Proceedings of ICRC 2001: 2325 © Copernicus Gesellschaft 2001

ICRC 2001

The automatized system for the collection, treatment and analysis of data in real time

A. A. Turpanov, S. A. Starodubtsev, V. G. Grigoryev, V. I. Kozlov, V. S. Nikolayev, and A. N. Prikhodko

Institute for Cosmophysical Research and Aeronomy SB RAS, Yakutsk

Abstract. The real-time neutron monitor data from different stations presented on the Internet are necessary for producing forecasts of space weather, but interactive access to data is not appropriate for programs performing this task. The work on short-term forecast of space weather showed that the most reliable results might be obtained when using data from several appropriate stations. Such an access might be possible via Internet provided there would be a common data representation format and program interface. A library developed for Yakutsk database access directly from Fortran programs is the first attempt to achieve this goal.

1 Introduction

At present data from many neutron monitors are available via Internet as real-time data plots. A lot of links to particular web sites may be found for example at Moscow neutron monitor web (Solar-Terrestrial Division of IZMIRAN, web site). While graphical representation of data provides quick and convenient description of the situation for a human, it's not usable for numerical analysis. Numerical data are also presented on the Internet, but by much smaller number of stations. Present interest to space weather forecasts require cosmic ray data to be taken into consideration as far as cosmic rays themselves are treated as significant component of interplanetary medium. As it was stated in Kozlov et al. (2001) data from several stations should be used in analysis in order to extract more information from cosmic ray data.

The Tixie Bay cosmic ray (CR) station operated by the Institute for Cosmophysical Research and Aeronomy had an automated system for data registration, processing and diagnostics of interplanetary disturbances since early 1984 (Kozlov et al., 1984). Now we have a new data registration and storing system for both Yakutsk and Tixie Bay stations, presented on the web site (Yakutsk Space Weather Group, web site), where not only graphics, but also numerical data are shown and are available for downloading. The usage of data in space weather forecast tasks requires a kind of application programming interface (API) to the database in order to give researchers an instrument for data access from their programs. The first attempt to implement such a facility for our database was made during last year and is described below.

Correspondence to: A. A. Turpanov
(turpanov@ikfia.ysn.ru)

2 The registration and data storing system.

The neutron monitors at Yakutsk and Tixie Bay CR stations consist of 18 standard counters in three sections. The counters are connected to the interface card in PC running OS Linux. The interface card and the low-level driver software was designed and made at the Institute for Cosmophysical Research and Aeronomy. A standard digital barometer connected to PC over serial port measures the atmospheric pressure. The time accuracy is supported by NTP synchronization to public servers stratum 1. The registration software consists of two programs coordinating their activity through operating system's IPC message passing facility. The first program polls the counters interface card and barometer once in 5 minutes, writes the data in raw hexadecimal format to a file and sends a message to the second program to read the data. The second program is responsible for passing data to the database server. It reads the new portion of data, stores it in memory buffer and waits for the next message. After a certain number of new data portions defined in configuration file were read, the second program calculates the MD5 digest to ensure the data integrity and sends an e-mail containing signed data to the sever machine. The SMTP and standard MTA — sendmail is used for data transport to avoid direct handling of possible failures of network connectivity.

The communication layer differs for Yakutsk and Tixie Bay stations. The Polar Geophysical Observatory is connected to Internet via leased line, but the station itself is located about 7 kilometers far from Tixie and has no cable connection to the observatory. The network connectivity is pro-

