5.20

PROJECT CRAFT: A TEST BED FOR DEMONSTRATING THE REAL TIME ACQUISITION AND ARCHIVAL OF WSR-88D LEVEL II DATA

†[1,2]Kelvin K. Droegemeier, [3]Kevin Kelleher, [4]Tim Crum, [1]Jason J. Levit, [5]Stephen A. Del Greco, [6]Linda Miller, [1]Carl Sinclair, [3]Mark Benner, [6]David W. Fulker, and [7]Harry Edmon

[1]Center for Analysis and Prediction of Storms, [2]School of Meteorology
University of Oklahoma
Norman, OK 73019
[3]NOAA/National Severe Storms Laboratory
[4]NOAA/Radar Operations Center
Norman, OK 73069
[5]NOAA/NESDIS National Climatic Data Center
Asheville, NC 28801
[6]University Corporation for Atmospheric Research
Boulder, CO 80307
[7]Department of Atmospheric Sciences
University of Washington
Seattle, WA 98195

1. INTRODUCTION

The US National Weather Service (NWS) has completed the installation of 120 WSR-88D (Weather Surveillance Radar, 1988-Doppler) systems (Crum and Alberty et al. 1993) as part of its $4.5B Modernization and Associated Restructuring Development Plan. (Another 26 Department of Defense and 12 Federal Aviation Administration systems also were installed, for a total of 158 radars, 141 of which are in the continental United States.) This unique observing tool provides nearly continuous single-Doppler radar coverage across the continental United States which, when coupled with the radar's superb sensitivity, sophisticated processing algorithms, and advanced user training, has produced substantial improvements in the identification and short-term warning of hazardous local weather (e.g., Crum et al. 1998); new opportunities for quantitative precipitation estimation and improved flood forecasting (e.g., Foufoula-Georgiou and Krajewski 1995; Fritsch et al. 1998; Droegemeier et al. 2000); a deeper understanding of mesoscale weather and opportunities for its explicit numerical prediction down to the scale of individual thunderstorms (e.g., Lilly 1991; Droegemeier 1997), and interesting non-meteorological applications of radar data (e.g., Gauthreaux and Belser 1998).

Although the WSR-88D radars originally were not equipped with archival devices, the value of their unique observations for scientific research soon was recognized. An interim solution for the recording and long-term archival of full-volume, full-precision Level II data (also known as base data) was achieved by outfitting each radar with an 8 mm robotic tape cartridge recording system (Crum et al. 1993). Since 1993, the National Climatic Data Center (NCDC) has been archiving the 8-mm tapes and using the same media to provide base data to the national community. This archival and re-distribution process represented an innovative solution several years ago, though by today's standards is human-resource intensive, costly, inefficient, and unreliable (the archive for the 120 NWS radars is only approximately 65% complete). Rapid advances in data transmission, compression, and storage technologies suggest that far better solutions now are available and can be implemented with high reliability at reasonably low cost using existing technology infrastructures.

In an attempt to begin addressing the long-term needs for WSR-88D base data transmission and archival, which have been noted as critical (NRC 1999), the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma joined forces with the organizations sharing authorship of this preprint to establish the Collaborative Radar Acquisition Field Test (CRAFT) (Droegemeier et al. 1999). Funded initially by a grant from the Oklahoma State Regents for Higher Education, the NSF, and more recently by the National Oceanic and Atmospheric Administration (NOAA) ESDIM and HPCC programs (see section 10), CRAFT is an experiment in the real time compression and Internet-based transmission of WSR-88D Level II data from multiple radars. The initial test bed of six radars, located in and around Oklahoma, has been delivering real time base data since early 1999 with virtually no interruptions. During the past year, 15 radars were added, with all 21 now serving as the prototype for a national data transmission and archival system.

2. CURRENT TECHNICAL STRATEGY

Project CRAFT is based upon three technologies. The first is the Internet and, more recently, the Abilene backbone (a high-speed research network). The second is the Radar Interface and Data Distribution System (RIDDS; Jain and Rhue 1995), which was created by the NSSL to extract base data from the WSR-88D radar product generator (RPG) for use in decision support systems. RIDDS consists of a Sun

† *Corresponding author address:* Prof. Kelvin K. Droegemeier, CAPS, Sarkeys Energy Center, Room 1110, 100 E. Boyd Street, Norman, OK, 73019-0628; kkd@ou.edu.

