Internet Delivery of Meteorological and Oceanographic Data in Wide Area Naval Usage Environments

Udaykiran Katikaneni
Naval Research Laboratory, Stennis Space Ctr, MS, USA 30529
ukatikan@nrlssc.navy.mil

Roy Ladner
Naval Research Laboratory, Stennis Space Ctr, MS, USA 30529
rladner@nrlssc.navy.mil

Frederick Petry
Naval Research Laboratory, Stennis Space Center, MS, USA 30529
fpetry@nrlssc.navy.mil

ABSTRACT
Access and retrieval of meteorological and oceanographic data from heterogeneous sources in a distributed system presents many issues. Effective bandwidth utilization is important for any distributed system. In addition, specific issues need to be addressed in order to assimilate spatio-temporal data from multiple sources. These issues include resolution of differences in datum, map-projection and time coordinate. Reduction in the complexity of data formats is a significant factor for fostering interoperability. Simplification of training is important to promote usage of the distributed system. We describe techniques that revolutionize Web-based delivery of meteorological and oceanographic data to address needs of the Naval/Marine user.

Categories and Subject Descriptors
C.2.5 Local and Wide Area Networks - Internet

General Terms
Design, Performance, Reliability

Keywords
meteorological and oceanographic data, resumable object streams

1. INTRODUCTION
Timely provision of meteorological and oceanographic (MetOc) data is essential for effective Naval operations. Access to and retrieval of MetOc data from heterogeneous sources in a distributed system such as the Internet presents many issues. Among these issues, effective bandwidth utilization is important for any distributed system. Bandwidth utilization is of particular concern to fleet operations. Assimilation of spatio-temporal data from Web-based sources means that differences in datum, map-projection and time coordinate must be resolved. Reduction in the complexity of data formats is a significant factor for fostering interoperability. Simplification of training is important to promote usage of the distributed system. All of these concerns directly affect the Naval/Marine user's (Warfighter) ability to effectively access, collect and share data/information across the Internet.

The future requires intelligence communities and military operations to rely more heavily on automated Web-based solutions for the delivery of MetOc data and products to the Warfighter. Many sources reference the growing need for this capability (e.g., the Department of Defense Net-Centric Data Management Strategy [1]).

These issues are being addressed by Tactical Environmental Data Services (TEDServices) [2]. TEDServices is being engineered by the Naval Research Laboratory (NRL), the Naval Oceanographic Office and the Naval Undersea Warfare Center, with sponsorship from Space and Naval Warfare Systems Command (SPAWAR) PMW-155. The Naval Research Laboratory's Geospatial Information DataBase System (GIDB™) serves as the prototype for TEDServices system components[3, 4]. It is currently implemented in Ozone, an open-source object-oriented database management system [5]. Data is accessible over the Internet via a Java Applet [6]. It includes a communications portal that enables users to obtain data from a variety of data providers distributed over the Internet in addition to the GIDB. The GIDB communications portal establishes a well-defined interface that brings together such heterogeneous data and provides a common geo-referenced presentation to the TEDServices user.

The system has been shown to be effective in reducing delays in data access. Data transmission is automated in a system-to-system publish/subscribe approach. Data is immediately available for transmission to other nodes in the system on receipt by any given node in the system. Daily updated data transmissions of 20 MB of highly compressed data occur in less than 3.25 minutes. Previous systems were not Web-based, provided no data compression advantages and required end-user reach back on a continuing basis to query for data updates.

Web technologies, primarily HTTP and Java Servlets, are used to create a distributed Web service in which services communicate with other services in an automated manner to maintain a data delivery system to support warfighters. The Web technology of HTTP and Java Servlets allow us to reliably transmit through firewalls in an authorized manner and provide an effective data stream for transmitting data compressed through advanced compression techniques. So HTTP was used for all data communications, requests and data transfers. Major motivations for HTTP usage were the firewall issues and building on previous GIDB Portal design. This provided compatibility with current and future DOD requirements and allowed a speed to capability enhancement rather than an ad hoc custom approach.

