6.4 THE UNIDATA INTERNET DATA DISTRIBUTION (IDD) SYSTEM: A DECADE OF DEVELOPMENT

Tom Yoksas*, S. Emmerson, S. Chiswell, M. Schmidt, and J. Stokes
Unidata Program Center/UCAR, Boulder, Colorado, USA

1. INTRODUCTION

The mission of the Unidata program of the University Corporation for Atmospheric Research (UCAR) is to provide universities with innovative applications of current computing and networking technologies to access and use atmospheric and related data for education and research.

Unidata's long-term communications goals have not fundamentally changed since their original formulation (Cooper, 1985, Domenico, 1989). These include employment of local area networking, access to supercomputer centers and national data banks, and the need for the community to exchange mail, software, and data among themselves. The vision for the last goal, the exchange of data among sites originally focused on use of two way satellite communications not unlike the satellite-based, commercial Internet access that is available today. Because of the high cost of two-way satellite communications until recently, this vision was put on hold, but never abandoned.

The development of NSFnet and its successors provided the substrate on top of which a multi-way communications system could be built. The Unidata Local Data Manager (LDM) evolved to be the vehicle that enabled the multi-way sharing of data in the Unidata community through a project known as the Internet Data Distribution (IDD) system. The IDD is an event-driven network of cooperating Unidata LDM servers that distributes discipline-neutral data products in near real-time over wide-area networks.

The IDD was developed in the early 1990s in response to challenges related to weather-data ingest via satellite broadcast (e.g., local sources of terrestrial interference, data outages caused by solar occultation, weather-related outages due to signal degradation, and the difficulty in locating satellite reception systems near departmental computing resources) and to provide access to datasets that were not commonly available. Starting with a modest goal of internet delivery of data available in the NWS Family of Services satellite broadcast, the IDD has grown to become the leading Internet2 advanced-application and one of the top bandwidth users (http://netflow.internet2.edu/weekly/), currently delivering about 20 terabytes (TB) of data per week in the aggregate to participating institutions. Stress testing conducted at the Unidata Program Center offices in the summer of 2005 demonstrated that a cluster approach to LDM data relay was limited only by the bandwidth available in the underlying (gigabit) network thus ensuring future IDD expandability at lest over the next few years.

The Unidata IDD has expanded from a US-centric delivery system to one that includes 13 countries on 5 continents. Additionally, the LDM is being used as the data distribution engine in systems akin to the Unidata IDD: by private industry; by several US government agencies including the National Weather Service and NASA; and by the national weather services of South Korea and Spain.

Corresponding author address: Tom Yoksas, Unidata/UCAR, PO Box 3000, Boulder, CO 80307; e-mail <yoksas@unidata.ucar.edu>

2. HISTORY OF THE LDM

LDM-1: 1987 - 1989

The goals for the original LDM prototype (Campbell and Rew, 1988) were modest by today's standards:

> "One of Unidata's primary goals is to enable the acquisition of meteorological data on a single computer and to allow access to these data by possibly dissimilar workstations within the same facility."

> "The primary goal of developing an LDM prototype was to refine our understanding of the issues and problems involved in implementing a production-quality system that would meet the expectations of the Unidata community."

The design principles adopted for LDM development, however, continue to this day:

- Extensibility: to provide a system architecture designed to be readily extended by users for the capture of new kinds of data from new data sources

- Generality: to allow for the simultaneous capture of multiple data streams with different structures
- Capacity to handle high-speed feeds: to permit the capture of data at speeds that are higher than used currently for conventional weather data
- Portability: to isolate system dependencies so that the resulting system will run under a variety of operating systems on a variety of workstations
- Performance: to capture data reliably from several sources without consuming a significant fraction of the resources on the host workstation
- Network functionality: to permit the access of data from other workstations on the network
- Robustness: to permit the unattended capture of data for long periods, in the face of data errors and limited disk space

LDM-1 included four basic modules: product processing manager, ingester, digester, and an administrative process. The LDM-1 prototype ran on DEC MicroVAX II/VMS and Sun 3/110 Unix workstations (Campbell and Rew, 1988; Green, 1988; and Fulker, 1988).

LDM-1 design decisions included:

- Used C as the implementation language
- Use of a coprocessor board for synchronous data
- Used one ingester for each feed
- Use of FIFO files for buffering
- Permit digester to be fed by one ingester
- Centralized control in a Product Processing Manager (PPM)
- Use of a mailbox abstraction for inter-process communication
- Use of a common statistics and error-reporting abstraction
- Use of Remote Procedure Call (RPC) mechanisms for network access

LDM-1 was developed in the era of NWS satellite broadcast of the Family of Services where data rates ranged from 2400 baud for the asynchronous DD+ textual data stream to 9600 baud for the synchronous NPS model data stream.

LDM-2: 1989 – 1991

LDM-2 was developed using lessons learned from the LDM-1 prototype. The objective was still the capture of data from the satellite-broadcast, NWS FOS IDS, DDS, PPS, and NPS data streams and the satellite-broadcast Unidata-Wisconsin data stream provided under contract by the Space Science and Engineering Center (SSEC) of the University of Wisconsin-Madison. The Unidata-Wisconsin data stream contained current weather data and GOES satellite imagery in a format directly usable by the personal computer implementation of the Man-computer Interactive Data Access (PC-McIDAS) application.

