A Data Collection System with Integrity Verification for Wetland Monitoring Projects

Anton Riedl, Brian Lamprecht and Rob Atkinson
riedl@pcs.cnu.edu, blampy@pcs.cnu.edu, atkinson@cnu.edu
Christopher Newport University, Newport News, VA, USA

Abstract -- Hydrologic monitoring is critical to the restoration and creation of wetlands and, thus, is important when overseeing wetland compensation sites. At present, the collection of hydrologic data requires visits to numerous water level measurement devices installed on the sites. This includes traveling to remote field sites, locating each station, downloading stored data onto a handheld device, and returning to the office or lab for data analysis. As this process is rather time and resource intensive, it is not surprising that infrequent and incomplete data collection is the consequence, leading to the delay of management efforts should hydrologic conditions be inappropriate. Furthermore, the extended cycle times between data measurement and data analysis increase the risk of losing critical measurements due to undetected equipment failures. Another weakness of the current methodology is that there is no chain of custody for data measurements in order to provide auditable validation of the collected data. Once the data have been read out from the measurement device, it is impossible to tell whether they are accurate or whether they have been modified somewhere along the way – either due to equipment malfunction, human error, or malicious intent.

To solve for these problems we are implementing a remote hydrologic monitoring information system capable of reporting measurements from the field in near real time. The benefits of this system include an overall cost reduction in performing assessments, a reliable chain of custody on hydrologic measurements, and risk reduction due to the opportunity for timely adaptive management strategies at wetland compensation sites. In this paper we discuss both the processes and the mechanisms involved in the implementation of our solution focusing on aspects concerning data integrity and data stewardship.

I. INTRODUCTION

In determining the presence or absence of wetland conditions, three key features must be established as present for a given area: predominance of hydrophytic vegetation, anaerobic soil conditions, and the hydrology of the terrain [1]. While all three of these are required for a terrain to be considered a wetland, hydrology is the most important because the soil conditions and the presence of hydrophytic vegetation depend on the regular presence of water. Understanding hydrology has been shown to be important to better understand wetland restoration and creation efforts, to detect regional impacts of climate change, and to understand wetland morphology [2][3][4].

The traditional means for monitoring site hydrology requires digging shallow wells and installing monitoring devices that measure the water level and store data locally. This method requires an individual to physically visit a site in order to download the data. In recent years, wireless technology and, in specific, wireless sensor networking are increasingly being developed and deployed for environmental monitoring purposes, focusing mainly on the automation of the data collection process [5][6][7][8]. The new technologies make it possible to raise the frequency with which ecological data can be collected and, thus, allow more detailed research into natural processes. Even though there are still issues to be solved before such systems can be deployed by non-experts in a straight-forward manner (e.g., issues related to data storage/loss, network resilience, or battery lifetime), it is expected that wireless sensor networks will be widely used for environmental monitoring.

In this paper, we present a system for hydrologic monitoring, which is based on wireless sensor networking. We propose an integrity framework for reporting data in a way that removes individuals from the chain of custody and allows the verification of the data's integrity. However, it must be noted that integrity checking can only be applied to data after they have entered the system. It cannot protect against deliberate or accidental misuse of the monitoring equipment, such as wrong installation or incorrect calibration (both leading to inaccurate data).

The rest of the paper is organized as follows. In section II we discuss the monitoring procedure and associated policies. Furthermore, the individual players in the process are presented and their relationship is discussed. Section III illustrates the overall architecture of our system and outlines our experimental implementation. Section IV is the main part and focuses on data integrity and its verification. Section V concludes the paper.

II. MONITORING PROCEDURE AND POLICIES

Current practice of measuring wetland hydrology involves the deployment of hydrologic measuring devices purchased from various manufacturers. Installed across a wetland site they take daily hydrology readings and store the information locally. The devices are deployed by environmental consultants that are typically hired by land owners who have a need to create wetland areas, e.g., as part of a mitigation agreement. The environmental consultants return to the site throughout the growing season to capture the measurements taken by the unattended devices and to process information. Once the objectives of the mitigation project have been achieved, requirements may necessitate the environmental consultants to file the results with a regulatory agency such as the US Army Corps of Engineers. Figure 1 illustrates this process and identifies the relationship between the involved entities.