2. TEDSERVICES
NRL's extensive work on GIDB was leveraged for the design and development of TEDServices. TEDServices is a new, scaleable and modular environmental data delivery system, designed to support Warfighters, weapon systems, and expert MetOc data users. It includes a middleware infrastructure that enables the interoperable transport and transform of data. This is accomplished in a manner consistent with WGS84 datum and universal time coordinate, facilitated by a MetOc/Mission Rules Based Data Ordering scheme (MRBDO).

TEDServices provides a new Internet-based architecture within the Oceanographer of the Navy's (N096) Operational Concept 2002 [7

WWW 2004, May 17-22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

]. This is organized in the following manner. Production Centers (e.g., FNMOC and the Naval Oceanographic Office (NAVO)) produce Numerical Weather & Ocean Prediction (NW&OP) data by assimilating global In-Situ Data. Domain Authorities (DA) use NW&OP within an expert knowledge context, to derive the "MetOc Answer" and populate the Domain Authority Virtual Natural Environment (VNE). A Domain Authority can be co-located within a Production Center. Centers of Expertise (CoE) will use data from the Domain Authority VNE to produce global CoE Products. A CoE can be co-located within a Production Center or Domain Authority. A Remote User can be ashore, afloat or mobile and will use data/products from Domain Authorities and Centers of Expertise. Also a Remote User will also collect in-situ data to be used by Production Centers and Domain Authorities. Automated ingest and publish, together with data subscription capabilities provide the means for data delivery throughout the system.

The TEDServices design supports the automated management and bi-directional transport of meteorological, oceanographic and other environmental data/information. TEDServices offers a lightweight, forward deployed data cache, which provides Warfighters, MetOc professionals, Tactical Decision Aids (TDAs), applications and weapon systems immediate access to the Virtual Natural Environment (VNE), a 4-dimensional representation of the User-defined battle-space environment. TEDServices' Clients will use a new MetOc/Mission Rules Based Data Order (MRBDO) process to subscribe to relevant data by mission, platform, TDA/application, parameter or product. The design tenants of TEDServices include: Data Transport (to reduce bi-directional bandwidth use), Data Management (to simplify data ordering and forwarded deployed data administration), Data Representation (implementation of a unified Geospatial and Time Coordinate Process), and DoD Joint Interoperability (supporting standards defined by the Joint MetOc Interoperability Board).

TEDServices offers a pure Java implementation for platform independence. It also provides planned support for the Joint MetOc Interoperability Board XML Interface Standard (Joint MetOc Broker Language – JMBL). A feature of TEDServices is the provision for remote administration of the system by authorized users.

3. TEDSERVICES COMPONENTS

The conceptual components of TEDServices are shown in Fig. 1. These include GateWays, Local DataBrokers, Local DataStores, and Interface support. These components are explained below.

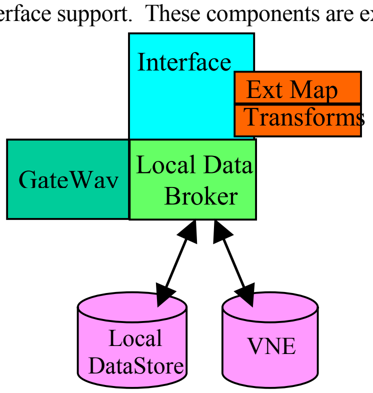

Figure 1: TEDServices Conceptual Components

The Local Data Broker (LDB) embodies the "smarts" of the system to pre-stage needed data at a particular location. It "knows" how to contact other TEDServices GateWays to request needed MetOc parameters/products over particular areas of interest. The LDB also monitors data usage and cancels further delivery of data that is not being used. The LDB also works to mitigate redundant reach back requests for the same data by multiple users. It does this by caching data so that the data is available to multiple users.

The GateWay component encapsulates the software that streamlines the process of integrating data from heterogeneous sources to a Common Transport Format (CTF). CTF is a well-defined, serialized Java object that is capable of describing a variety of data (gridded volumes, point data and imagery) in a consistent manner. This CTF assures a uniform datum, uniform projection and universal time coordinate. A CTF for all data types within TEDServices simplifies format transformations to end-users and is a significant factor for fostering interoperability. The MetOc/Mission Rules Based Data Ordering system allows data requests to be aligned with relevant mission specific packages and platforms. This reduces the likelihood of requests for data that are not essential to a particular task. It also offers a means of simplifying training.

