Persistent Security, Privacy, and Governance for Healthcare Information

W. Knox Carey
Intertrust Technologies
955 Stewart Drive
Sunnyvale, CA 94085
knox@intertrust.com

Jarl Nilsson
Intertrust Technologies
955 Stewart Drive
Sunnyvale, CA 94085
jnilsson@intertrust.com

Steve Mitchell
Intertrust Technologies
955 Stewart Drive
Sunnyvale, CA 94085
mitchell@intertrust.com

***Abstract*—A fundamental tension between accessibility and governance exists in the design of healthcare information systems. In order to be useful in practice health information must be distributed, but as the information moves between systems — and different information governance policies — the risk of privacy and security violations increases. The lack of a persistent policy enforcement mechanism thus inhibits the dissemination of health information, making it less useful for research and treatment. In this paper, we argue that trusted computing and policy management technologies are required to allow for broad distribution of health information while preserving security and privacy. We also introduce the concept of *derived resources*, which helps to address many of the unique challenges in the governance of health information.**

I. Introduction

Medical information is everywhere, but only rarely where it needs to be to make a difference. A public health researcher studying the propagation of a new disease needs timely, comprehensive information drawn from front-line healthcare institutions across the country. A family doctor seeing a new patient might avoid repeating expensive tests if he could only access the patient's lab results from last year. Health data recorded by home monitoring systems sit locked in a personal computer, when they should be informing diagnosis and treatment in the hands of a clinical practitioner. All of these examples depend on the frictionless flow of medical information through a heterogeneous network of devices and systems. Unfortunately, this information flow is not happening.

In an ideal world, comprehensive electronic health records would be available instantly to doctors, researchers, and other stakeholders, regardless of the original source of the information. In practice, sharing data across systems involves significant risks to the security and privacy of patient data. When patients and institutions *do* release information, they have little assurance that it will be governed in a manner that is consistent with their policies.

The lack of persistent governance throughout the lifecycle of medical information means that every interface between two systems is a potential point of compromise. In a technologically diverse environment where governance of medical information cannot be guaranteed across institutions, ensuring compliance with regulations requiring security and privacy becomes an intractable problem.

In this paper, we describe some of the elements necessary for a solution to the problems of governed information sharing, and discuss how new developments in trusted computing enable new applications for patient privacy, data security, and medical research.

II. Elements of the Solution

Many of the technologies required to provide persistent governance of medical information have been in use for many years in other fields. Applying these technologies to healthcare will require a multidisciplinary approach, a rethinking of basic assumptions about security and governance for healthcare information systems, and a set of new technologies specifically adapted to this application area.

A. Persistent Governance of Medical Information

Providing for governance of medical information across heterogeneous systems requires an expanded perspective on the nature of data security, one that takes into account not only *access* to governed information, but also the *managed use* of that information.

Since the earliest days of computing, sensitive information has been secured primarily by keeping it isolated within a carefully guarded perimeter that admits only authorized individuals [CSTPS]. Access control models have evolved in step with technology development (e.g. [XACML]), but all such systems still retain the property that the use of sensitive information — once access is granted — is relatively unrestricted.

In the 1990s, the first digital rights management (DRM) systems [DBOX] introduced the notion of *persistent governance* of information. Not only were data protected cryptographically, but use of those data were subject to certain rules that were securely associated with the data. DRM and other trusted computing technologies ensured that the policies specified for the data would be enforced consistently wherever the data traveled. By contrast, the policies governing information in older systems depended as much on the location of the data (at rest, in transit over a SAC, buffered in an intermediate system, etc.) as they did on the intention of the data originator.