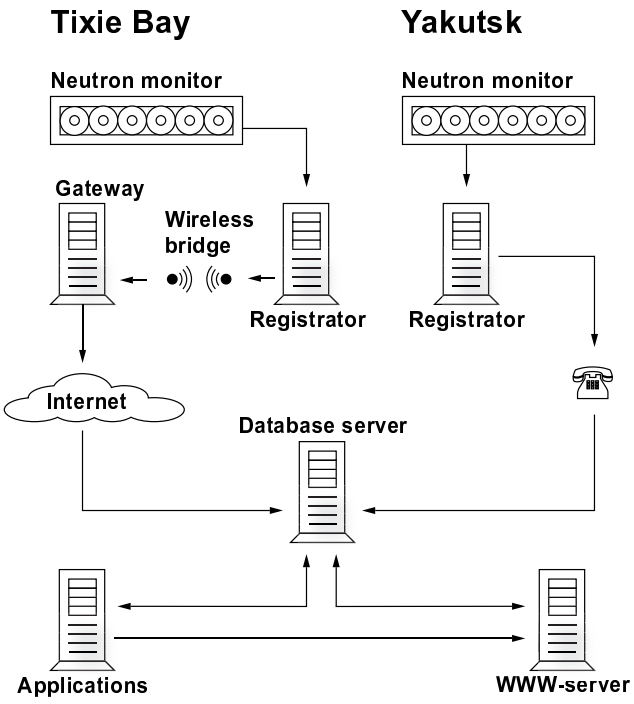
vided by a three-point wireless bridge, which was installed in October 2000. Thus the Tixie neutron monitor is now always online and sends the data each 5 minutes. In Yakutsk the station has only a dial-up connection to the Institute via a telephone line shared with the voice phone, so it sends the data once per hour.

On the server side the mail is checked for integrity, decoded and inserted into primary data tables. Responsible investigator revises the primary data once a day. The failures found are either replaced by interpolated values or zeros, if the interpolation is not reasonable, and then the portion of revised data is stored in the main tables. Upon data insertion the stored procedure at the database server calculates the pressure-corrected count rates. The database "Interplanetary Medium" (ipm) consists of original data from Yakutsk and Tixie Bay neutron monitors and a number of mirrored data from Lomnitski Stit neutron monitor, Oulu neutron monitor, Omni data set, Proton Flux Data from Helios-A and Helios-B Spacecrafts (Experiment E7). These data were collected together to have necessary information for many research tasks with single access method.

Tables for neutron monitor data have the following fields:

Field	Type
Date and time	timestamp
uncorrected count rate	float4
atmospheric pressure, mb	float4
corrected count rate	float4
flag	int2

Tixie table contains about 1.5 million records of 5-minute averaged data starting from 1980-01-01. Yakutsk table contains about 250000 records of 5-minute averaged data from 1999-01-01. The DBMS we use is the PostgreSQL server running on dual Pentium III 650 machine with SCSI disk subsystem on Linux platform. Though PostgreSQL doesn't directly use the multiprocessor system, processes it creates for each session can easily run on different processors and share the cache and loaded functions code, so the overall performance of the system is fairly high. The topology of the whole system is shown in the figure.



3 The software for data access.

In general, the relational data model is not very well suited for handling time series experimental data. For example this model doesn't have a concept of the "neighbor" data element, so to calculate, say, a derivative one has to run a rather high-cost query to find data element next to the given one. Still RDBMS' have an advantage of providing very effective access to data portions within a given range of indexed field, what is one of the main tasks for data storage systems. To provide convenient way to retrieve data and some aggregate values the database must have specific interfaces for different tasks. For interactive on-line access we have a WWW-interface (Yakutsk Space Weather Group, web site). It provides access to near real-time raw non-revised data from Yakutsk and Tixie Bay station, the revised daily updated data from the above stations and the mirrored data, which are extensively used in research work on methodology of the short-term space storms forecast. Our WWW- interface to the database provides the data retrieval based on time limits and presents data as graphic or in ASCII both on-screen and downloadable. The work is in progress on development an interface for cross–tables queries with additional con

While this kind of interface is quite acceptable for interactive access and is common for many stations, it's not suited for research work involving scanning of large amounts of data. The best way from our point of view is to enable investigators to have direct access to binary data from user's program without any disk operations. To achieve this goal we've developed a library *libipmg77* of routines, which implement basic classes of queries. As far as Fortran is traditionally used in scientific programming, this library is oriented on Linux GNU Fortran-77 and soon will be ported to Win32 platform. The library is written in 'C' language using the PostgreSQL API - *libpq*.

The PostgreSQL is an open-source system, so it was possible to create some useful SQL extensions such as commonly used aggregate functions - variance and standard deviation.

The general scenario of data access with use of *libipmg77* functions consists of the following steps:

> The user's program calls the function and passes it the parameters for constructing the query and the references to arrays for results.
>
> As far as number of fields in query and the number of respective arrays are not fixed, the function first parses

the strings and numerical parameters, defines the number of optional arguments and constructs the text of SQL query. Then it connects to the database server and sends the query.

