Challenges and Opportunities in Deeply Embedded Systems Security

Madhukar Anand and Insup Lee
Department of Computer and Information Science,
University of Pennsylvania,
Philadelphia, PA, 19104, USA
{anandm,lee}@cis.upenn.edu

Abstract

Deeply embedded systems present a number of new challenges and opportunities in security. In this essay, we introduce some of them and explore potential ideas for addressing them.

Introduction The advent of low-powered wireless networks of embedded devices has spurred the development of new applications at the interface between the real world and its digital manifestation. A distributed computing platform that can measure properties of the real world, formulate intelligent inferences, and instrument responses, requires a new class of techniques in distributed computing, real-time systems, artificial intelligence, databases, control theory, and security.

Before these intelligent systems can be deployed in critical infrastructures such as emergency rooms and power plants, the security properties of such systems must be fully understood. Existing wisdom has been to apply the traditional security models and techniques to these networks of deeply embedded systems: as in conventional computing environments, the goal has been to protect physical entities: devices, packets, links, and ultimately networks. However, deeply embedded networks are not traditional computing devices, and as a result, existing security models and methods are insufficient. These networks have unique characteristics that warrant novel security considerations: the geographic distribution of the devices allows an attacker to physically capture nodes and learn secret key material, or to intercept or inject messages; the hierarchical nature of many of these networks and their route maintenance protocols permit the attacker to determine where the root node is placed. Perhaps most importantly, most deeply embedded networks rely on redundancy (followed by aggregation) to accurately capture environmental information even with poorly calibrated and unreliable devices. This results in a fundamental distinction between a physical message in the network and a logical unit of information: a message with a single reading may reveal very little information about the real environment, whereas a message containing an aggregate or collection of readings may reveal a great deal more. These characteristics open the door for an entirely new security paradigm: one that acknowledges that there is a fundamental distinction between physical messages and logical information, and that focuses on how to minimize the correlation between the two in order to limit opportunities for compromise. In the rest of this essay, we highlight some of the challenges in producing a comprehensive security model that is tailored for these low-powered distributed devices and some new approaches that hold promise in solving some of these problems.

Measuring confidentiality Existing literature has proposed the use of computationally inexpensive cryptographic techniques to handle message confidentiality and authenticity in deeply embedded networks. The difficulty of ensuring confidentiality and authenticity is not, however, due solely to the energy constraints imposed on devices. A network that is comprised of many small computing devices, is subject to physical capture. Any cryptosystem must therefore tolerate the compromise of devices and their keys. However, the compromise of some nodes need not result in a total loss of security. Rather than providing all-or-nothing guarantees about privacy or security, there is a need for providing probabilistic guarantees with respect to compromise. The primary challenge is therefore to define models and metrics along these lines, for different protocols logical-level information privacy and security properties.

The approach we advocate takes advantage of the fact that any real-world attacker is limited by the properties of the system he or she is attempting to compromise (c.f., [2]). Based on this model, in an earlier work [1], we have presented an initial framework, taxonomy, and methodology for quantifying the privacy and security of deeply embedded applications, under the assumption that some nodes may be compromised, and based on the networks size, protocols, and computations. Rather than providing all-or-nothing guarantees about privacy or security, the goal is to examine probabilistic guarantees with respect to compromise, and to

understand and improve existing aggregation strategies with respect to these guarantees.

As an instance of the above approach, we can define the eavesdropping vulnerability based on several important parameters for a network of sensors. First, there is the probability that a compromised set of devices, S_A, greatly resembles the set of nodes that our application is sam- pling, S_C . This probability is a function of the size of S_C , the specific aggregate function σ, and the data distribution of the sensors S. For example, if all sensors produce the same reading, then the adversary can compromise the system from a single reading. We formalize the probability based on these parameters.

Definition 0.1 *(Eavesdropping Vulnerability) The eavesdropping vulnerability (γ) relative to a set of compromised nodes is de- fined as $\gamma(\sigma, S, S_A, S_C, \delta) = p(|\sigma(S_C) - \sigma(S_A)| \leq \delta)$, where σ is the aggregating function and δ the adversarys error tolerance.*

Although we have considered a single aggregate computation here, the eavesdropping vulnerability can be generalized to sup- port multiple aggregate computations over different attributes. We finally note that, if the underlying distribution of sensor values and the link vulnerability are known, then we can compute the expected eavesdropping vulnerability. Once the vulnerability is quantified, then we can design networks which minimize it.

Context and topology obfuscation Many networked embedded applications involve data aggregation from a variety of sensors before actuation of some control action. In such a scenario, data is pulled towards the decision making unit, through the cooperation of a few intermediate nodes. For the sensor values to have meaning then, context is needed. Where the value was recorded, and at what time, are necessary for interpretation. Conversely, if the time and location of one reading are known, it may be possible for an adversary to infer a great deal about other readings nearby. Aggregation of data by the intermediate nodes also leads to a non-uniform distribution of information across nodes. Therefore, attacking a leaf node in a tree-structured network gains little influence (for disruption) or information (for eavesdropping); attacking a node near the root gains significant influence and information about the aggregate value. For eavesdropping, there is an interesting third case of attacking nodes in the middle of the tree: intermediary nodes perform enough aggregation to compen- sate for inaccurate sensors, but their values may be local enough to reveal private data. The network must therefore be aware of these metadata and their role in security. The second security challenge for deeply embedded systems is therefore to identify cost-effective schemes for hiding network timing and obfuscate the underlying topology.

Possible solutions to address this challenge might be based on sending messages at regular intervals, disassociating a reading from a physical event by adding a random delay to message transmission, or adding spurious messages to mask the legitimate send times. Such *confusion* techniques have been used, for instance, to thwart eavesdropping on the internet [2].

Secure aggregation In networks where aggregation occurs at intermediary nodes, end-to-end encryption is not possible because each node must be able to compute with the data. The standard security doctrine that the network should not be trusted and that all messages should be encrypted and decrypted at the source and destination is incompatible with aggregation. Unfortunately, the alternative of trusting each link between the source and the destination is unappealing.

Unlike traditional computing platforms, end users who are identified by sensor nodes have little ability to set policy. When browsing the Internet, for example, users can use anonymizing proxies to protect their privacy. When being sensed by a sensor, however, the end user has no input as to the level of information disclosure, and must trust in the decisions made by the sensor network. Since being sensed can be a passive act and can be done without the knowledge of the observed party, designing networks with privacy guarantees is an arduous task.

The security challenge with respect to data in deeply networked systems is therefore to develop aggregation techniques that are secure, scalable and also ensure the desired level of privacy.

Conclusion Existing systems and methodologies have largely applied the internet model of security to cyber physical systems. However, these approaches fail to make the distinction between physical messages and logical information, focus on all-or-nothing security guarantees, and are ill-equipped to deal with the inherent asymmetry in the distribution of information across the network. In this essay we have highlighted some of the new challenges and also opportunities with respect to security of such systems. We believe that once these challenges are surmounted, applications with intrinsic security considerations will become immediately realizable.

References

[1] M. Anand, Z. Ives, and I. Lee. Quantifying eavesdropping vulnerability in sensor networks. In *Proceedings of the Second International Workshop on Data Management for Sensor Networks*, pages 3–9. ACM Press, 2005.

[2] E. Cronin, M. Sherr, and M. Blaze. On the reliability of internet eavesdropping. In *Technical Report, University of Pennsylvania*, February 2005.