SparcStation 5/110 workstation that ingests base data from the radar's wideband-3 port and outputs them as UDP packets to an Ethernet hub. A circular buffer on the RIDDS holds several volume scans of data. At present, 51 RIDDS sites are in place at various NWS Forecast Offices around the country.

Figure 1. WSR-88D systems that are delivering Level II data in real time, via the Internet, as of 1 October 2001.

The third technology is the Local Data Manager (LDM) software developed by the UCAR Unidata program. Used by numerous universities as well as government laboratories and operational centers, LDM (Rew and Davis 1990; Davis and Rew 1994) is a system for event-driven data distribution that acquires data from and shares data with other networked computers.

As shown schematically in Figure 2, CRAFT involves linking to the RIDDS workstation (which is located in the NWS Forecast Office that "hosts" the WSR-88D radar) an inexpensive (approximately $1,000) personal computer running the Linux operating system and LDM. This PC requires only nominal memory (64 megabytes) and disk space (2 gigabytes) and is purchased without a monitor. The data then travel from the PC to an optional multi-port router ($1,500) and onto a standard phone line (e.g., 56K bandwidth, though under some circumstances, such as squall lines, this is insufficient; see below). Recently, at select sites, the LDM computer has been connected directly to the local area network within the NWS Forecast Office.

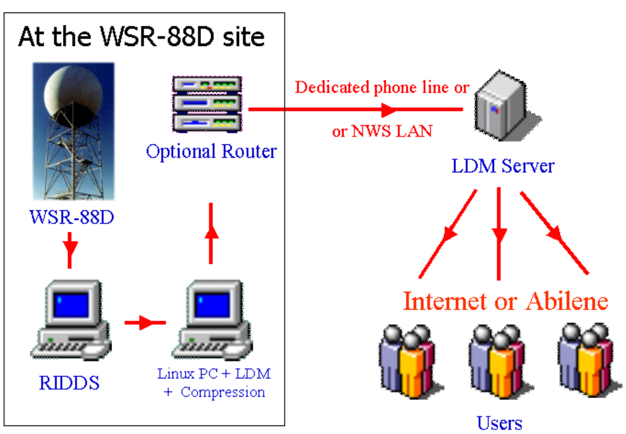

Figure 2. Schematic of the data compression and transmission strategy presently used in Project CRAFT.

A key aspect of CRAFT is the addition to LDM of an off-the-shelf, public domain, no-loss data compression algorithm known as BZIP2. It compresses the base data, in packets of 100 radials, down to an average of 1/10th their original size. (The maximum compression achieved is about 16:1 for clear air data, and the minimum to date in heavy weather has been about 6:1. This translates to a data volume of between 40 and 90 gigabytes/day for the 120 NWS radars. The aggregate compressed base data rate for the entire NWS WSR-88D network is only 30-40 megabits/sec, and consequently bandwidth is not the principal issue with regard to acquisition or distribution.) At the end user site, another computer running LDM ingests and decompresses the base data. In the event of a communications failure, the LDM PC at the radar site automatically stores and can retransmit several days worth of base data.

3. SYSTEM PERFORMANCE AND RELIABILITY

In early June, 2000, the NCDC began receiving compressed base data in real time from the original six CRAFT radars via the commodity Internet at T1 bandwidth (1.55 mbits/sec). The system proved exceptionally stable, with most data outages caused by crashes to the RIDDS system or the radar itself. Although only one phone line was used per radar for this early experimentation phase, thus providing a single point of failure, the communication links have been remarkably stable. In August, 2000, the NCDC began receiving and transferring base data directly, and automatically, to its hierarchical data storage system (HDSS) -- completely bypassing the 8-mm recording system. Transmission stability (Level II data receipts at the NCDC are well above 95% of the total data available) has led to the suspension of 8 mm recording at 18 sites.

Using the Network Time Protocol (NTP), CAPS currently is generating data latency (time delay between transmission and receipt) and other statistics in an effort to evaluate the performance of the CRAFT infrastructure. At the present time, latencies range from a few seconds to well over a minute and are a function of radar mode and weather detected. In some cases, data transmitted from radars close to the LDM aggregation site take longer to arrive than data transmitted from more distant sites. Packet traces and other analyses are being performed to understand this behavior, and network modeling is expected to provide insight into network response under various conditions of load and configuration.