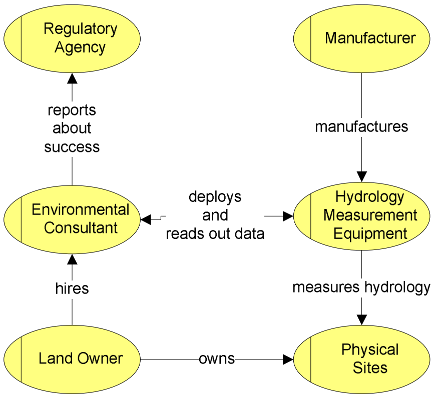

Fig. 1. Current Monitoring Practices

The process of data collection is very time and resource intensive, which is the reason why it is done rather infrequently, despite the fact that extended cycle times between data measurement and data analysis increase the risk of losing critical measurements due to undetected equipment failures. Another weakness of the current practice is that there is no chain of custody for data measurements in order to provide auditable validation of the collected data. Once the data have been read out from the measurement device, it is impossible to tell whether they are accurate or whether they have been modified somewhere along the way – either due to equipment malfunctions, human errors, or malicious intent.

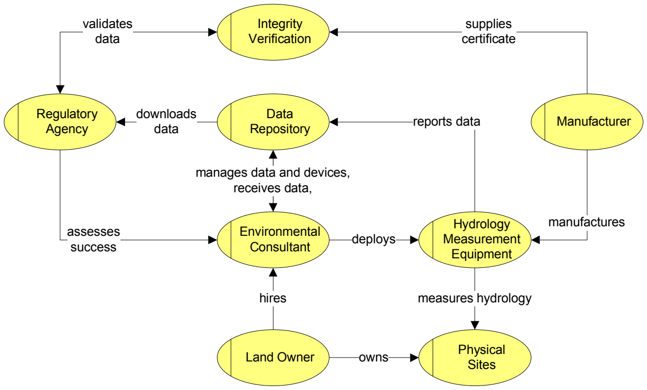

Fig. 2. Proposed Monitoring System

To solve for these problems, we propose a system as depicted in Figure 2. Instead of having to download the data by hand, the monitoring equipment reports them via sensor networking to a data repository, which is run by a manufacturer or by a third party. In either case, the consultants receive the data through the repository and are able to release them to regulatory agencies. In addition to data distribution, the overall system includes a service, which is employed for verifying the data's integrity.

III. System Overview

Figure 3 illustrates the architecture of the data collection and integrity verification system. Its two main parts are the wireless sensor network on the wetland site and the backend service for data storage, distribution and integrity verification.

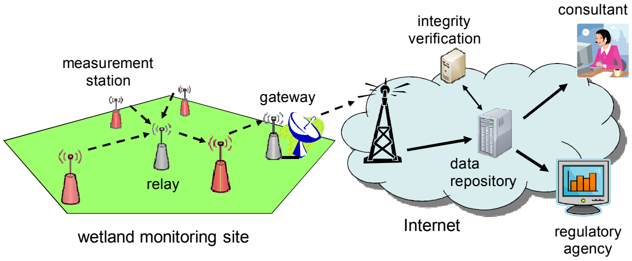

Fig. 3. Architecture Overview

In our system, each measurement station is to be equipped with a low-power processing unit and a radio module, which together provide the necessary functionalities to allow data integrity verification and wireless data collection. In addition to storing the measurements locally, as it is the case with today's devices, data are processed and sent on to a gateway device. Due to the characteristics of the terrain with its seasonally changing, densely leaved flora, the radio coverage is limited. It is therefore necessary to install relay stations in the field whose sole purpose is to guarantee connectivity between stations and the gateway. It should be noted that the location of the measurement stations is determined by environmental factors inherent to the wetland monitoring process, while the location of the gateway is predominantly a question of accessibility to power supply, if at all available, and to Internet connectivity. Various options exist to establish the link from the gateway to the Internet, the most prominent ones currently being wireless LAN, cellular, and satellite technology. Among these three, wireless LAN is the one least likely available at typical wetland locations.

As outlined above, the data are uploaded to a repository server, from where they can be accessed by the consultant and also by auditors if permission is granted. We envision the data repository to be run by an independent entity, which specializes in data storage and presentation. As consultants are not necessarily well-versed in networking technologies, data retrieval and access management have to be made as simple as possible. For data integrity assurance it has to be possible to keep track of data sets, even after they have been modified by the consultants. New presentations of data have to be linked to their original values in order to allow the verification of their integrity or to detect alterations. In the simplest case, a regulatory agency will be granted access to the original data sets and then has to trace through any changes made to the data in order to verify the integrity of the final, presented values. A more sophisticated data modification tracking methodology would embed meta-information into the data sets themselves and, thus, enable the verification of their integrity at each step along the way. Integrity-related functionalities can either be integrated into the data repository service or are offered as an independent service.

