Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2018-24
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 9 July 2018
© Author(s) 2018. CC BY 4.0 License.

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access
EGU

Near real time environmental monitoring and large volume data collection over slow communication links

Misha B. Krassovski[1], Glen E. Lyon[2], Jeffery S. Riggs[3], Paul J. Hanson[1]

[1]Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6290, USA
[2]Campbell Scientific, Inc., 815 W 1800 N, Logan, UT 84321-1784, USA
[3]Integrated Operations Support Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6290, USA

Correspondence to: Misha B. Krassovski (krassovskimb@ornl.gov)

Abstract. Climate change studies are one of the most important aspects of modern science and related experiments are getting bigger and more complex. One such experiment is the Spruce and Peatland Responses Under Climatic and Environmental Change experiment (SPRUCE, http://mnspruce.ornl.gov) conducted in in northern Minnesota. The SPRUCE experimental mission is to assess ecosystem-level biological responses of vulnerable, high carbon terrestrial ecosystems to a range of climate warming manipulations and an elevated CO2 atmosphere. This manipulation experiment generates a lot of observational data and requires a reliable onsite data collection system, dependable methods to transfer data to a robust scientific facility, and real-time monitoring capabilities. This publication shares our experience of establishing near real time data collection and monitoring system via a satellite link using not very well-known possibilities of PakBus protocol.

1 Data acquisition

1.1 SPRUCE experimental site

The SPRUCE experimental field site is an 8.1 ha Sphagnum–Picea mariana (black spruce) bog within the US Forest Service (USFS), Northern Research Station's, Marcell Experimental Forest (MEF; 47o 30.17' N, 93o 28.970' W). Access to experimental plots is provided by three, 2.5 m wide, aluminum-framed, composite-decked boardwalks above the bog surface. Boardwalks are supported by helical piles anchored into mineral soil to a depth between 12 to 18 m. The boardwalks provide support for electrical power, propane, and CO2 distribution lines, and data transmission infrastructure. In total, 17 plots were initially established (Fig. 1). Each plot contains a 10m high instrumentation tower with environmental sensors at 0.5, 1, 2 and 4m above the ground (nominal heights of typical shrub and tree foliage and branches). Ten plots are isolated from the aboveground environment by 8m high sidewalls and can be supplied with warmed air at ground level and subsequently mixed throughout the vertical space of the enclosure. Of the 10 enclosed plots, 2 are operated without heating (control plots), and the remaining 8 plots are equipped with propane-fired heating and air handling units to deliver duplicate elevated temperature treatments at 2.25, 4.5, 6.75, and 9.0o C above recorded readings in control plots. Five of the 10 plots, including each of the temperature treatments, are equipped with CO2 injection equipment to deliver an elevated CO2 (eCO2)

Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2018-24
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 9 July 2018
© Author(s) 2018. CC BY 4.0 License.

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access
EGU

treatment (500 ppm above ambient). The seven remaining non-enclosure plots are also instrumented and monitored to provide information on undisturbed, background ambient conditions. More details about warming manipulation studies of ecosystems can be found in Hanson et al., 2017.

1.2 Onsite LAN

The SPRUCE onsite data acquisition system is a local area network that is wired using fiber optic and Cat6 Ethernet cables. This system was designed to handle high volume sources of in situ observational data and function in a harsh, remote environment. Air temperature in northern Minnesota can be up to +40º C at summer and drop down to -40º C in winter. Although all network and data acquisition equipment are located inside enclosures that provide protection from the weather elements, they still may suffer extreme temperature and humidity variations. All networking equipment deployed outside the control building is industrial grade and has extended operating temperature and humidity ranges. Ethernet switches are rated for an operating environment of -40 to +70o C and 10–90% relative humidity. Fiber/Ethernet switches are rated for -20 to 70o C and 5–95% relative humidity.