The ability to persistently govern information across systems enables new possibilities for the dissemination of sensitive information, possibilities that are not realizable with more traditional forms of access control:

- Data can be transmitted through a heterogenous network of systems, even untrusted systems, with no degradation in security. Data can be copied and distributed freely.
- Data can be consumed offline — rules are evaluated locally at the point of access. Consuming systems need not contact a remote policy decision point.
- Data and the rules governing it can be distributed separately. As a result, data may be distributed in advance of any rules, and new rules may be associated with the data at any time. It is not necessary to know all of the rules that will govern use of a particular set of data *a priori*.
- Data can be packaged with rules that enforce very fine-grained usage policies. For example, access to data can depend upon time, the accessing principal, the membership of the accessing device in a group, and so forth.

The capabilities described above are essential for ensuring consistent governance of healthcare information. For example, institutions that collect and maintain medical data have certain legal obligations to secure patient records, especially when the records contain individually identifiable information [HIPAA]. Under HIPAA security rules, certain disclosures are authorized, but the original institution has no mechanisms at their disposal to ensure that the information disclosed is not being misused, or indeed to have any ongoing relationship with the medical information once it leaves their facility.

In a world of uniform policies, in which every institution or system that handles medical information is governed by the same rules — the same combination of legal requirements, corporate policies, patient preferences — simple exchange of information over a secure channel might be sufficient. Unfortunately, such policy uniformity does not exist in practice. This is especially true when medical information systems incorporate data collected or uploaded by patients themselves.

The first prerequisite, then, for persistent governance of medical information is to incorporate ideas pioneered in trusted computing, policy management, and digital rights management into the handling of medical information. Older access control models are simply insufficient to meet the challenge at hand.

B. Consistent Trust Management

Trust management systems fulfill two primary functions in a secure system:

1) They associate names and other attributes of a principal with the verifiable use of secret information by that principal. Each principal is assigned security credentials that assert these attributes.
2) They ensure that credentialed principals have met certain criteria set by the trust management system. This means that corresponding principals that cannot realistically verify compliance with these criteria can rely upon the trust management system as a trusted third party.

As applied to the persistent governance of information, trust management systems ensure that credentialed systems with access to governed information evaluate and enforce the rules associated with that information. A consistent trust management framework is required for medical systems so that the originator of medical data — the entity that associates policy with the data — can rely upon any credentialed recipient without the necessity of establishing a relationship in advance.

The healthcare community has long recognized the urgency of addressing the problem of providing trust management for medical information systems [MTMIS], and advisory bodies to the US federal government have begun to address requirements in this area [FHITS]. The context for these recommendations however, has tended to focus on securing communications between endpoints rather than providing assertions of compliance to certain policies. This focus will need to be expanded to facilitate persistent governance of information across heterogeneous systems.

C. Derived Resources

Applying trusted computing technologies in healthcare is not simply a matter of adopting existing techniques used in other problem domains — governance of medical information presents unique requirements that have no parallels in other fields.

- The information being protected concerns an individual with legal rights as to how the data are disseminated and used. These rights vary by jurisdiction, and, to the degree possible, should be under the control of the patient.
- Various stakeholders require (or are entitled to) different views of the same data. For example, a patient may be interested in every data point in a series of health metrics, whereas his physician is only interested in a summary of the trends. Epidemiology researchers might see the data, but they should see it at perhaps a lower resolution, and certainly in an anonymized form.
- Different aspects of the governed data may be important over time. Initially, for example, a doctor may be interested only in the overall trend for a particular health metric. When an anomalous circumstance is discovered, however, the full data series may become important. This phenomenon is also important in research, where revisiting data years later can shed new light on older studies [CURRY]. It should not be necessary to repackage older information to enable this feature.
- To maximize utility, the data need to be distributed as broadly as possible, but the data should not be distributed in an ungoverned manner — it should always be possible for the owner or originator of the information to control data access policies and to audit actual usage.