The Interface component is responsible for a number of tasks, including:

- Receiving user requests for data and products.
- Handling user requests to obtain data in a number of supported file formats (e.g., netCDF, draw for FalconView, ShapeFile, MIFF, etc.).
- Interpolating gridded data to a user-specified spatial resolution.
- Establishing an order (subscription) for data to be forward-deployed at the platform or location.

The Local DataStore/VNE provides a forward-deployed object-oriented cache for data and products, which are accessible via the Interface. This cache uses the Java Ozone OODBMS and allows for remote administration.

Together, the TEDServices components form a TEDServices GateWay. There is ideally one GateWay per platform or location which serves all users and applications at that platform or location. This obviates the need for multiple reach backs for the same data by multiple users. Obviating this type of reach back serves to reduce bandwidth usage. TEDServices GateWays communicate with each other to forward deploy needed data to the end-user. User applications access data only from their local TEDServices GateWay.

4. RESUMABLE OBJECT STREAMS

Large scale data transfer can be difficult when network communications are unstable. TEDServices employs Resumable Object Streams (ROS) for all data traffic between major components across the network to achieve fail-safe data transportation under these conditions. ROS allows either the client or server side of a request to lose network connection, regain it, and the request will continue where it left off. In the event of a server shutdown and restart, server side processing of requests does not require the client to resend the request. Retransmission of the previously transmitted portion is not necessary in either case. Data requests can still be wrapped in compression and/or encryption. The ROS transmission controls add almost no storage overhead to the communication (approximately 13 bytes).

ROS utilizes control request types, including ones to obtain a process id, execute a process, obtain a failure index, resume execution of a process, obtain process status, resume a response and end transmission. For each new process or method invoking a ROS transmission, ROS provides a unique process id that is used by all other control request types. When lost network communications are re-established, this process id is used to identify the correct data stream. The execute process control type sends the interface command name and parameters to the server for invocation. The failure index lets the client know where to restart transmission in case of a failure. Resume process execution restarts the parameter transmission in case of failure. Process status checks the status of a running request. The resume response control type allows the client to resume downloading a response from a request at a specified index in transmission byte stream. Finally, communications are ended with the end transmission control type.

5. COLLABORATIVE APPLICATION SHARING

A Collaborative Application Sharing Process (CASP) is implemented in TEDServices to enable remote application users to share the state of their applications as well as to share information across the Internet. This means that some of the mission planning requirements can be placed at "Centers of Expertise" where experts can perform some of the less time-critical planning and provide results to the field. This allows, in a U.S. Navy setting, heightened situational awareness in a distributed environment.

When CASP is used to share an application's state, users send to TEDServices a Java object that encapsulates the state of their application. This state is stored within TEDServices in a non-application specific manner. When the object is retrieved from TEDServices, the remote user may open the CASP object. This will restore the application's state to the state contained within the CASP object. The user may then make any appropriate modifications and re-submit the object back to TEDServices for further dissemination and sharing.

The model for CASP is a publish and subscribe paradigm. A typical CASP scenario follows. One or more parties subscribe to a particular CASP product from a remote TEDServices GateWay. When the CASP product is published to that GateWay, the GateWay automatically pushes the product to all subscribing parties. All subsequent updates to the CASP product are automatically pushed to the subscribing parties as well. Subscribing parties that received the CASP product modify the CASP product based on local knowledge. Then, they re-publish the updated CASP object back to TEDServices for further dissemination and sharing

6. RESUMABLE OBJECT STREAMS IN THE FLEET BATTLE EXPERIMENT

In April 2003, TEDServices was demonstrated in Fleet Battle Experiment – Kilo (FBE-K). TEDServices GateWays were installed at the following locations: the aircraft carrier Carl Vinson, NRL Monterey (FNMOC), the Naval Oceanographic Office (NAVO), NPMOC-Yokosuka, NPMOC-Pearl, the Naval Undersea Warfare Center (NUWC), the Fleet MetOc Advance Concepts Lab (FMACL), Guam and NRL Stennis Space Center. Atmospheric data was transferred from FNMOC to the TEDServices GateWay at NRL Monterey where it was automatically ingested into TEDServices. Similarly, oceanographic data was transferred within NAVO to the TEDServices GateWay there, where it was ingested into the TEDServices VNE. All TEDServices GateWay to GateWay communications occurred on the SIPRNET.