1.3 Data collection

Data from an instrument or a sensor is recorded by dataloggers located in the data acquisition panels that are assigned to each plot. Logger data then is copied via the LAN to the data storage server located in the control building. All dataloggers deployed are Campbell Scientific, Inc. (CSI) CR1000 models commonly used for environmental science experiments and long-term meteorological observation stations. The main software component is the LoggerNet product produced by CSI. It supports programming, communication, and data transmission between dataloggers and PCs. Use of this package is especially convenient in applications that require telecommunications and scheduled data retrieval within large datalogger networks. LoggerNet runs on a computer in the control building and queries each datalogger every 30 min. Collected data is copied to an onsite file server and then through a satellite link transferred to ORNL servers. The Fig. 2 diagram shows all components together and gives a full overview of the SPRUCE data acquisition system. A more detailed description of the entire data acquisition system including measurements, instruments, sensors, software, and hardware deployed can be found in Krassovski et al., 2015. This publication will look in detail on how the data is transferred from the experimental site to ORNL.

2 WAN data transfer

2.1 Satellite connection

As mentioned above, the experimental site was built in a remote location that had no utility lines in the surrounding area. About 5 km of power cable was buried to supply the site with electricity. Because of low regional population density, the region has no communication lines or reliable cellular coverage. The closest wired Internet access point is more than 16 km

© Author(s) 2018. CC BY 4.0 License.

away. Estimates showed that the expense to extend wired Internet access to the site was initially too high, but may be provided in the future. This reality left only one possibility for Internet connection – a satellite link. Modern technologies can provide fast speed satellite connections up to 1 Gb/s, but they are very expensive and their use as a permanent connection for data transfer was cost prohibitive for SPRUCE. Consumer oriented satellite Internet access, however, was affordable and available in low dense population areas. The drawback of consumer grade connections is the upload speed. Most consumer users browse the Internet, download movies and music and do not upload a lot of data. The fastest upload speed that was found on the affordable market was 1Mb/s. Another requirement was to find a service that provided static IP address. Although there is a staff member on site to monitor and manage the experiment, an ability to have a remote access to the local area network is very important.

2.1 FTP transfer

FTP is one of the most popular ways to transfer data. It is easy to implement, quite reliable and easy to automate. A limitation is that it deals with files as a whole. Experimental output data files are managed in 1-year increments (i.e. they contain data from January 1st to December 31st). In the beginning of a year they are small, but as the year progresses, files sizes grow, and transfer time gets longer. Another limitation is that satellite connections are sensitive to weather conditions. Heavy rain, snow or very dense clouds can interrupt link and data transfer. Every interruption forces the ftp to reinitiate interrupted transfers. Copying only the changed blocks could save time and bandwidth. Many commercial and free software implement partial transfers, but only for block-oriented file types like database files, drive images, virtual disk images, etc. Stream-based files, on the other hand, will usually cause all blocks to be changed whenever they are modified (for example text documents, spread sheets, zip files, and photos). Dataloggers generate comma separated value (CSV) text files, which can't be partially transferred without special and software that runs on both sides of the communication line.

2.2 PakBus protocol

As stated before, the main control and monitoring tool at the experimental site is the LoggerNet software product distributed by CSI. The LoggerNet software product offers the tools to remotely accomplish many datalogger tasks, including writing datalogger programs, transmitting programs to dataloggers, collecting data, and analyzing data either in real time or after the data has been saved to a computer. It consists of a communications server application, several client applications and supports telecommunications and scheduled data retrieval used in large datalogger networks. LoggerNet supports the TCP/IP protocol and can transmit the CSI PakBus protocol over it (LoggerNet manual, 2015). PakBus is a switched network protocol developed by CSI that works with PakBus enabled dataloggers and other devices. It is a routing protocol and a datalogger or the LoggerNet software can be configured as a router. PakBus is mostly used when a situation does not require setting up a TCP/IP network, but can be implemented over it as well. The LoggerNet server uses PakBus to communicate with dataloggers and collect data. The data collection is based on an implemented schedule (every 30 min in this particular case, it is a good resolution for recording diurnal patterns.) and works in the following way. Measurements from sensors and