To meet these requirements, we introduce the concept of a *derived resource*. In most governance systems, the resource to be protected is a static object such as a video file or a document. Typically, granting access to a resource consists of applying a set of conditions to determine if access is possible and then producing the key that allows the consuming system to decrypt the resource, which is uniform for all users. As the requirements above indicate, this uniformity is not sufficient in

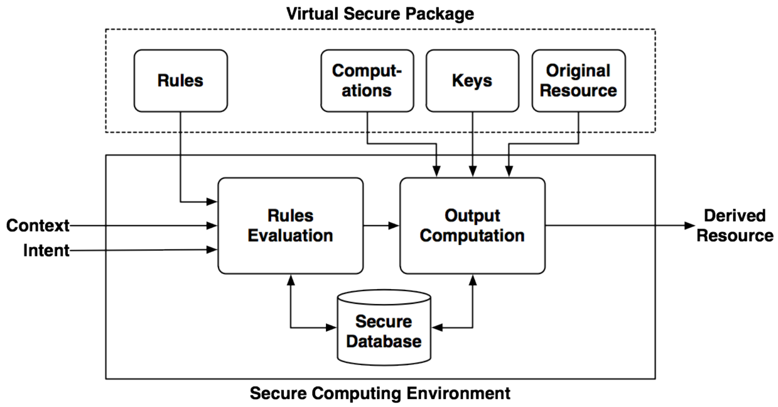

Fig. 1: A system that provides access to derived resources. In practice, each of the elements shown in the *virtual secure package* may be distributed separately. The dotted line surrounding the package indicates that the elements described are securely associated with one another.

healthcare applications — different stakeholders have different interests in the data that may change over time.

Derived resources address this problem by securely associating a set of specified computations — performed on the resource itself — with the packaged resource. As shown in Figure 1, a resource to be protected is associated with a set of rules governing access, keys to allow decryption, and computations to be applied to the original resource before returning it to the requesting system. The computations may depend on several factors, including the identity of the principal that will access the data, environmental considerations at the point of evaluation of the computation, or state information maintained by the system at the client accessing the derived resource.

For some types of media, it is possible to partially address this problem by pre-encoding resources with a multiresolution encoding scheme, such as SVC [SVC], and creating rules that govern each of the pre-computed static resources as if they were separate resources. This approach is less applicable in healthcare applications, however, as the particular required view of the data may change over time.

The derived resources scheme has several properties that enable the applications described in the next section:

- The precise view required of a set of raw information need not be computed in advance; the packager simply associates a computation that produces the derived resource. These computations can be reusable for different data sets, e.g. produce a five-day trailing average over the enclosed data series.
- Since computations may depend on conditions such as the principal accessing the information, different stakeholders in the packaged resource may obtain different views; different computations are associated with each principal.
- New views of the resource can be provided after the fact. If a new type of derived resource becomes important after the original resource is already protected, the packager can simply provide a new set of associated computations rather than recomputing and repackaging the entire data set. These new computations can of course be generated by the original packager, but perhaps equally importantly, they may be proposed by the users of the data and selectively authorized by the owner of the resource.
- Derived resources can be distributed broadly, without necessarily granting unlimited access to all recipients. As such, this technique provides a solution for the tension at the heart of many healthcare data management problems — the need to publish information versus the need to manage usage of the information.
- Creating a derived resource can be lossless. Using derived resources, the original data need not be repeatedly filtered and repackaged, so no information is lost that may be of use in the future.
- The computations performed as a condition for rendering can be expressed for a standardized machine (as in the Octopus system [OCTO]), or declaratively, using agreed-upon semantics. The packager may validate, audit, and rely upon the results computed at the accessing system.

Adding derived resources to existing trust computing models enables new uses that are difficult or impossible to realize with older technologies. The next sections describe some of these applications specifically for healthcare information.

III. Applications

A. Protecting Patient Privacy

One of the most important applications for derived resources is protecting patient privacy while ensuring that information is distributed to the points where it is required. Consider, for example, a diabetic patient that is recording blood glucose levels at home using an non-invasive glucose monitor [NGM].

The data, collected once per hour, are synchronized with an online service that allows the patient to chart his blood glucose over time. The service also actively packages and forwards the information to the patient's physician, who can use it to evaluate the effectiveness of the prescribed regimen. The packages sent to the physician are associated with computations that grant access only to a set of authorized principals. The computations can be specified such that the physician herself has access to the full set of data, whereas colleagues may see the data only in a partially anonymized form.