The effectiveness of TEDServices bandwidth utilization is highlighted in this fleet battle experiment. NAVO delivered 3.2 GB of raw oceanographic data to the TEDServices software located at NAVO over the eight days of the exercise. TEDServices then compressed this data down to 1.2 GB. Total transmission to the TEDServices node at Pearl Harbor based on needed parameters and area-of-interest totaled only approximately 500 MB. TEDServices data compression and "smart" system push of only needed data resulted in an 80% reduction in bandwidth utilization.

NPMOC-Pearl subscribed to parameters being published at the NAVO TEDServices GateWay. Upon receipt of these parameters, NPMOC-Pearl used them as first guess fields to run the Modular Ocean Data Assimilation System (MODAS). The value-added data was then published at the NPMOC-Pearl GateWay and was then pushed to other TEDServices GateWays (Carl Vinson, Fleet Mac Lab, GUAM and NUWC) based on their data subscriptions for particular parameters and areas of interest.

During FBE-K, ROS was configured so that it would try to complete a transmission for up to two hours. According to these numbers, that was not long enough. FBE-K drove home the point that network centric warfare is not a perfect world. Loss of network connectivity for <u>extended</u> periods of time must be considered. In the course of FBE-K, the following incidents occurred that affected connectivity for intervals of more than two hours at time. Each of these incidents occurred in different locations.

- On an afloat platform, the local SIPRNET went down.
- A network cable was unplugged from a TEDServices GateWay in order to share with another machine
- A cleaning crew accidentally turned off the power to a TEDServices GateWay.
- A catastrophic hardware failure occurred on one of the TEDServices GateWays.

This exercise highlighted that particular attention must be paid to development for an automated system. ROS was sufficient for a window of opportunity - if the network went down or the receiving server went down, ROS guaranteed delivery within a two hour window.

After the configurable retry limit is exceeded, ROS times out and drops the transmission. Something more is needed in addition to this retry interval. The exercise helped in identifying a weak point in the delivery process. This has been fixed within the TEDServices Local Data Broker. After a timeout, the LDB puts data to be sent to the remote gateway into a queue and periodically monitors the network for restored connectivity. The transmitting GateWay will resume transmission of data (using ROS) once the receiving GateWay comes back on-line. In the latter case, only the most current data is sent.

The version of ROS used in FBE-K worked well for network hiccups and network outages that were shorter than two hours. During FBE-K, eight cases were recorded where network connectivity was lost in mid data transmit. When network connectivity was regained, ROS continued the data transmission at the point where it was interrupted. It did not retransmit the entire data transmission, only the portion of the data that was not transmitted initially. In other cases, ROS guaranteed delivery by establishing initial connectivity prior to sending data as the remote TEDServices GateWay was unreachable upon initial attempt to transmit data.

Figure 2 is a scatter plot that compares the sizes of the NAVO TEDServices GateWay transmissions and amount of time required to successfully transmit the data. This plot includes transmissions of oceanographic data for a total 711 transmissions. While some of the transmit time variations can be attributed to network latencies, it is reasonable to conclude that the transmissions in the plot that required longer than 2.5 minutes to transmit are the cases where ROS was guaranteeing the delivery of data within a specified maximum try time interval. That is ROS continued attempting to deliver transmission data for the time interval specified and encountered network connectivity issues.

There were forty-two transmissions that took longer than 2.5 minutes to successfully transmit. Note in particular the transmission on the upper left of the graph. The transmission was about 500 kb and took 10 minutes to guarantee delivery. This transmission would most likely have been dropped without ROS.