LDM-2 introduced the abstract product data type as the basic unit for processing. It also introduced support for table-based, pattern-action construct where regular expression patterns were matched against data product headers to determine what, if any, actions the user wanted to occur. The actions supported in LDM-2 included:

- FILE: write the product into a file on disk
- EXEC: start another program and pass the product to it as input
- GRIB: decode NPS numerical products
- SAO: decode surface airways observations
- UPA: decode atmospheric soundings

LDM-2 was, in many ways, a monolithic system. The ingest interface was defined as an ordinary or remote procedure call at compilation. The action subsystem was composed of subroutines for each supported action.

LDM-2 was designed to be extended as painlessly as possible by Unidata users that wanted to merge their software with the LDM. This effort did, however, require those users to include their code as subroutines run by the LDM itself.

The design of LDM-2 allowed processing of “a dizzying volume of information – on the order of 100 Mb/day, aggregate” (Davis and Rew, 1990).

The LDM-2 provided the lowest layer of the Unidata Scientific Data Management (SDM) system, a system for UNIX- and VMS-based workstations composed of two distinct systems: one for data management, and the other for data analysis and display (Fulker, 1990).

LDM-3: 1991 - 1993

LDM-3 was designed to extend and refine the concepts contained in LDM-2. In particular, LDM-3 changes included (Davis, 1991):

- The string used as a product identifier was regularized. This allowed for simpler regular expression patterns in the server initialization file.
- Elimination of the “stuck in queue” problem. LDM-2 users complained that the last product in the *ingester-server-outputfile* processing pipeline would not be available when the feed was idle for long periods of time. This situation was particularly bad for the NPS data stream since the last product received from a model run would not be made available until data from the next model run was received.
- Servers were enabled to pass on data to clients as the data were received.
- Data ingesters could feed multiple servers.
- Data ingesters would configure themselves based on the name they were invoked by. This allowed the same code to be easily used for different feed types. All that was needed was the creation of a symbolic link to the ingester of the appropriate name.
- A new RPC call, *sendme*, was added to enable data sharing among LDM servers. The same call could be used by decoders to get data directly from the LDM.

The vision for use of “high speed” networks for LDM distribution of research datasets was first articulated (Domenico, 1992):

> “We are also beginning to experiment with a system that potentially could use the Internet to distribute certain research datasets (Davis, 1992). The idea is to have cooperating Unidata LDM systems ‘fan out’ the data from sites where it is injected to other sites on the Internet.” “Each of these hub sites would in turn be capable of passing the data along to others, processing it locally, and making the processed data available to other nodes on the local network.”

The difficulty of accomplishing this goal (one must remember that this was in the infancy of the Internet) was well recognized:

> “Distributing anything over the Internet means communicating in a complicated computing environment. Because computers of different architecture have different communication requirements, sending information from a computer to another type can be technically challenging. One of Unidata’s goals is to provide access to the power and capabilities of all systems on the network while making the network node appear to be simple extensions of the workstation on the scientist’s desk. Our hope for achieving this goal lies in close compliance with computing standards.”

Furthermore:

> “Networking at the Unidata Program Center is a concrete example of Thelonius Monk’s observation: ‘Simple ain’t easy.’ The goal of having at your fingertips all the power of all the computers on your network is becoming feasible, but it does require piecing together many components that were not designed to fit together. In fact many of the components are still not available.”

Thus, the conceptual notion of the IDD was born.

LDM-4: 1993 - 1996

LDM-4 was the first LDM implementation designed to support movement of data between servers connected by the Internet. The application was recast into a model where installations could function both as client and servers.

The goals articulated in the original LDM-1 prototype continued and were extended in LDM-4 (Davis and Rew, 1994):

- Enhance portability by using a layered, standards-based approach: This requires the use of the most generally applicable (abstract) interface from the following choices: ANSI C, POSIX.1, ONR RPC 4.0, and BSD sockets.
- Support functional backward compatibility with LDM version 3. This means that anything a site could do with version 3 can be done with version 4. Additionally, version 3 clients could interoperate with version 4 servers.
- Include protocols for handling large products.
- Include (new) protocols and facilities for event-driven data dissemination and notification, to be used for Internet data dissemination and external decoders.
- Support distributed error handling. A new interface library allowed more manageable logging in a distributed system (via the syslogd(8)) interface).

The abstract product data type introduced in LDM-2 was generalized to include an identifier (a simple strings limited to 255 bytes) and a body that consisted of a counted array of bytes. Since the LDM does not look at or modify the contents of the body of the product, it is discipline-neutral. The products identifiers were also very general in that they were not required to conform to any standard.

For the purpose of coarse discrimination between products of similar types, the concept of a feed type, an enumerated type designed to inform the LDM about how a product identifier should be interpreted, was adopted. Thus, every product was composed of an identifier, body, and associated feed type. The product identifier and feed type control its routing through the distributed system.

LDM-4 roles were most easily understood in terms of the flow of data. A data source was a process that provided data to the system. A data sink was a process that received data from the system. LDM-4 processes could receive products and redistribute them to others in the distributed system. In this way, they could be both data sinks and sources. The notion that data flows *downstream* from data sources to data sinks was first expressed.