The server runs the query and returns the result to calling function.

The function extracts the result, determines the binary types of its fields, sets the appropriate pointers to user's arrays, fills them, closes the connection to the server and returns to user's program.

At present *libipmg77* includes three main functions and several small utilities. The main functions are:

ipmbas() - the function that performs the queries for retrieval of user-defined data set matching the optional user-defined condition from the user-defined time range:

```
SELECT FieldsList FROM Table
WHERE DateTime BETWEEN DateTime1
AND DateTime2 AND Cond
ORDER BY DateTime
```

ipmagg() - the function that performs similar query for aggregate values.

ipmseg() - the function that searches the user-defined time range for the first continuous segment of data matching the user-defined condition.

Testing of library showed the dramatic increase of the overall performance in comparison with traditional way of working with data files. The further improvement of performance may be achieved for the class of programs that scan the large time interval of data with a window of certain width. If such a program keeps the opened connection to the database server, the scenario might be as following:

The program calls a function that performs the first query; returns result to the calling program and stay resident.

The main program makes the next function call and starts working with data.

The function gets the data from server, stores it in the main program's memory, and waits for the next call.

After processing the first portion of data main program makes the next function call and starts processing the data from second query already stored in it's memory.

This algorithm was tested and showed an outstanding performance: in fact the main program waits for data only once and then continuously works with data already stored in memory.

The library *libipmg77* is already used in research work on studying short- and long-term variations of CR; processes of CR interaction with solar wind turbulence near fronts of the interplanetary shock waves; ground level enhancements; for calculation of barometric coefficients for Yakutsk and Tixie Bay CR stations, etc. The results presented in Kozlov et al. (2001a,b) were obtained using this library. The program for short-term forecast of space storms, working in automatic regime also uses the *libipmg77* for getting new data from database server.

4 Discussion

Real-time neutron monitor data presented on Internet make it potentially possible to use data from many stations in research and production, especially in projects concerning space weather. But programs that should run automatically need common convention for external data representation format and common interface for data access. While the first task is rather simple to accomplish, the second one requires the certain work to be done on design and implementation. In fact, the ideal model should be the distributed database of neutron monitor data. As far as different DBMS's are used at different locations, the distributed database should have either replication servers or gateways, say at WDC's for Solar Terrestrial Physics, for providing single interface for binary data access. Further, the library for retrieving data from replication server or gateway should be developed and ported to all commonly used platforms and made available for scientific community.

The Yakutsk group is ready to participate in such a work and share it's experience if the movement will actually arise. The INTAS 2000 project 752 "Key Parameters for Space Weather" includes task T1.1 with nearly the same goal " . . . development of common standard of experimental data storing and handling, . . . ", so the discussion may be held together with participants of this project.

Acknowledgements. We are grateful to all corresponding data sets owners for kind permission to mirror data for convenience of local use.
We are also grateful to the PostgreSQL Global Development Group for providing public access to high-quality open-source database management system.
This work was supported by Russian Foundation for Basic Research grants (the leading scientific school by G.F. Krymsky 00-15-96669 and R98 - the Arctic 00-02-96204).

References

Solar-Terrestrial Division of IZMIRAN web site
http://helios.izmiran.rssi.ru/cosray/main.htm

Yakutsk Space Weather Group web site:
http://teor.ysn.ru/rswi/

Kozlov, V.I., Borisov, D.Z., Tugolukov, N.N., The diagnostics method of interplanetary disturbances on cosmic ray fluctuation changes and its realization in the automation system of scientific investigation at the Tixie station, Izv. AN SSSR. Ser.fiz., 48, 2228–2230, 1984.

Kozlov, V.I., Starodubtsev, S.A., Grigoryev, V.G., et al. The early diagnosis of the solar wind large-scale disturbances by registration of GCR scintillation in the regime of real time, Izv. RAN. Ser. Fiz., 65, 385–388, 2001.

Kozlov, V.I., Starodubtsev, S.A., Turpanov, A.A., et al. Short-Term Forecast of Space Weather in Real-Time, This conference proceedings, 2001a.

Kozlov, V.I., Starodubtsev, S.A., Turpanov, A.A., et al. Long-Term Forecast of Space Weather in Real-Time, This conference proceedings, 2001b.