An experimental system is currently being implemented using wetland monitoring equipment from Remote Data Systems, Inc. and sensor networking devices from Crossbow. Each water level monitor is paired with a mote (either Micaz [9] or IRIS [10]), which provides data processing and radio functionality. Micaz and IRIS motes both include a low-power 8-bit microcontroller from Atmel (ATmega 128L and ATmega 1281, respectively) and an IEEE 802.15.4 compliant RF transceiver from Texas Instrument (CC2240) with a transmission rate of 250 kbps. The operating system running on the motes is TinyOS [11]. The motes interface with the water level meters via an RS-232 serial line. The connection between the sensor network and the fixed network will be established via a Crossbow Stargate Netbridge [12]. On the backend side, data storage and integrity verification are implemented as simple web services.

IV. Data Integrity Verification System

The underlying principle of the data verification process is a digital signature, which is associated with the measured data before they are stored in the data repository. This signature makes it possible to determine whether data that are later being presented to the oversight agency are indeed based on the original values or whether they have been tampered with. Conceptually, the data integrity system consists of two components: a signature generation component, which signs the original data, and a signature verification component, which checks the integrity of the data based on the provided signature later on. Both component types can exist in single or multiple entities.

In the following sections we will discuss implementation issues according to the three aspects:

- Location of signing: measurement site vs. gateway device vs. backend server.
- Type of cryptographic algorithm: public-key vs. symmetric-key.
- Data aggregation level: signature for individual values vs. aggregated data sets.

A. Location of Signing

Various possibilities exist of where the signature can be applied, each leading to a different level of strength of the integrity verification system:

- at the location of measurement, i.e., inside the monitoring devices or the attached motes,
- at the link between the wireless network and the fixed network, i.e., in the gateway device, or
- in the backend of the system, e.g., in the data repository server before storing the data, or in a specialized verification server.

Signing the data in the end system, directly at the location of measurement, provides the strongest solution regarding integrity verification. In this case, a malicious attack, which aims at the modification of the data without the ability to detect these modifications, would require tampering with the end devices themselves. One would have to intercept the data between the measurement module and the processing unit in order to modify them before signing. Negative aspects of this solution are limitations in processing power and energy consumption in the motes.

The second option, i.e., signing the data in the gateway device between the wireless and the fixed network has the advantage that the device is typically more powerful and power supply might be available. At the same time, the location is still close to the point of measurement and, thus, the confidence in the verification process is still high.

The final option, applying the signature on the server side inside the fixed network, has the advantage that computation power and energy consumption are practically not an issue. However, the shortcoming is that the data have already traversed a range of devices, opening up opportunities for tampering.

For our system we have chosen the end device as the location of signature generation. In order to prevent replay attacks, the data are combined with a station ID and a unique time stamp before being signed.

B. Cryptographic Algorithm

There are generally two methodologies to verify the integrity of data: symmetric cryptography and public-key cryptography. In both cases, the data are run through a hash function in order to produce a fingerprint, which is then used as input for the signature algorithm.

In case of symmetric encryption, a secret key is shared between the sender and the receiver. After the sender encrypts the fingerprint of the data, the receiver can verify its integrity by decrypting the ciphertext and comparing the resulting value with a newly generated hash value. It should be noted that the devices applying the signature on the one side (i.e., the motes, the gateway or the server) and the entity that verifies the signature on the other side (the verification server) have to be under the same administration in order to assure the privacy of the key. Any other entity that wants to check the integrity of data sets has to do so by submitting the data together with the signature to the dedicated verification server. This solution is conceivable; however, due to the need of the close relationship between the devices that perform the signature and the ones that verify it, it is rather impractical. As a consequence, we base our solution on the public-key cryptography scheme.

With public-key cryptography the data are signed using a private key, which is not known to anyone but the signer itself, in our case the mote. For our system to work, we assume that each monitoring device comes pre-configured with at least one private key, for which the manufacturer provides a corresponding public key (e.g., in form of a certificate). This public key is used to verify the signature. The advantage of this approach is that anyone can perform the verification process and it is not a necessity to have a dedicated verification server. Nevertheless, we propose to have such a server, as it makes data handling easier and allows, in interplay with the data repository, a solution that is transparent for the users.