Today, the NCDC and CAPS are receiving Level II data, at T3 bandwidth (45 mbits/sec) via Abilene (see below), from all 21 radars shown in Figure 1; another 20 radars (Figure 3) should be added in the next several months. Efforts now are underway to test data delivery to multiple user sites, including private companies, universities, and government laboratories. The only requirements are an LDM system and sufficiently large disk space and Internet bandwidth.

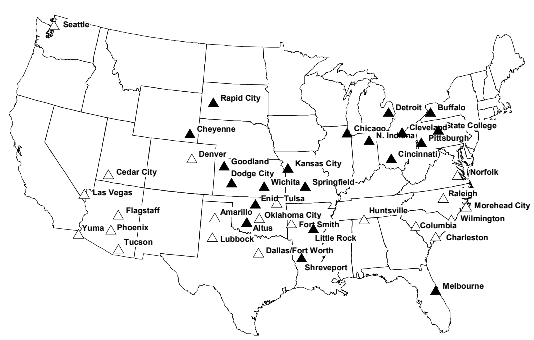

Figure 3. Potential additions to Project CRAFT (black triangles), with current radars shown by white triangles (compare Figure 1).

The NCDC is undertaking another related initiative to upgrade its Level II service. Specifically, all Level II data presently stored on some 76,000 8-mm tapes are being moved to the HDSS (data from 1999 through 2001 already have been moved, and the complete Level II archive should be in place on the HDSS by late 2003). A new web-based system linked to the HDSS has been developed to allow remote users to query the Level II database and download compressed, archived files automatically via the Internet. It is expected to be available soon to the entire user community.

4. VISUALIZATION AND ANALYSIS TOOLS

In anticipation of an Internet-based system for delivering Level II data to large numbers of users, the NSSL and NCDC are developing or adapting web-based tools for perusing data at the NCDC prior to download. Further, they are adapting more sophisticated packages for quantitative analysis of Level II data that can be run on a user's individual computer.

5. TOWARD AN OPERATIONAL SYSTEM

The NWS presently is updating its requirements for real time Level II data, and thus the specific infrastructure by which these data will be delivered remains to be determined. The goal of CRAFT is to explore a variety of options, and below we present one of several under consideration.

Expansion of the CRAFT concept to the entire WSR-88D network will require not only additional bandwidth, but more importantly a variety of security and quality of service capabilities for meeting mission-critical needs, such as data assimilation for numerical weather prediction models. In addition, it must accommodate the move to Open Systems (e.g., Saffle et al., 2001), which involves significant changes to existing radar hardware and software (section 8).

To meet these challenges in a way that engages the entire user community, including the private sector, the authors of this preprint organized a National Level II Data Stakeholder's Workshop in Boulder, Colorado on 14-16 February 2001. Approximately 70 attendees represented a broad cross section of the user community, and via several presentations and break-out discussion sessions, a two-phase plan was established.

In Phase I, shown in Figure 4 and designed to evaluate the stability, security, and redundancy of an operational framework, a single communications link to each participating radar (either a phone line or the NWS LAN as in Figure 2) would provide compressed data to an LDM server. The latter, known as LDM "super sites," already exist in scaled-down form at many universities and even NOAA laboratories, and could easily be upgraded to become a super site. Located on Abilene but connected to the commodity Internet as well (Figure 4), the super sites would ingest data from a dozen or so nearby radars (only one set of radars is shown for clarity) and then relay these data to all other LDM super sites (shown in Figure 4 by the ovals). In this manner, each super site would contain data from all radars, with the Abilene network serving as a high-bandwidth "bus" to provide tremendous redundancy and options for essentially unlimited growth.

Any user (indicated schematically in the upper right corner of Figure 4) could obtain data via the commodity Internet, or Abilene, depending upon their connectivity and available bandwidth. NOAA and similar organizations would have the *option* to receive, though not re-transmit data, at their LDM super sites *if so desired*. This would reduce the load on NOAA servers and thus help provide a higher quality of service for time-critical operations. As networking capabilities continue to expand, bandwidth limitations on the commodity Internet are likely to vanish entirely, along with the current distinction between it and Abilene.

