;login:

DECEMBER 2013 VOL. 38, NO. 6

ELECTRONIC SUPPLEMENT

Conference Reports

HealthTech '13: 2013 USENIX Workshop on Health Information Technologies

Washington, D.C.
August 12, 2013

A Healthy Dose of Privacy

Summarized by Michael Rushanan (micharu1@cs.jhu.edu)

Privacy-Preserving Computation of Disease Risk by Using Genomic, Clinical, and Environmental Data

Erman Ayday, Jean Louis Raisaro, Paul J. McLaren, Jacques Fellay, and Jean-Pierre Hubaux, École Polytechnique Fédérale de Lausanne

Erman prefaced his presentation with the disclaimer that "the talk will be mostly dominated by Genomics," and the discussion quickly lead into the proverbial security-oriented rabbit hole; genomic sequence data is yet another sensitive source, perhaps more so than others, and thus requires adequate protection by way of confidentiality. But just how sensitive is the genomic data? Erman led the definition of sensitivity with the question, "Why do we care about protecting genomic data?" The answer is more jarring than the usual financial ruin and public display of intimate correspondence. The genomic sequence of any physical actor represents an irrevocable fingerprint that not only uniquely identifies the individual, but also reveals a great deal about the person's family lineage. Worst yet, services that utilize genomic data to trace personal ancestry and preemptively identify genetic disease risk, such as Ancestry.com and Counsyl.com, are not subject to regulatory compliance with respect to confidentiality.

The talk attendees were all in agreement that services such as the aforementioned have a high utility. This is when Erman solidified his purpose: to provide a practical system for operating on sensitive genomic sequence data, and nongenomic metadata, without compromising the confidentiality of the patient (i.e., the physical actor sharing her genomic data). Erman and his team structure their system on a cryptographic encryption scheme known as "homomorphic encryption" coupled with privacy preserving protocols. An example, albeit abstract, was a pharmacist calculating the risk of a drug with respect to a patient's genomic data.

As for the implementation, Erman was interested in encrypting the single letter difference of single nucleotide polymorphisms (SNP, or "snips") with homomorphic encryption. Though this provides significant overhead, SNPs are directly used to compute disease risk, an odds ratio, and thus the malleable property of homomorphic encryption would allow untrusted actors to perform computations over the encrypted SNPs. Additionally, Erman's system model must consider the confidentiality of an entire DNA sample (on the order of 50 GB/user) and nongenomic metadata. Finally, there are numerous parties that vary in the level of trust: the patient; a certified institution that is trusted and thus performs the encryption; an untrusted storage and processing unit; and a medical unit that can perform disease risk computation.

Marcel Simone (J&J) asked about the scalability of the authors' proposed implementation. Erman replied that the storage overhead, per user, is approximately 51.2 GB. Additionally, Erman concluded that this is due to the code state, proof-of-concept, and that future optimizations should reduce the overhead by a factor of 10.

Understanding the Challenges with Medical Data Segmentation for Privacy

Ellick M. Chan, Peifung E. Lam, and John C. Mitchell, Stanford University

Ellick Chan started by providing the attendees with context about data segmentation and privacy. The example used to begin the discussion involved the sensitive topic of AIDS. Ellick stated that we, the health community, would like to strike a balance between relevancy for both the patient and health care provider, but also prevent sensitive information from being discerned by a third party. This is deemed problematic by Ellick, who went on to describe the following scenario.

A patient has AIDS, and thus the provider attempts to redact meaningful data such that a third party may not infer the patient's full condition. First, the ICD-9 code 042 (AIDS) is redacted. Upon closer inspection of the record, azidothymidine (AZT) is listed under prescriptions and this leaks information about the patient's condition as it is used exclusively for AIDS treatment. Additionally, record data such as short/long-term medical conditions (e.g., anemia) and sexual orientation can enable a third party to infer the AIDS condition.

Ellick codifies this problem into a threat model where our proposed attacker is passive; she has direct access to the redacted medical record, and is computationally bounded (i.e., she won't be able to break NIST-recommended encryption standards). To better understand her capabilities, Ellick defines a model in which to map diseases to manifestations. This approach allows Ellick to quantify what the attacker is able to discern from a redacted medical record as well as structure segmentation that is relaxed with respect to proximity and relevancy. However, if this predicate-reducer algorithm fails to provide accurate and reproducible results, it could result in a misdiagnosis.