© Author(s) 2018. CC BY 4.0 License.

instruments connected to a datalogger are written to the internal memory of the data logger. The LoggerNet communications server contacts dataloggers, retrieves data from their memory, and writes it to files. The communication server determines what data is new since the last data collection and retrieves only uncollected data, although datalogger memory may have days' worth of data stored. This way the data files are growing incrementally and only the last 30 min of data is transferred over the LAN.

2.3 PakBus over WAN

After exploring the routing capabilities of the PakBus protocol, it was decided to use it over the satellite connection and take advantages of the incremental data transfer provided by the LoggerNet communication server. Each device in a PakBus network has a unique address. Valid addresses are 1 through 4094. Communications from PakBus devices in the PakBus network with addresses greater than 3999 cannot be ignored by other PakBus devices in the PakBus network and must be responded to. LoggerNet default PakBus address is 4094. To simplify orientation, dataloggers at the experiment site have PakBus addresses assigned according to the plots they are installed in and the LoggerNet server is assigned the default address of 4094. Port 6785 is the default port used for PakBus/TCP communications in LoggerNet. Port forwarding for port 6785 is set up at the SPRUCE router, and the LoggerNet server is configured to bridge PakBus ports thus functioning as PakBus router. When bridging PakBus ports in LoggerNet, all dataloggers in the network must have unique PakBus addresses, which they do in this case, or a PakBus address conflict will exist. The Fig. 3 details the LoggerNet data logger network settings at experimental site. Another instance of the LoggerNet server software was installed at ORNL and assigned a PakBus address of 4093 to avoid PakBus address conflict with the LoggerNet software at the experimental site. The configuration depicted in the Fig. 4 is implemented at ORNL. The IP address of the IPPort in Fig. 4 is the WAN IP address of the SPRUCE router. All of these arrangements allow the LoggerNet server at SPRUCE site to act as a PakBus router and allowed the LoggerNet server at ORNL to communicate with devices on remote LAN and retrieve data.

2.3 Near real time data collection and monitoring

As previously mentioned the LoggerNet communication server gathers only data stored in the dataloggers since the last scheduled data collection and transfers only those small data packages over the satellite link. At the present time, it takes about 30 minutes to move all new data from the experimental site to ORNL. For redundancy and as an additional backup, LoggerNet data collections are scheduled on both onsite and ORNL servers and data accumulated in both locations. Although technically unnecessary, to optimize data collection, simultaneous data collection by the two LoggerNet communications servers is avoided by offsetting their data collection schedules by 15 minutes. At SPRUCE site, it happens twice in an hour at XX:00 and XX:30, at ORNL it scheduled at XX:15 and XX:45 It allows monitoring of data with a thirty-minute delay on site and about a one hour delay at ORNL (it varies slightly based on the satellite link speed). The data received are also used as a constant feed to data visualization software (Vista Data View, Vista Engineering), for plotting all site data in variable time series formats. Such visualization package enables remote assessments of equipment performance

© Author(s) 2018. CC BY 4.0 License.

and data quality assurance. Figure 5 shows an example of measured events at experimental plot 19. Ecosystem-scale manipulation experiments are getting more complicated and require innovative approaches that help manage high volumes of in situ observations. New large-scale experiments in remote locations will become common in the future and will require reliable data acquisition systems. The presented approach shows an example of non-standard network technologies to overcome difficulties caused by slow speed communication link. We hope that the provided details about network technologies, hardware, software configuration, and the logic behind these choices will help the future design and development of similar systems for other experiments.

