# 11.6 DESIGNING AN FTP SERVER FOR NOAA/NESDIS MODIS NEAR-REAL-TIME DATA

Kristina Sprietzer* and Paul Haggerty
Science and Technology Corporation, Suitland, Maryland

Gene Legg
National Oceanic and Atmospheric Administration, Suitland, Maryland

## 1. INTRODUCTION

The Near Real Time MODIS processing system was conceived as a proof of concept program, and implemented under the Computer Sciences Corporation (CSC) Central Satellite Data Processing (CSDP) contract. It was designed to provide NOAA with rapid access to MODIS environmental data from the TERRA and AQUA satellites and to provide risk reduction for future high data rate satellite systems by identifying the risks and problems associated with high data volumes and Near Real Time operations.

## 2. THE MODIS INSTRUMENT ON TERRA/AQUA

The MODIS sensor is sensitive in 36 wavelengths of the spectrum, ranging from the visible to the thermal infrared, with spatial resolutions of 250 meters, 500 meters, and 1 kM, depending on the particular band. MODIS has a greater spectral and spatial resolution than previous NOAA satellite instruments. This greater resolution produces more data and increases the size of the data sets. Each data product covers 5 minutes of satellite observation. Data compression reduces data storage requirements to one third of original size. (See Table 1 for data set size). Multiplying the total of all products by 288 granules per day results in a global coverage consisting of approximately 545,000 MB (or 545 GB). With two satellites, the global processing system requires over 1 TB of storage per day.

## 3. MODIS DATA PROBLEMS

The larger data sets produced by MODIS would strain the capacity of NOAA's distribution centers, such as CEMSCS which has relatively slow transfer rates, which discourage handling large data products, and a high overhead fee associated with data storage. However, archiving of MODIS data products is limited and generally unnecessary. Since NOAA NRT MODIS is a near-real-time system, new data are constantly available. Furthermore, NASA's MODIS processing system serves the science community with research-quality archived data, which is easily accessed by the general public. Because of this, NOAA's NRT MODIS program saves only as much as is necessary to give customers a reasonable period of time to acquire it, without exceeding our limited storage capacity. Because of the storage and associated bandwidth problems, the NOAA NRT MODIS project decided to try to handle the distribution of data with an Origin 3200 SGI with an attached 5 TB RAID system as a short-term solution.

| PGE | Description | File Size (MB) | Compressed File Size (MB) |
|---|---|---|---|
| 1 | L1A | 450 | NA |
| 2 | L1B | 1000 | 500 |
| 3 | Cloud Mask | 61 | 25 |
| 4 | Atmosphere | 12 | 1.8 |
| 6 | Cloud Properties | 13.5 | 4.1 |
| 7 | Snow Cover | 32 | 6.8 |
| 8 | Sea Ice | 13.5 | 3.1 |
| 9 | Ocean Color | 278 | 14.8 |
| 10 | Sea Surface Temperature | 33 | 10 |
| Total | | 1893 | 565.5 |

**Table 1: Data Storage Requirements for MODIS data**

### 3.1 *Storage and Bandwidth*

The file sizes listed in Table 1 are for 5 minutes of observational data, which is referred to as a granule. Approximately 6 daytime granules cover the continental United States (approximately 11 GB uncompressed). Global coverage involves 288 granules (approximately 550 GB). If 10 customers requested global coverage, and tried to download 550 GB daily, this would impose a network load in excess of 5 TB of data. Obviously, once customers are accessing the storage device, bandwidth becomes a serious issue in addition to simple storage.

### 3.2 *Bandwidth*

Proper hardware for increased bandwidth needs to be present on both the local machine (NOAA MODIS) and the remote machine. The local machine's gigabyte network card is of no use to the possibly old network card on the customer's machine relying on antiquated and saturated network lines. Also, the routing devices used throughout the global network between the two machines must be properly sized as well. Unfortunately, this means you must contact every system administrator of each router to see if you can gain increased performance access (where available). When working with government or military hardware, the information for these devices may be hidden by firewalls, making contacting system administrators difficult. A test of sending products over regular

* Corresponding author address: Kristina Sprietzer, STC, NOAA/NESDIS/IPD Federal Center FB-4, 4401 Suitland Road, Suitland, MD 20746; e-mail: Kristina.Sprietzer@noaa.gov

internet access proved how limited ordinary routing is for this project (150 MB 3-11 hours).

### 3.3 *Product Size*

Using gzip compression has decreased the file size by half or more. Another way to decrease the file size is to remove channels that the user does not intend to use. The MODIS instrument collects 36 channels, and many of the customers do not use all of them. For optimum bandwidth performance, it would be advisable to subset the channels on the server. However, each customer may require a different subset. Other questions that arise are how much processing time and power does it take to subset? These factors can also affect bandwidth. Once the file is subset, should it be stored for future customers who might wish to download it? Storing the files would add even more to the current storage volume, but it may save in bandwidth by not reproducing the files. These questions are still being investigated for this project to determine optimum performance.

### 3.4 *Area of Interest*

Area selection could also reduce the amount of data that are sent to customers. For example, customers interested in ocean products do not need to be sent land images. Additionally, customers interested in imaging data do not need nighttime data. These are all possibilities of how data can be requested. Currently, customer information is stored in text files. Each additional option given to a customer results in a decrease in transfer bandwidth, but increases the complexity of the customer's information. Increasing the customer's options of data selection decreases the amount of data sent, and the customer's data selection information quickly grows to the point where it will be useful to integrate a database system.

### 3.5 *Web Form Options*

Most requests are handled manually. However, staffing for the project is only during business hours. Customers have requested a web form to add new areas, products, or subsetting products. Perhaps they want to start an ongoing subscription or perhaps they want a one-time download. This form would list all the possibilities to the customer in order for them to contemplate it off-line and give task personnel the necessary information to input their selections into the processing system.

### 3.6 *Security*

In an ideal world, the above form would instantly make the appropriate changes and the customer would start receiving those products. Currently, we are not doing that for security reasons. We have a processing server that processes data and ftps the data to our clients. The NOAA MODIS FTP server is now used for storage and for customers logging in to transfer files. In the near future, our current system will shift the responsibility of transferring files to customers to the dedicated ftp server.

Clients are welcomed to compete with each other for bandwidth on the FTP server without slowing down our processor. Anytime a server is open to an incoming client, security will need to be lowered to allow access. In order to have a secure interactive form, all input data from the form must be checked carefully to not allow hacking, and task personnel need to stay abreast of the latest information for new bugs in case the currently accepted method has been shown to fail. Furthermore, even if the information is genuine and correct, it is possible for a customer to ask for more data than it is possible to provide over the networks. Without review, a customer could easily bring the entire network to a halt simply by asking for all available data.

## 4. SUMMARY

When processing large data sets there is a concern in transferring the product to customers. The MODIS data is on the technological edge of current communications. The problem does not involve inexpensive changeable parts like thousand dollar PCs and cable modems. The problem involves $5 million super computers and intercommunications for buildings that are several decades old. Customers need to be highly selective in requesting the data products they need. When designing ftp servers, one must be aware of the security risks involved and continue to be up to date on security bugs and in various methods to improve the system's security.