Assuming success in Phase I, Phase II would involve the installation of dual (redundant) communication links from each radar to distinct points on the Internet. This would remove the single point of failure present in Phase I, with the links to be of sufficiently high bandwidth to accommodate future growth (e.g., to dual-polarization). Furthermore, the differing needs of the academic/research, private sector, and operational communities may suggest that the network be separated into three independently-managed logical components or buses, one for each of the above groups. Though operated separately, the networks would provide redundancy for one another.

Although the NCDC will continue to serve as the permanent archive for Level II data, we envision the establishment of one or two national "data stores" that would contain a revolving catalog of several months of the most recent Level II data from all radars. These data, stored on solid state or rotating disk (i.e., not on slow tape) could be accessed virtually instantly for all types of research. Further, these servers would provide short-term redundancy in the event of a catastrophic system failure at the NCDC.

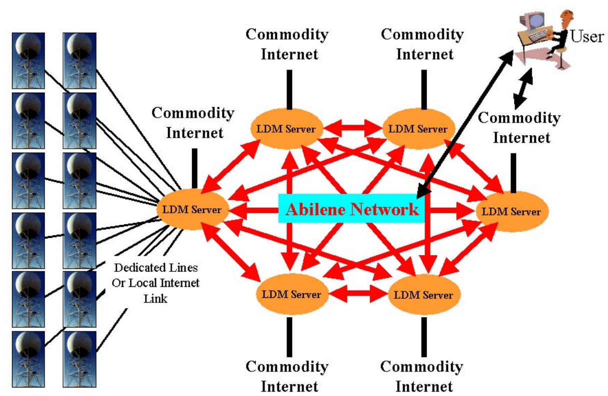

Figure 4. Schematic of Phase I experimental infrastructure designed to evaluate the use of Abilene as a bus for delivering Level II data in real time. See the text for further details.

6. THE VALUE OF AN INTELLIGENT NETWORK

Because of the time-critical nature of weather prediction and warning, the network infrastructure used to deliver Level II data must be able to respond rapidly to changing load. Nowhere is this requirement better illustrated than in storm-scale numerical weather prediction (e.g., Droegemeier, 1997), where Level II data will be assimilated continuously in prediction systems operating at grid spacings of 1-3 km. By using a series of nested grids, each having successively finer mesh spacing, the model will be capable of capturing the details of local, high-impact weather while at the same time representing the larger-scale environment. With automated nesting, in which new grids are initiated over regions of developing weather and removed over regions where active weather is dissipating, the computation becomes even more dynamic because the model configuration changes with time. Further, as storms grow and decay across the nation, the networking requirements for Level II data transmission change markedly.

By the time storm-scale prediction becomes operational, it will be practicable to run part of the calculation (e.g., the outer grids) at one computing facility and another part (e.g., the inner grids) at a facility perhaps hundreds or thousands of miles away in a manner that maximizes throughput and load balances the entire computation (a concept known as the "virtual machine room"). Because an operational forecast model must be *guaranteed* real time access to base data from all WSR-88D radars, the network must respond by providing increased bandwidth in areas of active weather (i.e., in locations of locally high data volumes). This concept of dynamic networking linked to dynamic computation and data acquisition in numerical weather prediction is the essence of the national computational grid (Foster and Kesselman, 1999), and is the reason CRAFT is emphasizing the use of Abilene and Internet2 -- which together represent research and development on high-bandwidth intelligent networks (a capability not available in the commodity Internet).

7. BENEFITS AND LIMITATIONS

A principal benefit of CRAFT, particularly in terms of cost and reliability, is its utilization of existing software (e.g., LDM, BZIP2, Linux) and networks (Abilene and the commodity Internet). As described above, this framework is laying the foundation for the cost-effective expansion of CRAFT to all WSR-88Ds. It provides for the inclusion of other radars (e.g., from the FAA); can accommodate the transmission of base data in real time to virtually any user in the world; and allows for linkages with other networks (e.g., AWIPS). Further, it allows expansion to larger data sets via enhancements to the existing radars (e.g., dual polarization, denser scanning strategies) or the introduction of new ones (e.g., phased-array).

Although the Internet itself might be viewed as unreliable for use in time-sensitive meteorological operations, especially because stakeholders lack ownership of or control over every link involved, our experience suggests just the opposite. Indeed, despite the possibility of major Internet disruptions (e.g., inadvertent excavation of primary trunk lines, terrorist attacks), the Internet is so pervasive and so redundant that it probably represents one of the most reliable means of data transmission currently available.