Abbreviations

SPRUCE, Spruce and Peatland Responses Under Climatic and Environmental Change experiment;
MEF, Marcell Experimental Forest;
DOE, US Department of Energy; TES, Terrestrial Ecosystem Science;
ORNL, Oak Ridge National Laboratory;
CSI, Campbell Scientific, Inc.;
LAN, Local Area Network;
WAN, Wide Area Network;
CVS, Comma Separated Value;
FTP, File Transfer Protocol;
TCP/IP, Transmission Control Protocol/Internet Protocol;

Acknowledgement

This research is supported by the US Department of Energy, Office of Science, Biological and Environmental Research (BER), and conducted at Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC under contract no. DE-AC05-00OR22725 with the US Department of Energy. The publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

© Author(s) 2018. CC BY 4.0 License.

References

Hanson, P., Riggs, J., Nettles, W., Phillips, J., Krassovski, M., Hook, L., Richardson, A., Aubrecht, D., Ricciuto, D., Warren, J., Barbier, C. (2017). Attaining whole-ecosystem warming using air and deep soil heating methods with an elevated CO2 atmosphere. Biogeosciences, doi: 10.5194/bg-14-861-2017.

Krassovski, M., Riggs, J., Hook, L., Nettles, W., Hanson, P., Boden, T. (2016). A comprehensive data acquisition and management system for an ecosystem-scale peatland warming and elevated CO2 experiment. Geoscientific Instrumentations, Methods and Data Systems, doi:10.5194/gi-4-203-2015.

LoggerNet ver. 4.3. instruction manual. (2015). Campbell Scientific, Inc., 815 W 1800 N, Logan, UT, 84321-1784

Figure 1: Aerial view of the SPRUCE experimental footprint.

Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2018-24
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 9 July 2018
© Author(s) 2018. CC BY 4.0 License.

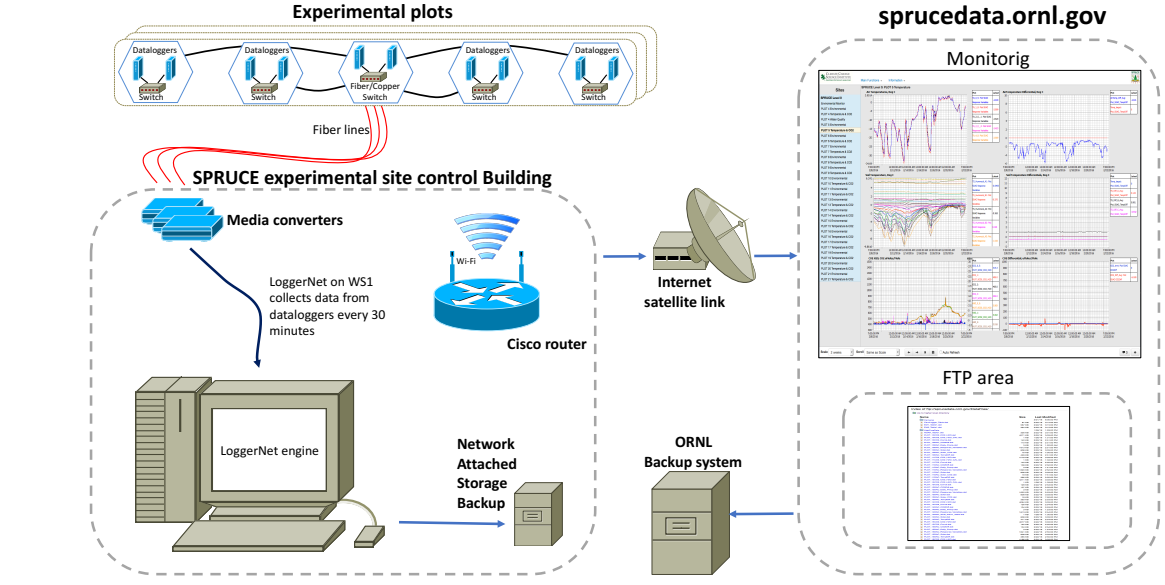

Figure 2: SPRUCE data system.