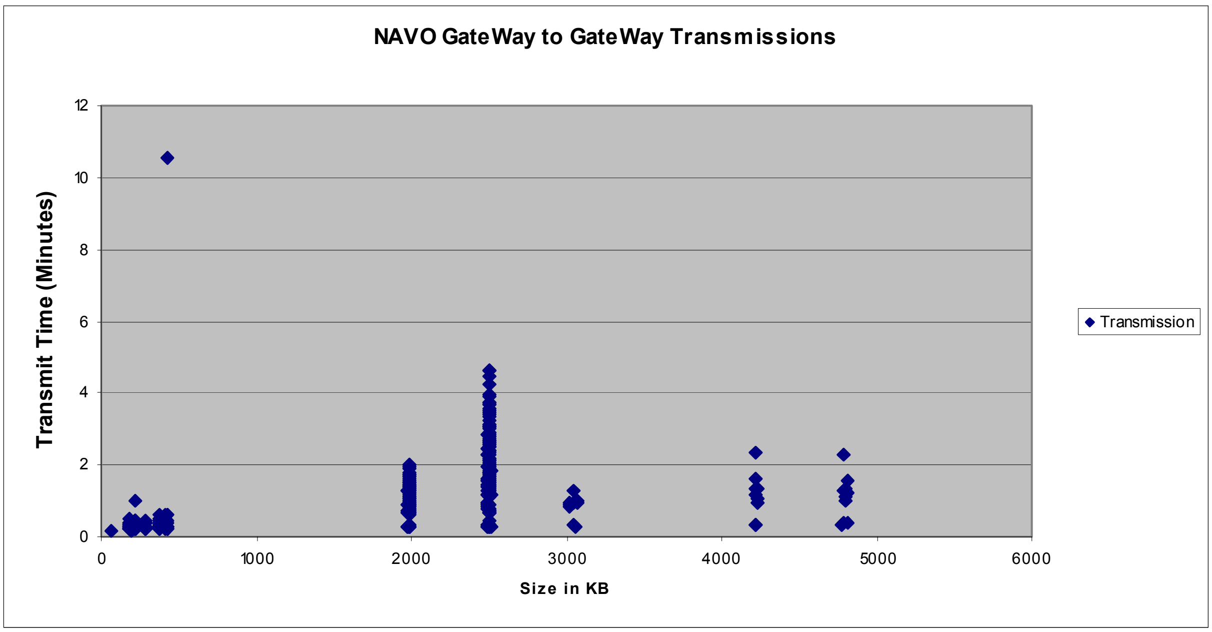

Figure 2: NAVO TEDServices GateWay Transmissions

ROS is a built-in feature of all GateWay-to-GateWay transmissions. Therefore, ROS was utilized in each successful transmission. The amount of ROS assistance provided was dependent upon the connectivity issue encountered. If no connectivity issue was encountered, then minimal ROS assistance was provided.

Of the G2G transmissions where ROS guaranteed delivery, most required that ROS assist in establishing initial connectivity prior to sending data as the remote TEDServices GateWay was unreachable upon initial attempt to transmit data. Some of the G2G transmissions lost connectivity in mid data transmit. In those cases, ROS regained connectivity and transmitted the remainder of the data to the remote TEDServices GateWay. There were eight of these incidents during FBE-K. In the case of NAVO NCOM data, the transmission could have been as large as 6.6 MB. Whether there was no initial connectivity or whether connectivity was lost in the course of transmission, without ROS, the transmissions would have been dropped.

Figure 3 is another scatter plot for the Pearl Harbor TEDServices GateWay transmissions. Here there is a total 5463 transmissions of oceanographic and atmospheric data. On the left side of the plot, note that there are some small

transmissions that took just over 20 minutes. Again for transmissions taking more than 2.5 minutes are where ROS assistance was required in order for the data to be delivered. Note the many 1 kb transmissions that required ten minutes or more. ROS continued attempting to deliver transmission data for the time interval specified and encountered network connectivity issues. Within the 10 - 25 minutes, ROS was able to gain initial connectivity or regain broken connectivity to deliver the data.