C. *Data Aggregation*

When signing the measurements, the data aggregation level has to be chosen carefully. On the one hand, the objective of keeping the delay between the time of measurement and the time of data delivery to the consultant as small as possible demands a low aggregation level. At best, each data point is signed individually so it can be sent on to the data repository immediately. On the other hand, this fine granularity adds a significant overhead respective to the amount of data that need to be handled and transmitted as well as to the computation time necessary for signature generation. Both represent a considerable strain on the power consumption, especially if the signing is performed within the end devices at the wetland site. Therefore, it is advisable to combine measurement points into data sets and apply the signature to these aggregates. The aggregation level in this case is determined by the maximum reporting delay that one is willing to accept.

To illustrate this tradeoff, assume that a single measurement sample is encoded as an m-byte value and that a samples are aggregated before being signed. In addition to the set of samples, the final message, which needs to be transmitted, contains a time stamp of length t, a station ID of length id, as well as the signature of length s. Table 1 lists typical values for the given parameters. It is assumed that the water depth samples are simply represented in ASCII notation, thus, being relatively long. The signature algorithm is assumed to be based on Elliptic Curve Cryptography (ECDSA), which achieves a sufficient security level using key sizes as small as 160 bits (which result in a signature size of 40 bytes) [13].

m	sample size	10 bytes
a	aggregation level	>=1 data samples
t	time stamp	4 bytes
id	station ID	4 bytes
s	signature length	40 bytes

Table 1: Data Overhead Example

Figure 4 shows the overhead added to the payload for integrity check purposes as a percentage of the actual payload size. By aggregating just a few samples, the overhead can be significantly reduced. Looking at Table 1, one should note that the overhead is mainly determined by the size of the signature.

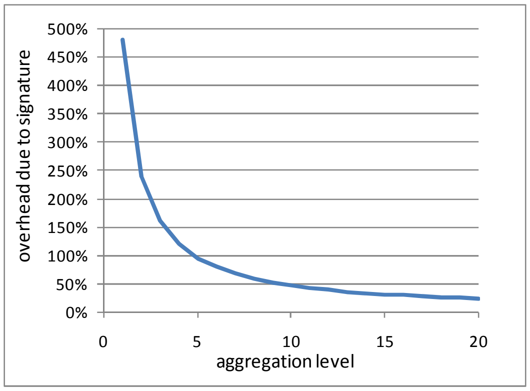

Fig. 4. Data overhead over aggregation level

With respect to power consumption and battery lifetime, the process of signature generation is crucial. Its computation time exceeds the activity period of any other individual monitoring-related task, such as downloading a measurement sample from the water level monitor, putting it into local storage or transmitting it to a neighboring station. According to [14] and [15], which both provide an ECDSA implementation for TinyOS, signature generation times are in the order of seconds. The minimum values given in the papers for Micaz motes are 2.00 s and 1.35 s, respectively. Considering that motes need to operate at very low duty cycles (< 1%) in order to achieve satisfactory battery lifetimes, signature generation can become a significant load if done too frequently. For example, if signature generation takes 2 seconds and it is carried out every 10 minutes, the duty cycle increase is over 0.3%. For a typical sensor network configuration (Micaz with 2000 mAh battery, radio in sleep mode for 99% of the time), this increase in duty cycle would result in a drop of the battery lifetime from about 12 months to 11 months. Therefore, when designing the system one has to choose the interval between signature generations large enough as not to cause unacceptably low battery lifetimes. The exact values, however, depend on the type of motes and the details of the monitoring objectives.

D. *Verification Service*

Once collected and signed, the data are sent to a repository server for storage and distribution. For each sample set, the context including the data values, the station ID and the timestamp is stored together with the digital signature. We envision the repository being operated by an independent agency such as a non-profit organization or a federal/state agency such as the US Army Corps of Engineers.

For integrity verification, the data are retrieved from the repository and their signature is checked against a newly generated hash. In case of public key cryptography, this can be done directly by the retrieving entity, e.g., by determining the public key from a certificate and then applying the appropriate signature verification algorithm. Another option, one that we pursue in our system, is to submit the data to a verification service, which is implemented as a simple web service. This service retrieves the appropriate certificate, extracts the public key and then checks the signature for its correctness.

