

An Anti-Entropy Protocol Suitable For Managing Data Deletion in an Epidemic Data Transmission System

Eric A. Freudenthal, PhD, Virgilio Gonzalez*, PhD, Brian A. Carter
Computer Science Department, University of Texas at El Paso
*Electrical and Computer Engineering Department
500 W. University, El Paso, Texas, 79968

ABSTRACT

Gossip-based epidemic transport protocols promiscuously replicate all information among participants to disseminate information in contexts where ubiquitous connectivity is not available. Potential applications include transmission of measurements and observations from personnel- or asset-tracking systems and health monitoring devices to infrastructure-connected upload stations. This paper describes an anti-entropy protocol suitable for efficiently deleting uploaded data.

We report on a nascent effort to investigate gossip based data collection protocols that disseminate measured data from sensors to infrastructure-connected upload stations in environments where ubiquitous connectivity is not available. Distinct measurement samples generally must be preserved until a copy of them is uploaded and samples that have already been uploaded can be safely deleted from devices participating in the transmission protocol. Our anti-entropy technique exploits the sequential nature of data collection and upload events; later upload events associated with a particular data sequence supersede all previous upload events associated with that sequence. Thus, a device that is aware of multiple upload events associated with a single sequence store and epidemically propagate only the most recently issued upload receipt.

It is possible to exploit the disconnected but controlled nature of health-care settings, such as field hospitals, to epidemically transmit data among small radio sensor devices carried by personnel. In addition to the collection of data from monitoring devices, this protocol can be used to track the potential spread of contagious disease by recording the sequences of interpersonal rendezvous. In such a scenario, all medical facility personnel wear small short-range radios that record the timing and sequence of nearby encounters with other devices. When this information is disseminated to a infrastructure, it can be used to determine potential infection vectors during an outbreak. If these histories are communicated epidemically, each radio potentially serves both as a sensor and as a “data mule” that carries others’ data.

A variety of low-cost communications standards are appropriate for this type of protocol including Bluetooth, UWB, Wi-Fi, and RFID/NFC. RFID/NFC is particularly interesting because it permits communication both among active (powered) devices and between an active and a passive (field-powered) device. For example, a system where personnel carry active devices could be made location aware by recording communication histories with passive devices attached to (and thus associated with) particular physical locations such as doorways.

Epidemic transmission of anti-entropy messages indicating sequence timestamps for already uploaded data permits devices to purge their memories. This frees resources to collect and transmit additional data. Anti-entropy messages associated with a particular data stream are monotonically ordered using a Lamport-style timestamp, and thus a later message indicates the expiration of all previous data and messages associated with that stream. Previous approaches using RFID do not have these capabilities [5,6].

A prototype system without anti-entropy mechanisms to record inter-personnel contact history already has been implemented. Initial results indicate that epidemic communication is suitable for collecting sequences of interpersonal rendezvous. We have not explored scalability challenges and, more importantly, have not included mechanisms to propagate upload events; instead, each datum is explicitly marked with an expiration time. Each datum is stored and epidemically communicated by sensors and data mules until expiration. We observe that this preset expiration time may occur prior to or long after upload. Either condition can be problematic: too-short

lifetimes result in data loss, and too-long lifetimes result in unnecessary storage and communication. This waste limits the scale of deployment. In contrast, our anti-entropy protocol enables participating devices to identify and delete copies of samples that have already been uploaded.

We will also investigate dynamic characteristics of realistic environments such as hospitals, nursing homes, and private residences in order to determine operational parameters. We are commencing study of the proposed protocol's operating characteristics both under simulation and through construction of prototype implementations. We will investigate the strategic positioning of upload stations to increase scalability. We also will investigate potential alternative operating modes that can be activated when large numbers of devices are close together for a sustained period of time such as "flash crowds" that can occur near elevators or on stairwells.

REFERENCES

- [1] A. J. Demers, K. Petersen, M. J. Spreitzer, D. B. Terry, M. M. Theimer, and B. B. Welch. "The Bayou Architecture: Support for Data Sharing among Mobile Users," Proceedings of the Workshop on Mobile Computing Systems and Applications, Santa Cruz, California, December 1994, pages 2-7.
- [2] M. M. Theimer, A. J. Demers, K. Petersen, M. J. Spreitzer, D. B. Terry, and B. B. Welch. "Dealing with Tentative Data Values in Disconnected Work Groups," Proceedings of the Workshop on Mobile Computing Systems and Applications, Santa Cruz, California, December 1994, pages 192-195.
- [3] J. J. Kistler, M. Satyanarayanan. "Disconnected Operation in the Coda File System," ACM Transactions on Computer Systems, Carnegie Mellon University, February 1992, pages 3-25.
- [4] D. B. Terry, K. Petersen, M. J. Spreitzer, M. M. Theimer. "The Case for Non-transparent Replication: Examples from Bayou," IEEE Data Engineering, December 1998, pages 12-20.
- [5] RFID Journal: Taiwan uses RFID to combat SARS. August 1, 2003. Available at <http://www.rfidjournal.com/article/articleview/520/1/1/>. Accessed July 28, 2004.
- [6] RFID Journal: Singapore Fight SARS with RFID. June 4, 2003. Available at <http://www.rfidjournal.com/article/articleview/446/1/1/>. Accessed June 18, 2004.