LDM-4 provided protocols that allowed processes to request transfers of products as a source of data to a *downstream* or to request the transfer of products from an *upstream* source as a data sink. In the former, the process to which products would be sent acted as a server; in the latter, the process from which product transfers are requested acted as a server. In both cases the set of products requested for transfer was defined by feed type and a regular expression pattern that would be matched against product identifiers. The object of the transfer request was defined by a host name. A simple access control mechanism based on the triplet of host name, feed type and product identifier pattern was instituted on the object of the request. The set of products that would be transferred from a source to a sink was defined to be the intersection of the set allowed on the object of the request and the set on the source of the request. If the intersection of the set allowed and the set requested was empty, no data would be sent. The set of products allowed in lieu of an explicit configuration entry was empty so that unexpected requests would be rejected.

LDM-4 provided two mechanisms for transfer of data products. *Reasonably* sized products were sent in a single transaction. Delivery relied on the reliability of the underlying TCP transport. *Large* products, on the other hand, were broken into a series of *reasonably* sized blocks, and each block was sent sequentially to the downstream process. In the large product case, the upstream process needed to wait for confirmation of receipt from the downstream process before sending the next block. A *reasonably* sized product was defined in code to be 16 KB.

In implementations previous to version 4, the LDM server contained actions for specific types of data instantiated as subroutines. LDM-4 moved data processing to external programs that read from standard input and were connected to the LDM through a PIPE action that provided products on standard output. Moving the data processing tasks to processes external to the server increased system modularity, maintainability, and extensibility. LDM users could easily add custom processing without having to understand the design of the server.

LDM-4, while providing many innovations in the use of the Internet, contained some inherent limitations. If an upstream server was feeding multiple downstream data sinks and one of the data sinks had a congested network connection or was slow in handling the data it received, the other data sinks being fed data could eventually suffer data loss, and the situation could eventually backup to servers further upstream. This so called “*slow link problem*” was mitigated somewhat by application of appropriate timeouts and use of non-blocking I/O with large buffers, but there were still situations where the reliable delivery of data could fail.

Additionally, the LDM-4 could deal with limited outages of servers, hosts, and networks, but recovery of data lost during the outage had to be handled manually.

Although LDM-4 was designed to be a practical solution for a reasonable range of kinds of data products, sizes, and configurations, limitations were recognized to exist for the number of distinct feed types, number of pattern-action lines in a configuration file, numbers of connections possible to upstream and downstream sources and sinks, and the ability to handle very large products.

LDM-5: 1996 - 2003

LDM-5 was developed to address the known limitations in LDM-4. LDM-5 included a number of significant changes:

- The product queue was recast from a FIFO to a shared queue (*squeue*)

- MD5 product signatures were added for use in duplicate product detection and elimination.
- Role reversal in FEEDME requests: In LDM-4, FEEDME requests moved over a different channel than the data. In LDM-5, the server and client did a role reversal and then use the same communication channel for the data transfer.
- Consolidation of multiple feed type/pattern requests to the same upstream LDM. This was added to limit the number of processes on both the downstream and upstream systems.
- The LDM was modified to support architectures that support 64-bit file offsets so that the previous 2 GB limit for LDM product queues was eliminated.

Even in the earliest releases of LDM-5, it was recognized that the amount of time required inserting or deleting a product from the product queue varied as a function of the number of products in the queue. When the number of products in the queue was modest (less than or equal to about 10,000), this time was negligible. When the number of products in the queue grew to 50,000 or more, the time became appreciable. This problem came to the forefront when the UPC attempted to ingest all NEXRAD Level III products and some experimental streams not generally available to the community (Rew and Wilson, 2001). The increased insertion time resulted in data relay delays to downstream sinks and even local data loss during periods when the LDM was unable to insert products into the queue at the rate they were being received.

The LDM product queue was redesigned using a relatively new computer science development, the *skip-list* (Pugh, 1990). The result of adding *skip-list* technology to the LDM was a **dramatic** decrease in the time needed to add, delete, and find products in the queue. Additional benefits included elimination for the need to run the queue expiry program, *pqexpire*, to free space in the queue; space in the product queue could be created as needed as new products arrived. Also, the arbitrary limit to the amount of time that data can be stored in the queue was eliminated. The amount of time that data could remain in the queue became a function of the size of the queue, not the number of products in the queue. This provided needed, additional elasticity for the IDD and meant that a significant amount of data could remain available for processing even when connectivity to upstream data hosts was lost.

With product queues larger than 2 GB, a data archive could conceivably be represented and accessed as an LDM product queue, providing a convenient form of retrospective data access for other LDMs (Rew, 2000).

LDM-6: 2003 - Present

LDM-6 was developed in large part as response to LDM-5 limitations in relaying data to electronically distant nodes.