While doing this, it has to be guaranteed that privacy of data is upheld and that services are only offered to authorized users. Due to the nature of the process where privileges might change over time, access control has to be flexible and easily configurable by the consultants. While typically the ownership of the data stays with the consultants, access is granted to various entities such as oversight agencies or collaborators.

In this context, interesting questions arise when data integrity should be verifiably preserved even after the consultant modifies the values received from the field. This is necessary for analysis and presentation purposes. So far, our current concept only considers integrity verification for originally submitted data. For auditing purposes it is therefore necessary to store the original data, and to be able to justify the individual modification steps that were performed thereafter. For later stages it is conceivable that the data are re-signed once they are taken out of their initial context and modified.

V. Conclusion

Data collection for wetland monitoring projects is currently still a tedious task, requiring regular visits to the sites in order to manually download measurement data from the monitoring stations. To facilitate this process, we are developing a system that employs wireless sensor networking technology to automatically collect the measurements from the site and to store them in a repository. A particular feature of our system is its strong focus on data integrity and verification. Current practices do not provide ways for regulatory agencies to determine whether submitted data have been tampered with. To remedy this shortcoming, we attach a digital signature to the data sets right in the monitoring devices and provide a web service for users to verify the data's integrity at a later point of time. While this does not fully prevent the submission of phony data (as, for example, faulty use of monitoring equipment cannot be detected this way), it makes it impossible to polish data at any time between the point of measurement and the point of submission.

References

[1] *Corps of Engineers - Wetlands Delineation Manual*, Environmental Laboratory, Technical Report Y-87-1, Jan. 1987.

[2] Hunt, R. J., J. F. Walker, and D. P. Krabbenhoft. "Characterizing Hydrology and the Importance of Ground-Water Discharge in Natural and Constructed Wetlands." Wetlands 19.2 (1999): 458-72.

[3] Conly, F. M., and G. Van Der Kamp. "Monitoring the Hydrology of Canadian Prairie Wetlands to Detect the Effects of Climate Change and Land use Changes." Environmental monitoring and assessment 67.1-2 (2001): 195-215.

[4] Shaffer, P. W., M. E. Kentula, and S. E. Gwin. "Characterization of Wetland Hydrology using Hydrogeomorphic Classification." Wetlands 19.3 (1999): 490-504.

[5] Cardell-Oliver, Rachel, et al. "A Reactive Soil Moisture Sensor Network: Design and Field Evaluation." International Journal of Distributed Sensor Networks 1.2 (2005): 149-62.

[6] Coletti, LJ, SE Fitzwater, and KS Johnson. Wireless Network Allows Monitoring of a Dynamic Coastal Resource. Vol. 87. 2000 Florida Ave., N.W. Washington DC 20009 USA, [URL:http://www.agu.org/pubs/agu_joureos.html]; Transactions, American Geophysical Union: American Geophysical Union, 2006.

[7] Bogena, H. R., et al. "Evaluation of a Low-Cost Soil Water Content Sensor for Wireless Network Applications." Journal of Hydrology (Amsterdam) 344.1-2 (2007): 32-42.

[8] Selavo, L., et al. LUSTER: Wireless Sensor Network for Environmental Research. Sydney, Australia: ACM, 2007.

[9] MICAz – Wireless Measurement System, http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICAz_Datasheet.pdf

[10] IRIS – Wireless Measurement System, http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/IRIS_Datasheet.pdf

[11] TinyOS – Open Source Operating System for Wireless Embedded Sensor Networks, http://www.tinyos.net

[12] Stargate NetBridge – Embedded Sensor Network Gateway, http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/Stargate_NetBridge_Datasheet.pdf

[13] Certicom Research. Standards for efficient cryptography – SEC 1: Elliptic curve cryptography. http://www.secg.org/download/aid-385/sec1_final.pdf, September 2000.

[14] An Liu, Peng Ning, "TinyECC: A Configurable Library for Elliptic Curve Cryptography in Wireless Sensor Networks," in Proceedings of the 7th International Conference on Information Processing in Sensor Networks (IPSN 2008), SPOTS Track, April 2008, pp 245-256.

[15] H. Wang and Q. Li. "Efficient Implementation of Public Key Cryptosystems on Mote Sensors (Short Paper)," in Proceedings of International Conference on Information and Communication Security (ICICS), Dec. 2006, pp. 519-528.