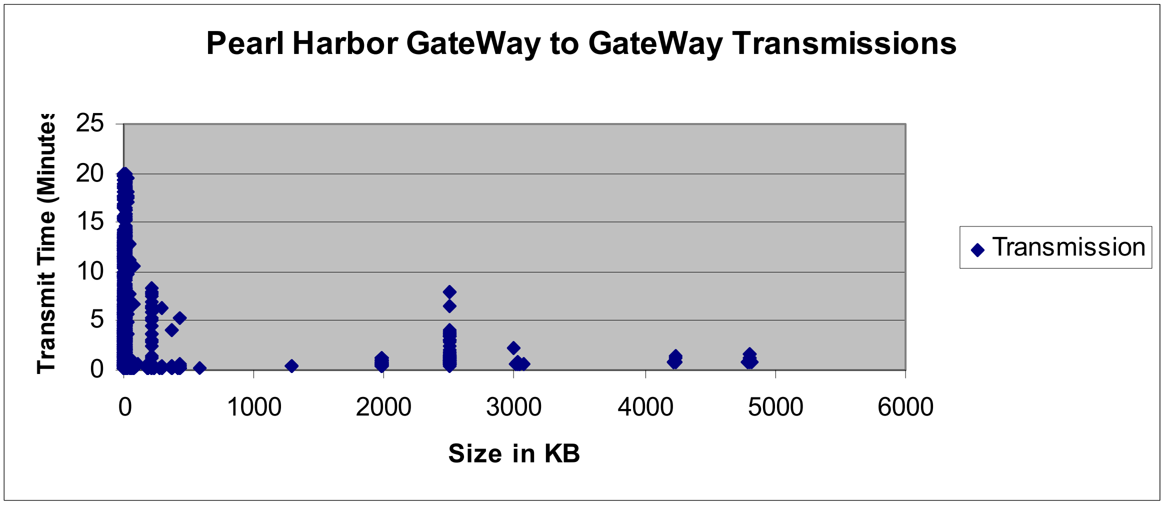

Figure 3: Pearl Harbor TEDServices GateWay Transmissions

7. SUMMARY

We have discussed issues pertinent to the Web-based delivery of MetOc data for network centric warfare. We have shown how TEDServices employs a number of advanced techniques for improved management and better Web-based delivery of MetOc data to the Naval user.

We have covered means for improving bandwidth usage and methods to resolve differences in datum, map-projection and time coordinate. Techniques for the reduction in the complexity of data formats were presented. System aspects that use a rule base to simplify training and also reduce bandwidth utilization were explained. We have also shown how TEDServices uses automated data ingest along with a publish and subscribe paradigm to reduce end-user interaction for acquiring data. Methods for handling large scale data transfer and collaborative application sharing were also discussed. Finally, we have shown how TEDServices and the effect of resumable object streams were demonstrated on the SIPRNET in Fleet Battle Experiment – Kilo.

8. ACKNOWLEDGMENTS

We would like to acknowledge the MetOc Systems Program Office, SPAWARSYSCOM PMW150, for sponsoring this research, and the TEDServices Engineering Working Group for their contributions to TEDServices.

9. REFERENCES

[1] DoD CIO, "Department of defense net-centric data management strategy", Draft, March 28, 2003.

[2] TEDServices: https://teds.navy.mil.

[3] Cobb, M., Foley, H., Wilson, R., Chung, M., Shaw, K., "An OO database migrates to the web," IEEE Software, vol. 15, pp 22-30, 1998.

[4] Needham, D., Wilson,R., and Shaw, K."An analysis of a CORBA-based approach to accessing geospatial information via the internet", Proc. 20th IEEE Int. Performance, Computing, and Communications Conf. (IPCCC 2001), pp. 236-243, April 4-6, 2001.

[5] Ozone: http://www.ozone-db.org.

[6] Chung, M., R. Wilson, R. Ladner, T. Lovitt, M. Cobb, M. Abdelguerfi, K. Shaw, "The Geospatial Information Distribution System (GIDS)", Chapter 17, in Succeeding with Object Databases, A. Chaudhri, ed., John Wiley , New York, 2001.

[7] Oceanographer of the Navy, "Naval Oceanography Program – Operational Concept", March 2002.