In addition to a significant overhaul of the code base, the major differences in LDM-5 and LDM-6 are:

- Selection of protocol to send products: In LDM-5 if a product was less than 16 KB, the HEREIS message was used; if greater, the COMINGSOON message was used. In LDM-6, the size the product must be to switch to the COMINGSOON message was made user-configurable.
- Waiting for a downstream reply before sending product. LDM-5 waited for the downstream to reply in the affirmative that a product/chunk had been received before sending the next product/chunk. LDM-6 uses batched RPC calls whenever possible so there the upstream server does not have to wait for the downstream reply.
- Handling of large products: LDM-5 would send as many 16 KB pieces of large products as needed in BLKDATA messages. LDM-6 sends the entire product in a single BLKDATA message.
- Maximum amount of time between RPC messages: In LDM-5 this was 5 minutes. In LDM-6 this was decreased to 30 seconds.
- Multiple downstream requests to the same server: In LDM-5 the requests were consolidated into a single request. In LDM-6 the requests are not consolidated. This allows greater throughput with current implementations of the TCP protocol.
- Action upon receipt of a HEREIS message: In LDM-5, if a new or duplicate product is received, then reply with OK message; otherwise, reply with RECLASS message. In LDM-6 there is no reply.
- Reconnection strategy: In LDM-5 if nothing was received in 12 minutes then reconnect. In LDM-6 if nothing is received in 1 minute then connect to top-level upstream LDM server and send an IS_ALIVE message.

Reconnect if and only if reply indicates sending LDM has terminated.
- Statistics: LDM-5 statistics gathered by *pqbinstats* were mailed back to Unidata hourly for analysis. The LDM-6 *rtstats* facility records statistics every second and sends them to an LDM server specified by the user every minute.

The removal of waiting for responses from downstream nodes coupled with the ability to send large products in a single transaction significantly decreases the time it takes to send products to downstream nodes. The effect was most notable when the nodes were significantly distant from one another. One of the most dramatic examples of this was data transfers from the Unidata Program Center offices in Boulder, CO to the Universidade Federal do Pará in Belém, Brazil where product latencies decreased from hundreds (if not thousands) of seconds down to a few seconds or less.

Some of the highlights of LDM-6 releases are included in the following:

LDM-6.1:
- a write counter was added to the LDM queue. This allows for a fast determination of whether the product-queue was properly closed and, consequently, self-consistent.

LDM-6.2:
- RPC sub-package was added to the distribution replacing use of the native RPC library. This was mainly done to work around a bug in the AIX 5.1 ONC RPC implementation.
- Corrected a bug in product queue module that prevented insertion of products that had the same insertion time as an existing data product.
- All programs that use regular expressions were modified to convert "pathological" (i.e., over constrained) regular expressions to non-pathological equivalents. Pathological regular expressions can use several orders of magnitude more CPU.
- The configuration section of the *ldmadmin* utility was moved into a separate file, ldmadmin-pl.conf. This eliminates the need for sites to modify *ldmadmin* in each new LDM installation to include site-specific configurations.
- The LDM was ported to MacOS-X.

LDM-6.3:
- Added ability of user to specify what network interface (IP address) the server should use. This allows the creation of directory/server clusters and the ability to run more than one LDM (as different users) on a single platform.
- Added the ability to set which logging facility to use in the ldmadmin-pl.conf configuration file.

LDM-6.4:
- Added the ability to use a port other than the default, 388.
- Added the ability to encode MD5 signature of last, successfully-received data-product in FEEDME product-class specification when connecting to upstream LDM-6. This prevents skipping of data-products that arrive at upstream LDM-6s out of order.
- Added the ability of the downstream LDM to automatically adjust the feed transfer-mode (primary vs. secondary) based on success of inserting data-products into the product-queue.
- Reduced CPU utilization by approximately 75%.
- Added upstream filtering. An upstream LDM can now filter data-products based the product-identifier and regular-expression in the LDM configuration-file.

The automatic switching of the feed transfer-mode by a downstream LDM was added to increase the reliability of data reception while, simultaneously, reducing its bandwidth use. This allows a downstream site to request the same data from multiple upstream sites (to improve reliability) without worrying about bandwidth usage.

Upstream filtering allows an upstream site to more closely control the products that downstream sites may receive. For example, an upstream site can now restrict downstream sites from receiving certain portions of data-streams.

3. HISTORY OF INTERNET DATA DISTRIBUTION

The driving force behind the creation of the IDD was the desire to develop a system for disseminating real-time, scientific data which would build on Internet facilities as the underlying mechanism for data distribution, and for broadening the community of users who can use the information (Domenico, Bates, and Fulker, 1994). Additionally, it was recognized that an IDD could provide data not commonly available to end users, and users would be able to share locally-held datasets with fellow IDD participants.

The initial IDD system was designed to:

- Enable scientists and educators to use their local workstations and personal computers to access scientific data from a wide variety of observing systems and computer models in near real-time.
- Allow data to be injected into the system from multiple sources at different locations
- Enable universities to capture these data, process them, and pass them on in easy-to-understand and easy-to-access forms (such as electronic weather maps in raster image files) to other institutions having more modest data needs as well as more modest equipment resources and technical expertise.