Jean-Pierre Hubaux asked Ellick to return to the threat model as he was interested in the health care provider; is the health care provider trusted? Ellick responded that the health care provider is trusted, but one might imagine transferring records to other practices that are domain specific, and thus the entirety of the record may not want to be divulged. Jean-Pierre then asked if this threat model considered broad-scale or individual attacks? Ellick responded that it is geared toward individuals at the moment, but has the potential to be applied more globally.

Privacy Aspects of Health-Related Information Sharing in Online Social Networks

Sadegh Torabi and Konstantin Beznosov, University of British Columbia

Lujo Bauer presented this work at HealthTech on the authors' behalf. The authors wanted to know how much online health information sharing is actually going on over social networks, what kinds of things were being shared, and how social network users perceived the privacy risks.

In terms of methodology, the authors' survey sample was collected via the Mechanical Turk (i.e., using crowd-sourcing to collect survey results). Questions on the survey consisted of: How often do you share health-related information (HRI)? Why do you share HRI? Why did you not want to share health-related information? What factors affect your perceived privacy risks (recipient of info, HRI type, HRI category, online social network where HRI shared)? How risky is it to share information in specific categories (select individuals, groups, all contacts, all other users)? How would you manage your HRI better (not sharing, manipulating shared data)?

The data collected suggested that HRI categories are closely correlated to how survey respondents evaluated risk. For example, healthy living was perceived as less risky (15.7% never shared) than HRI of people in the respondents' custody (54.8% never shared). For HRI that was shared, respondents indicated that the need to help others by sharing a personal experience (66.9%) and to seek help/support (51.8%) were leading causes for sharing HRI.

Tony Dahbura (Johns Hopkins) reasoned about the disparity of what people say they do and what they actually do. Lujo acknowledged this concern but reaffirmed that this initial study was the first of its kind and was more interested in collecting a baseline. Tony concluded his questions by implicating the disparity between the Likert scale and the sample population (skewed toward a younger sample of 18–26). Lujo agreed that the questions could be refined to collect more accurate data.

Medical Device Security from the Bottom Up

Summarized by Michael Rushanan (micharu1@cs.jhu.edu)

Using Bowel Sounds to Create a Forensically-Aware Insulin Pump System

Nathan L. Henry, University of Tennessee; Nathanael R. Paul, University of Tennessee and Oak Ridge National Laboratory; Nicole McFarlane, University of Tennessee

The authors' goal was to improve the security and medical condition of diabetics who use insulin pumps. Nathan Henry presented their new sensor system which involves the use of bowel sounds to perform what he describes as "[an improvement of] forensic security for insulin pumps." To start the discussion, Nathan mentioned prior work from the attack perspective—Radcliffe's "Hacking Medical Devices for Fun and Insulin: Breaking the Human SCADA System,"—and from the general non-invasive defense perspective, Gollakota et al., "They Can Hear Your Heartbeats: Non-Invasive Security for Implantable Medical Devices." The authors' approach to forensic security follows an out-of-band side-channel which, Nathan claims, will prevent negative events (i.e., intentional negative patient events that affect the patient's well being) incited by traditional security overhead to the embedded system.

The abstract implementation of this technique is described by Nathan as follows: To determine a negative event, interrupt a disperse call; measure bowel sounds from the subject tested; diagnose collected sound samples to evaluate time delta of last meal consumption; and assert whether to disperse insulin call was malfeasant or not. To achieve a runtime operation like the above, Nathan recounted his experimental setup. This setup required an electronic stethoscope, a significant collection of bowel sounds at different time deltas post eating, and targeted signal processing to eliminate noise. Additionally, Nathan required five subjects and admitted early on that his solution had to be tailored per subject.

Nathan informed attendees that false negatives in the preliminary work were high and that one of the greatest challenges was canceling out noise (e.g., subject movement and talking). However, he did report that bowel sound detection, when the subject remained still and quiet, would rise 1.5 times above the baseline measurement within the first five minutes after the start of a meal. In concluding, Nathan mentioned their future work is to provide the seamless integration of bowel sensing to the insulin pump embedded system.