One principal limitation of Project CRAFT, as currently implemented, is the single point of failure in the communications link to each radar (which, as noted above, will be addressed in Phase II). Further, although the 56K dedicated phone lines used at most sites are sufficient to handle the network load based on known file sizes, it was recognized recently that the effectiveness of BZIP2 is predicated not upon file size, but rather upon the structure of the data to be compressed. It is now clear that both large tropical systems -- with imbedded mesoscale structures -- as well as systems containing large gradients, e.g., squall lines, render the compression algorithm less efficient. Thus, even though a squall line fills fewer range gates with non-zero data values compared to, say, a hurricane, and thus produces a *smaller uncompressed* file, the relatively large reflectivity *gradients* in the line reduce the effectiveness of run-length encoding in BZIP2, thus producing a *larger compressed* file.

Although the frequency with which compressed Level II data exceed the 56K bandwidth limit is not yet known, we believe it to be less than 5%. It was recognized early in the project that large data sets would on occasion approach the 56K limit. However, anticipated improvements in compression were thought to provide a sufficient margin of safety. With BZIP2 nearly optimal for radar data, the NSSL has developed *pre-processing* techniques that may improve the efficiency BZIP2 to the point where 56K lines once again become sufficient in all cases for the present WSR-88D configuration. Unfortunately, any enhancements to the radar (e.g., more dense volume scan strategies, introduction of multi-moment data), coupled with the occasional need to re-transmit Level II data that were saved on the local LDM during communication system outages, ultimately may render 56K lines unsuitable for operational use.

8. THE MOVE TO OPEN SYSTEMS

Because Project CRAFT is a proof-of-concept activity, it makes use of technology, like RIDDS, that fulfills present needs but does not have a long-term future in the WSR-88D program. If the CRAFT concept is to be implemented in a stable manner for future operations, it must be fully integrated into the WSR-88D framework. The opportunity to do so now exists by virtue of a long-standing effort within the NEXRAD Product Improvement Program to improve overall system functionality, reduce cost, and provide greater flexibility for the future. Known as the Open Systems Architecture (Saffle and Johnson 1999), this project involves replacing existing hardware and software with off-the-shelf "open" components that make use of industry standards such as Unix.

In the new Open Systems Architecture, the deployment of which is underway and will conclude in late summer, 2002, the functionality of the RIDDS has been replaced by the Base Data Distribution System (BDDS). In summer, 2001, the Oklahoma City WSR-88D received the first Open Systems Radar Product Generator (ORPG)/BDDS. The LDM and compression software, appropriately modified to accommodate the new TCP/IP protocol, worked flawlessly.

Although the functionality of the RIDDS is contained within the BDDS, the LDM data distribution system, and its associated compression algorithm, are not. As BDDS systems are deployed at existing RIDDS/LDM sites, the CRAFT functionality will be maintained. However, at all other sites, additional LDM systems will have to be installed until such time that the LDM and compression software can be integrated into the BDDS itself. The decision to proceed in this manner is predicated upon the outcome of current experiments, and upon our ongoing critical evaluation of networking options for the delivery of Level II data, including via networks operated by the NWS.

9. RESEARCH PROSPECTS

The opportunities for using Level II data in research are enormous, and we refer the reader to the following web sites for some creative examples: http://mesocyclone.ou.edu/ and http://www.nssl.noaa.gov/swat/Cases/cases_pix.html.

10. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Russ Rew, Woody Roberts, Guy Almes, Jim Belville, Jack Hayes, Walt Telesetsky, Bob Saffle, Warren Blanchard, Esther Atkins, and Greg Mandt. The advocacy of improved access to WSR-88D data by the US Weather Research Program Science Steering Committee and, in particular, by Rit Carbone and Bob Gall, is gratefully acknowledged.

This research was supported by the National Science Foundation under Grant ATM91-20009 to CAPS, and by Grant ATM99-81130 to the first author. Additional support was provided to CAPS and the NSSL by NOAA through Cooperative Agreement NA67RJ0150, which involves funding from the NOAA ESDIM (Environmental Service Data and Information Management) and HPCC (High Performance Computing and Communications) Programs.

11. REFERENCES

Online at http://kkd.ou.edu/iips2002.html.