Deployment of the IDD was spurred by three factors:

- The switch to KU-band satellite technology by the commercial entity through which the great majority of users received satellite-broadcast data. The costs being faced for conversion from C-band satellite reception to KU-band by over 60 sites made deployment of LDM-4 a priority.
- Several sites who wanted to participate in the Unidata-subsidized satellite broadcast of data had been unable to mainly for two reasons that were out of their control:
    - local terrestrial interference from sources who had airway "right-of-way" (e.g., the military and the phone company)
    - an inability to locate satellite receiving equipment near departmental computing resources (e.g., campus beautification committees)
- The eagerness of the NOAA Forecast Systems Laboratory (FSL) to take advantage of the event driven data distribution system that LDM-4 offered.

With the release of LDM-4.1 in November, 1994 came the push to transition users from satellite-broadcast NWS FOS data to IDD deliver of the same data. The Unidata Program Center's assistance to sites in installing and configuring the new technology helped speed the acceptance and use of the IDD in the community. By early-1995 essentially all sites that were willing and able to transition were receiving data via the IDD.

The community of US IDD participants grew throughout the years as the LDM continued to evolve to be better able to relay ever increasing volumes of data. By the time that LDM-5 was deployed, over 100 institutions in the US and Canada were participating in and benefiting from the real-time data flowing in the IDD.

CRAFT

In 1998, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma joined forces with Unidata, the University of Washington, the National Severe Storms Laboratory, and the WSR-88D Operational Support Facility to establish the Collaborative Radar Acquisition Field Test (CRAFT). The principal goal of CRAFT was to demonstrate the real-time compression and internet-based transmission of WSR-88D base data from multiple NEXRAD radars. The internet-based transmission technology employed was the Unidata LDM-5, and the resulting system could be regarded as a closed IDD (Droegemeier, 2001). Through stakeholder cooperation, a combination of leveraging technology and creative partners forming useful collaborations, the technology was transferred to the National Weather Service in 2004, and the data are now available to the broad community of users (Miller, 2006).

CONDUIT

Also in 1998, with support by the U.S. Weather Research Program (USWRP), the combined efforts of the UPC, the National Weather Service's Office of Meteorology and Office of Systems Operations (OSO), the National Centers for Environmental Prediction (NCEP), and the NASA Goddard Space Flight Center (GSFC), resulted in the creation of the Cooperative Opportunity for NCEP Data Using IDD Technologies (CONDUIT) project (Chiswell and Miller, 1999; Miller, 2006). The goal of CONDUIT was use of LDM/IDD technology to provide access to and distribution of high resolution model output that was only available from NCEP and NWS/OSO FTP servers. Making the CONDUIT model data available to the broader community enabled researchers to obtain the model datasets as soon as they were available on the NCEP and NWS/OSO servers using the LDM's "push" technology. CONDUIT data volumes were and continue to dominate the total flow in the Unidata IDD.

SUOMINET

At about the same time, the UPC was engaged with its community and the NOAA Forecast Systems Laboratory in the creation of a large network of GPS receivers (SuomiNet) and to use the resulting data for estimating water vapor in the atmosphere and total electron content in the ionosphere (Ware, et al,, 2001). LDM-5 was employed to collect and disseminate project data in an IDD modeled after the existing Unidata IDD.

NON-US IDDs

The LDM was also adopted for use in constructing data distribution systems in foreign countries. In 2000, the Instituto Nacional de Meteorologia of Spain used the LDM-5 to build an IDD to distribute model output and METEOSAT image data to weather offices throughout Spain. In 2003, the National Weather Service of South Korea used LDM-6 to build an IDD to distribute internally-produced model data to their forecast offices.

OTHER IDDs

The LDM was also adopted by other groups for internal data distribution networks. The Johnson Space Center used LDM-5 to distribute data used in their operations (Batson, 2002). The Weather Underground, Inc. also employed the LDM-5 to distribute data within their site since it was view as being faster or more efficient than using NFS or FTP.

SOUTH AMERICAN IDD EXTENSION

The international expansion of the Unidata IDD began in earnest as the first phase of the ***MeteoForum*** pilot project (Yoksas, et al, 2004) conducted by the Unidata and COMET programs of UCAR. The first phase of ***MeteoForum*** was the provision of real-time flows of hydro-meteorological data delivery to WMO Regional Meteorological Training Centers (RMTCs) in WMO Regions III (South America) and IV (North America). The UPC's role in this effort was particularly well suited to its primary mission since the RMTCs involved in the pilot project are co-located, or closely aligned with prominent national universities:

Argentina	Universidad de Buenos Aires (UBA)
Barbados	University of the West Indies (UWI)
Brazil	Universidade Federal do Pará (UFPA)
Costa Rica	Universidad de Costa Rica (UCR)
Venezuela	Universidad Central de Venezuela (UCV)

Brazilian participation in the IDD was inaugurated in fall, 2001 simultaneously at the Universidade Federal do Rio de Janeiro (UFRJ) and the Universidade Federal do Pará (UFPA). LDM-5 was installed in the Laboratório de Prognósticos em Mesoescala (LPM) at the UFRJ to ingest the real-time, meteorological data available in the IDD.

In summer, 2002 Unidata installed the LDM-5 at the UFPA to test the feasibility of delivering GTS observational data, model output, and GOES-East satellite data in near real-time to the RMTC co-located on the UFPA campus in Belém. Results reinforced previous observations that the data delivery engine behind the IDD, the LDM-5, was inefficient when relaying data between machines that are electronically distant. Counter-intuitively, relaying data to a sequence of intermediate hosts actually improved the end-to-end performance of the IDD.