Michael Rushanan asked about the sampled bowel noise distribution and whether it was deterministic? Nathan replied that they hadn't considered the randomness of the noise collection, only that certain thresholds of bowel sounds indicate having eaten or not.

WattsUpDoc: Power Side Channels to Nonintrusively Discover Untargeted Malware on Embedded Medical Devices

Shane S. Clark, University of Massachusetts Amherst; Benjamin Ransford, University of Washington; Amir Rahmati, University of Michigan; Shane Guineau, University of Massachusetts Amherst; Jacob Sorber, Clemson University; Kevin Fu, University of Michigan; Wenyuan Xu, University of South Carolina and Zhejiang University

Shane described the purpose of their research as gaining greater visibility into medical device malware infections. The motivation came from the Manufacturer and User Facility Device

Experience (MAUDE), which is the FDA's solution for monitoring and logging adverse events occurring in medical devices. These adverse events, described as "tension over availability and safety of a device, confirm the fears of security researchers—there is indeed malware impacting these devices." Unfortunately, the practice of discovering malware infections on medical devices is not straightforward.

Shane provided us our first barrier: software changes are disallowed due to concerns about certification/manufacturer interactions; the resistance to update due to negative exposure. The second barrier is network availability and arduous manual configuration. The question then becomes, how do we discover malware on these medical devices? The authors' solution: WattsUpDoc, which provides a way to validate the embedded system via power-line monitoring. Power analysis is the process of making inferences over time, and this works well for medical devices as the devices are limited in scope and physical deployment.

To evaluate this approach, two devices—a Baxa ExactaMix 2400 compounder and a Schweitzer SEL3354 substation computer—were tested for malware infection via power analysis, both running on Windows XP Embedded. Power analysis was done via a trace collection that requires the above medical devices to be plugged into a special outlet that provides sampling, data acquisition, and offline storage for later analysis. WattsUpDoc continued to grab large training sets of normal and abnormal conditions that involved emulated and real malware (e.g., key logger). Once these traces were collected, the team could identify a confirmation of the intuition that the extra workload of malware introduced odd power fluctuations.

Next, they moved forward with a quantitative and automatic method for verifying the power fluctuations that first involved extracting a meaningful feature vector (skewness, variance, and root mean square). The experimental setup was described as partitioning the power traces into training/testing sets and training using stratified 10-fold validation. Shane described the calculation of precision and recall to identify false positive and false negative percentages to validate the accuracy of their approach. They found there was 94% accuracy on the compounder and 99% accuracy on the substation.

Jean-Pierre Hubaux asked about any guarantees against stealthy malware that might evade power analysis. Shane responded that the goal here is to protect devices that are primarily collateral damage and thus infected by generic malware; this technique is not applicable to personal computers. Ann Cox (DHS) asked about medical devices in the context of a patient's home. Shane responded that they have yet to look at devices in a patient's home, but imagines that these devices are far easier to experiment with due to accessibility. Marcel (J&J) asked about the ability to get rid of malware on medical devices—what do you do with it? Shane replied that currently the approach is to wipe the device clean. Someone asked about the sustainability of this approach—do we have to train for each device? Shane said that it should scale for the same hardware, but they do not have the ability to test this right now.

Daren Lacey (Johns Hopkins) asked if he had plugged this analysis device in at his medical campus, would he be out of spec? Had he "voided the warranty?" Shane's answer, as best he could tell, was probably not. Shane had spoken to someone off-the-record and that person's best guess was that the analysis device was not classified as a medical device, so there was probably no way to run afoul of the FDA. Dennis Schneider (Baxter Health Care) asked about Windows Embedded and staged updates. It seemed to him that updates would change the analysis; how did Shane propose to keep traces updated? Shane answered that he had done a few tests on Windows boxes at the lab and that updates had not affected the analysis, though he didn't claim this would generalize. Dennis ended with the question, "Wouldn't a drastic change to the GUI or embedded antivirus dramatically change traces?" Shane replied that this was possible, though major changes to the GUI seemed unlikely.