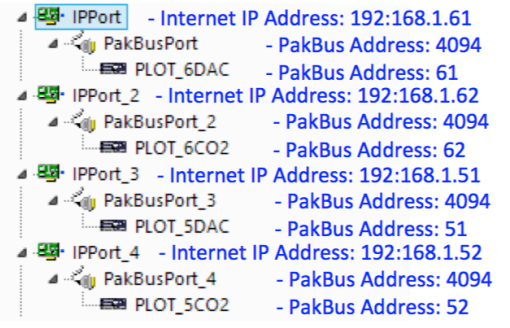

Figure 3: Example PakBus and IP addresses assigned at experimental site.

Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2018-24
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 9 July 2018
© Author(s) 2018. CC BY 4.0 License.

Geoscientific Instrumentation Methods and Data Systems Discussions

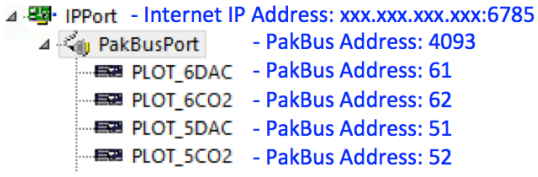

Figure 4: Example PakBus addresses assigned at ORNL.

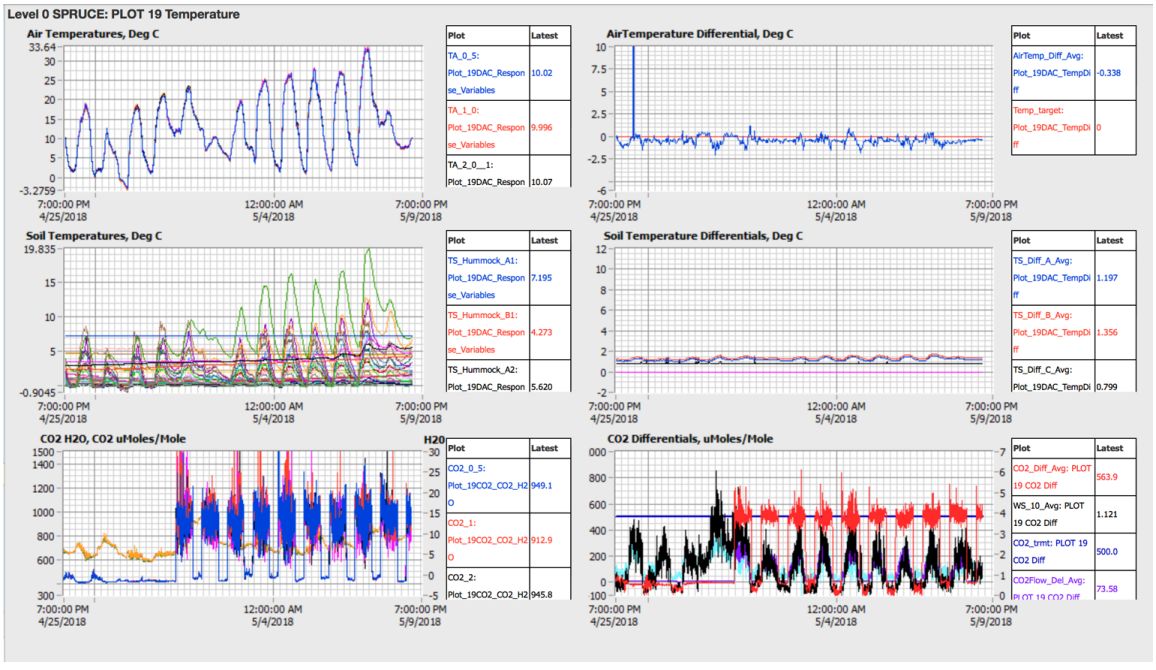

Figure 5: Monitoring processes at plot 19 via VDV software at ORNL.