Since the UFRJ had access to the Internet2 connection in Rio de Janeiro and was already ingesting data as an IDD receive-only node, they were approached with a proposal that they act as a top-level IDD relay node, initially for the RMTCs at the UFPA and the UBA in Buenos Aires and then throughout Brazil (Yoksas, Coelho, 2002). The UFRJ continues to act as a top-level IDD relay node.

Lessons learned in the UFPA data relay tests were combined with independent efforts at the UFRJ, the Hong Kong University of Science and Technology, and at the University of Melbourne (Melbourne, Victoria, Australia) in an LDM redesign that resulted in the creation of a next generation LDM, the LDM-6 (Emmerson, 2003), that is able to relay substantial volumes of data to both local and remote sites with little to no latency (the time difference from when a product is first injected in the IDD and the time the product is received).

The ability to relay virtually all of the data available in the IDD to Brazil was demonstrated in a series of "stress tests" between the UPC offices in Boulder, CO and the UFRJ in Rio de Janeiro. Over a ten day period at the end of December, 2003, all non-proprietary IDD data streams were relayed over Internet2 to the UFRJ IDD node housed in the campus Network Operations Center (NOC). An average of 1.5 GB of data, with peaks exceeding 2.7 GB, was relayed to the UFRJ each hour. During this test, product latencies (the time difference between a product first entering the IDD and when it is received) remained in the sub-second to a few seconds range. This test convinced us that the UFRJ could assume a leading role in real-time data dissemination in Brazil.

IDD-BRASIL

In late 2003, Brazilian data relay capabilities were bolstered when the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC, a division of INPE, http://www.cptec.inpe.br/) joined the UFRJ in providing data relay to sites in Brazil and Argentina. The IDD offered a means by which CPTEC could receive a reliable stream of real-time hydro-meteorological data. Like the UFRJ, CPTEC is also well connected to Internet2 (de Almeida et al, 2005).

The datasets being moved routinely to Brazil include high resolution NCEP model output (the IDD CONDUIT and HRS streams), high resolution GOES-12 satellite imagery (the IDD UNIWISC stream), and GTS global observation data (the IDD IDS|DDPLUS stream). The relay system established at the UFRJ and CPTEC was named the ***IDD-Brasil***, and has evolved into a peer of the Unidata IDD (Yoksas, et al, 2004).

Part of the establishment of the ***IDD-Brasil*** was the drafting of a set of *principles of participation*:

- For the most part, participants must be educational institutions
- Participants must acquire and maintain appropriate computer hardware and Internet access
- Real-time data must be relayed free-of-charge
- Cost of participation is sharing of locally-held datasets with fellow participants
- Top-level relays must take ownership of the expansion and support processes

The first institution to receive ***IDD-Brasil***-relayed data was the UFPA and its associated RMTC. Soon thereafter, data relays were established between the UFRJ and CPTEC so they could act as each other's real-time data ingest backups and share data-relay duties.

Efforts aimed at broadening participation in the ***IDD-Brasil*** have been ongoing since its inception. CPTEC (Waldenio Gambi de Almeida) and the UFRJ (David Garrana Coelho) have been promoting the benefits of participating in the ***IDD-Brasil*** and in the usefulness of Unidata display and analysis systems through discussions with a variety of Brazilian universities. This effort has been very successful: currently, Brazil ranks second only to the US in ***IDD*** participation.

The first university outside of Brazil to connect to the ***IDD-Brasil*** was the Universidad de Aveiro in Portugal. The second international site was the Universidad de Buenos Aires, in Buenos Aires, Argentina. The third international site was the Universidad de Chile in Santiago, Chile.

Additional information on the expansion of the ***IDD-Brasil*** can be found in de Almeida, et al, 2005. Updated information on CPTEC data products being made available in the ***IDD-Brasil*** can be found in de Almeida, et al, 2006.

IDD-CARIBE

Where Internet delivery of real-time data is not practical, and when a university site is within the NOAAPORT broadcast footprint, the UPC has recommended installation of satellite-based data reception systems. In February 2004, the UPC worked with the Universidad de Costa Rica to install a UPC-designed and built NOAAPORT satellite ingestion system on the UCR campus in San Jose, Costa Rica. Since the installation, the UCR has been able to ingest real-time global observations and NCEP model output for use in education and research. The UCR has agreed in principle that, as its Internet connectivity improves, it will assume a leading role in extending access to its real-time meteorology data to Central American universities that also have sufficient Internet connections. The first steps in this effort are just being taken.

In fall, 2005, the UPC began working with the Caribbean Institute for Meteorology and Hydrology (CIMH), a WMO RMTC, to test real-time delivery of data to Barbados. We have observed that IDD-delivery of data is possible, but not spectacular given the limited network connection (a dedicated 256 Kbps link) that the CIMH currently has to the Internet. This situation will improve as network bandwidth at the CIMH is increased.

The success of the incipient data distribution/sharing efforts among the UCR, CIMH, and Unidata university community, named the ***IDD-Caribe,*** will depend entirely on the quality of network connections available at participating sites.

