression Ratio***
Water Vapor	*1:50*
Infrared	1:35
Visible	1:15

Table 1: Wavelet Satellite Imagery Compression Ratios

4.2 Wavelet Compression for Gridded Data

Wavelet compression can achieve high compression ratios for gridded data. Wavelet compression can be adjusted to compress slices or layers of gridded data or entire volumes. Based on the parameters used, wavelet compression can be considered lossless based on precision requirements. The NWS has set required precision for any data compression algorithm to be considered lossless, based on the precision achieved by GRIB1 compression. **Table 2** provides a sample list of required precisions.

Variable	***Precision***	***Units***
Temperature	0.125	K
Geopotential Height	0.125	gpm
U and V wind components	0.125	m/s
Relative Humidity	1.0	%

Table 2: NWS Maximum Error Lossless Precision Requirements

Using maximum error tolerance as a parameter when applying wavelet compression, we can guarantee error values of the reconstituted grid will not exceed this error tolerance parameter. By specifying values from **Table 2** above, we can compress gridded data sets within the required precisions the NWS accepts as 'lossless'. **Table 3** lists the results of using wavelet compression on an entire temperature field (4 dimensions, x, y, z, and time) for two grids; the NAM 10km and GFS 80 km. Wavelet compression is able to achieve a 32.7 to 1 compression ratio for the NAM 10km, resulting in a much more manageable file size. Looking at the theoretical transfer rate of a ISDN line (about 128 kbps), you could transfer the resulting compressed file in just under 11 minutes.

Grid	***NAM 10km***	***GFS 80KM***
Variable	Temperature	Temperature
Grid Size (x,y,z,t)	310x250x36x29	65x43x11x21
File Size Before	323.6 MB	2.6 MB
File Size After	9.9 MB	0.40 MB
Max Error Tolerance	0.125 Degree	0.125 Degree
Resulting Ratio	32.7:1	6.5:1
Theoretical Transfer time at 128 k	10 minutes, 50 seconds	26 Seconds

Table 3: Wavelet Compression using Maximum Error Tolerance

Wavelet compression achieves a 6.5 to 1 ratio (**Table 3**) for the temperature field of the GFS 80km lower resolution grid. This is less than what was achieved for the NAM 10km because in an 80km grid there is greater variance and less correlation between neighboring values of temperature than a 10km grid. Wavelet compression is more efficient when applied to data sets with high correlation and more redundancy between values. In addition to grid point spacing, the dynamics of the field itself are important. Temperature is considered a rather smooth field in both vertical and horizontal planes because of less variance between neighboring grid point values, compared to fields such as relative humidity. Wavelet compression achieves lower compression ratios for these non smooth fields. **Table 4** list the results of wavelet compression using maximum error tolerance from **Table 2** as a the precision control parameter on relative humidity fields of the same two models. While the compression ratio is much lower than the temperature field, the resulting file size is more manageable than with no compression.

Grid	***NAM 10km***	***GFS 80KM***
Variable	Relative Humidity	Relative Humidity
Grid Size (x,y,z,t)	310x250x33x29	65x43x7x21
File Size Before	296.7 MB	1.6 MB
File Size After	32.0 MB	0.42 MB
Max Error Tolerance	1.0 %	1.0%
Resulting Ratio	9.3:1	3.8:1
Theoretical Transfer time at 128 k	35 minutes	28 Seconds

Table 4: Wavelet Compression using Maximum Error Tolerance

5. FUTURE ITEMS

To continue development on CRMS, we are developing 'pull technology' to allow clients to retrieve data. By allowing a user to retrieve data on their own schedule, we can further reduce burden on bandwidth and networks by downloading data in off peak hours. As we continue to add new data sets, including higher resolution models, text products, and local data sets, we will continue development of specialized compression techniques. Preliminary results show application of a lossless arithmetic encoder to radar imagery obtains higher compression ratio than GIF, Wang (2002).

We are also interested in exploring other transfer and network technologies. One idea is to use multicast technology, CISCO (2002), for file transfers. Multicast has the benefits of reducing network load on all paths to all clients but is limited by low adoption by routers and the time to live (ttl) specification in the Internet Protocol(IP). Files transferred via multicast must specify a ttl to prevent data packets from eternally bouncing around networks. The downside is if your client exists just beyond the number of hops specified by ttl, you will never receive the data.

6. ACKNOWLEDGMENTS

The authors wish to thank Steve Gilbert of the National Weather Service Headquarters for information on required precision for compressed data sets.

7. REFERENCES

Bullock, C. S., and U. H. Grote, 1994: FX-ALPHA: A new FSL workstation. Preprints, *Tenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Nashville, Amer. Meteor. Soc., 354-357.

Cisco Systems Inc. 2002: "Internet Protocol (IP) Multicast." http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/ipmulti.htm

Madine, S., N. Wang, E. Polster, J. Pyle, R. Brummer, 2002: FX-Net National: A Nonlocalized Internet-Based Meteorological Workstation. 18th Int. Conf. On Interactive Information and Processing Systems (IIPS), Orlando, Florida, American Meteorological Society, J339 – J341

Schranz, S., J. Stewart, N. Wang, E. Polster, 2005: FX-Net – Integrating Air Chemistry and Weather Data for Research and Operations. 21st Conf. On Interactive Information and Processing Systems (IIPS), San Diego, California, American Meteorological Society, 15.4

Schranz, S., J. Stewart, N. Wang, E. Polster, 2006: Gridded FX-Net Prototype Project for the National Interagency Fire Center. 22nd Conf. On Interactive Information and Processing Systems(IIPS), Atlanta, Georgia, American Meteorological Society, 6.8

Steffen, C. E., N. Wang, 2003: Weather Data Compression. 19th Conf. On Interactive Information and Processing Systems (IIPS), Long Beach, California, American Meteorological Society, 4.9

UNIDATA. 2005: "Unidata LDM" http://www.unidata.ucar.edu/software/ldm/

Wang, N., S. Madine, R. Brummer, 2002: Investigation of Data Compression Techniques Applied to AWIPS Datasets. 18th Int. Conf. On Interactive Information and Processing Systems (IIPS), Orlando, Florida, American Meteorological Society, J9.7