ANTARCTIC-IDD

More recently, a data relay network has been developed by the US Antarctic research community (Lazzara, et al, 2006). The ***Antarctic-IDD***, built on top of LDM-6, is fully compatible with the Unidata IDD. This system was setup in a test mode and demonstrated in the spring of 2005. The ***Antarctic-IDD*** is growing to include a variety of data sets from a variety of data providers for a variety of users. Currently, the ***Antarctic-IDD*** carries surface and upper air observations, satellite observations and products, as well as numerical model output.

4. CONTINUED DEVELOPMENT

The LDM has proven to be a robust, reliable and portable base on which to build data distribution networks. The LDM history previously presented demonstrates that as design or implementation limitations are identified, new, innovative developments have been employed to keep the LDM viable.

Most recently, the implementation of a four-node Linux cluster (composed of one director and three LDM data servers) as a top-level IDD relay at the UPC offices demonstrated the ability to relay significant amounts of data to downstream sites (Yoksas, et al, 2005). Live stress testing (testing conducted on an "operational" system already feeding data to 220 downstream connections) showed that the cluster was able to relay – on average – over 500 Mbps (5.4 TB per day) of data to downstream sites during a three day trial **without** introduction of product latency. The limiting factor in this stress test was not the LDM software or cluster node performance (in fact, the real servers were essentially idling), but, rather, not having more downstream connections. Peak relay data rates exceeding 900 Mbps convinced us that the limiting factor in the ability to relay data was the underlying gigabit network in UCAR. This test bolstered our confidence that the current implementation of the LDM coupled with cluster technology will be able to effectively relay all of the data desired by the expanding Unidata community for at least the next 2-3 years.

The successes of the LDM-6 have not deterred investigation of alternate approaches to data distribution by the UPC. An investigation of use of the Network News Transfer Protocol (NNTP) implemented by the open-source Internet News (INN) package (Wilson and Rew, 2002) showed promising results. Even though there are many similarities in INN and LDM functionalities (e.g., both provide a push approach to data relay), NNTP/INN addresses several limitations identified in the LDM (Wilson, 2004):

- NNTP routing relies on the *flooding algorithm* in which sites are highly interconnected. Articles flow to sites using massive redundancy such that an article will reach a site by the fastest route possible at that moment.
- News articles flow through the network through a *flooding* algorithm that uses redundant transmission by sending copies to many sites that, in turn, send copies to other sites. This eliminates the problem of maintaining a distribution topology, something that is done by-hand in the IDD.
- Under NNTP articles are classified using a virtually unlimited number of hierarchically structured newsgroups. Articles can be cross-posted to more than one newsgroup, providing multiple views of the same article.
- NNTP supports pull based article retrieval, so that clients can connect to a server and retrieve articles of interest on demand as long as they are available at the server.
- NNTP also supports *control* messages, messages that may initiate processing at a remote site, depending on how the site is configured. This provides a limited degree of network-level configuration. For example, control messages are used to inform sites about additions and deletions to newsgroup hierarchies.
- INN supports both batch and streaming transmission.
- INN supports dynamic creation and destruction of connections to peers based on relay volume.
- In INN, multiple spooling methods can be configured to address a variety of goals such as short and long term storage.
- INN supports authentication and PGP verification.

Use of NNTP and INN is not without problem, however:

- Since the NNTP protocol was originally developed for text products, binary products require encoding before transmission.
- Use of the existing Usenet network would open the possibility of attach in the form of spamming, spoofing, and sending control messages. This problem can be mitigated by developing a network separate from Usenet.
- INN is a large and complex package whose configuration is not for the faint of heart. The configuration impact can be minimized for those sites that have the least resources by their use of reader only software.

With the implementation of the *auto-shifting* feature in LDM 6.4, it is thought that the LDM has advanced about as far as it can given the constraints of the existing protocol. Further major advances in the LDM may require a new protocol that is not tied to the historical client/server approach but is, instead, based on more modern peer-to-peer concepts such as exemplified in applications like BitTorrent (http://www.bittorent.com/). A new protocol and implementation could allow for the following improvements:

- More dynamic creation and destruction of data-product streams.
- Support for access to "one-time" data products (i.e., data-products that are not continuously generated)
- Better load balancing of communication links.
- More adaptive and flexible dynamic routing of data-products with steady-state results that are relatively independent of the configuration of initial connections.
- A better user interface for obtaining data-products. For example, reception of a data-product stream could be started by clicking on a hyperlink in a web page.
- Support for the Windows operating system as well as the usual UNIX variants.

Naturally, minimizing disruption – both to individual sites and to the flow of data – will be a major concern of any new implementation and deployment.

The lessons learned in the NNTP/INN experiment, LDM-6 developments, and investigations of technologies like BitTorrent are being combined at the UPC in the design of a new data relay system (Wilson and Emmerson, 2005). This ongoing activity will provide the underpinnings of a next-generation IDD that will even better serve the international Unidata community.

5. ACKNOWLEGEMENTS

The Unidata Program Center is funded by the US National Science Foundation (NSF).

10. REFERENCES

Cooper, C., 1985: "An approach to Unidata Communications System Design". Final Report to the UCAR Projects Office, Unidata Program Center, September.

Domenico, B. A., 1989: "Unidata Two-Way Communication System". *Proceedings 5th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 208-212.

Campbell, D. P., and R. K. Rew, 1988: "Design Issues in the UNIDATA Local Data Management System". *Proceedings 4th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 208-212.

Green, R. N., 1988: "The Unidata System: Communication". *Proceedings 4th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 177-179.

Fulker, D., 1988: "The Unidata System: An Overview". *Proceedings 4th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 30-32.

Davis, G. P., and R. K. Rew, 1990: "Distributed Data Capture and Processing in a Local Area Network". *Proceedings 6th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 69-72.

Fulker, D. W., 1990: "Unidata: Facilitating Data Access and Use". *Proceedings 6th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 30-32.

Davis, G.P., 1991: "The Local Data Manager, Version 3". Report to the Unidata Implementation Working Group, Unidata Program Center, April.

Domenico, B., 1992: "Weather Information in the Office: The Power (and Perils) of Networking". *Proceedings 8th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 75-78.

Davis, G. P., 1992: "Using High Speed Wide Area Networks for Data Distribution". *Paper presented at the 8th International Conference on IIPS for Meteorology, Oceanography*, and Hydrology.

Davis, G. P., and R. K. Rew, 1994: "The Unidata LDM: Programs and Protocols for Flexible

Processing of Data Products". *Proceedings 10th International Conference on* IIPS *for Meteorology, Oceanography, and Hydrology*, 131-136.

Rew, R. K., and A. Wilson, 2001: "The Unidata LDM System: Recent Improvements for Scalability". *Proceedings 17th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*.

Pugh W., 1990. "Skip Lists: A probabilistic Alternative to Balanced Trees". *Communications of the ACM*, 33, pp 668-676.

Domenico, B., S. Bates, and D. Fulker, 1994: "Unidata Internet Data Distribution (IDD)". *Proceedings 10th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, J15-J20.

Rew, R. K., 2000: "A New Unidata LDM: Scaling up for a Data Deluge". *Unidata Program Center*, August.

Droegemeier, K., 2001: "Project CRAFT: A Test Bed for Demonstrating the Real-time Acquisition and Archival of WSR-88D Base (Level II) Data". *Proceedings 17th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*.

Miller, L. I., 2006: "CONDUIT and Level II Data Distribution: Leveraging that Works for Collaborative Projects". *22nd International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, January.

Chiswell, S. R. and L. Miller, 1999: "Toward Expanding Model Data Access Through Unidata's Internet Data Distribution System". *Preprints, 15th International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, 441.

Ware, R.H., D.W. Fulker, S.A Stein, D.N. Anderson, S.K. Avery, R.D. Clark, K.K. Droegemeier, J.P. Kuettner, J.B. Minster, and S. Sorooshian, 200: "SuomiNet: A Real-Tim National GPS Network for Atmospheric Research and Education". *Bulletin of the American Meteorological Society, 81(4)*, pp 677-694.

Batson, B., 2002: Personal Communication, October.

Yoksas, T. et al, 2004: "MeteoForum – Initial Successes in Data Sharing Leading to the Creation of the IDD-Brazil". *20th International Conference on IIPS for Meteorology, Oceanography and Hydrology*, Seattle, WA, 2004.

Yoksas, T. and D.G. Coelho, 2002: Personal Communication.

Emmerson, S, 2003: "Differences between LDM-5 and LDM-6". *Unidata seminar series presentation, Unidata Program Center, UCAR Office of Programs*, July, PPT.

de Almeida, W., L. A. de Carvalho, S. H. S. Ferreira, D. Garrana Coelho, M. G. Justi, and T. Yoksas, 2005: "Perspectives on Internet Data Distribution Expansion and Use in Brazil". *21st International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, San Diego, CA, 2005.

de Almeida, W., A.L. Cavalcante, A.B. Junior, A.L.T. Ferreira, A.S.A Pessoa, M. Vinicius, L.A. de Carvalho, S.H.S. Ferreira, G.O. Chagas, D. G. Coelho, M.G. Justi, and T. Yoksas. "Data Products from CPTEC Available on the IDD-Brasil". *22nd Conference on IIPS for Meteorology, Oceanography, and Hydrology*, San Diego, CA, 2006.

Lazzara, M. A., G. Langbauer, K. W. Manning, R. Redinger, M. W. Seefeldt, R. Vehorn, and T. Yoksas, 2006: "Antarctic Internet data distribution (Antarctic-IDD) system". *22nd International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, January.

Yoksas, T., S. Emmerson, S. Chiswell, M. Schmidt, and J. Stokes, 2005: "LDM Cluster Stress Testing." Unidata Program Center Internal Report, July.

Wilson, A. and R.K. Rew, 2002: "Exploring an Alternative Architecture for Unidata's Internet Data Distribution". Proceedings 18th *International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, January.

Wilson, A., 2004: "Using News Server Technology for Data Deliver". Proceedings 20th *International Conference on IIPS for Meteorology, Oceanography, and Hydrology*, January.

Wilson, A. and S. Emmerson, 2005: “Next Generation Data Relay Proposal.” Unidata Program Center Internal Report.