B. Filtering Diagnostic Information

In most cases, healthcare workers are not interested in all of the details of a particular data series; the salient information is contained in a few metrics that are computed from the data. To continue the previous example, a physician may be interested only in the peak blood glucose over a week period rather than the hour-by-hour data points. Default computations associated with the resource can therefore produce just the desired indicators.

On the other hand, when the situation merits a more detailed investigation — e.g. the blood glucose peaks predictably at certain times — the physician can apply alternative computations that produce higher-resolution data. Parameterized computations may allow users to view data at different resolutions within a predetermined range, as necessary.

C. Data for Medical Research and Epidemiology

The emergence of relatively low-cost home medical monitoring devices, especially when coupled with technology that ensures the integrity and trustworthiness of the data collection process, has the potential to fundamentally alter the nature of medical research. The overwhelming volume of data that can be collected in the course of patients' everyday lives (as opposed to a more artificial clinical setting) presents a massive opportunity to understand and transform health.

In order to preserve privacy, however, the data must be at least partially de-identified before it is distributed. Generally speaking, the more widely a data set is disseminated, the more anonymous that data should be. The patient himself should most likely have full access to all of his own biometrics, as should direct caregivers, but people beyond that circle should see less and less personally identifiable information, unless the patient chooses to provide it. Systems that collate medical information, for example, might allow users to opt-in to a program that automatically sends their data, with an associated anonymization computation, to public health authorities.

The same mechanisms may be used to create a market for selectively anonymized health information. For example, a researcher might offer incentives (financial or otherwise) to patients in a given demographic group who provide access to their data for a given period of time at a given resolution.

D. Integrity of Medical Research

The fact that information is not lost in creating a derived resource helps to solve an important problem in medical research, and in scientific research in general. When scientists prepare a publication, they typically filter the data that they have obtained in their research for clarity of presentation. Unfortunately, this filtering process makes it difficult to validate the original data or to reinterpret it using new algorithms. The progress of scientific research is served when the data are available, but scientists (and their funders) have justified concerns with the completely open publication of data that they have collected [TRAV].

Using derived resources, scientists can publish data that they have collected, along with the computations that produce their reported results. Publication of the computations allows other researchers to validate the results, and also to suggest new computations that illuminate different aspects of the same data. By associating a new computation with the data set, the original scientist can allow his work to be built upon by others without ceding control of the data.

IV. Conclusions

In healthcare, information that is not shared cannot be used to treat patients and improve health. A patchwork of interacting policy environments that were not designed to work together provides little assurance that shared health information will be used consistently with the wishes of all stakeholders in that information. On the other hand, the persistent governance afforded by trusted computing technologies facilitates the broad distribution of healthcare data while simultaneously providing for its security and privacy. Moving from static, uniform resources to derived resources — expressed as computations applied to the original governed resource — enables new use cases in medical privacy, clinical use, and research.

References

[CSTPS] J. P. Anderson *Computer Security Technology Planning Study, ESD-TR-73-51, ESD/AFSC*, Hanscom AFB, Bedford, MA, Oct. 1972, Volume 1.

[XACML] DeCouteau, Davis, and Stagges, ed. *Cross-Enterprise Security and Privacy Authorization (XSPA) Profile of XACML v2.0 for Healthcare Version 1.0.* http://docs.oasis-open.org/xacml/xspa/v1.0/xacml-xspa-1.0-cs02.html.

[HIPAA] U.S. Department of Health and Human Services. *The Privacy Rule*. http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/index.html

[MTMIS] Blaze, Feigenbaum, and Lacy. *Managing Trust in Medical Information Systems*. AT&T Internal Research Paper. http://www.eyetap.org/~maali/trust-papers/blaze96managing.pdf

[DBOX] Sibert, Bernstein, and Van Wie. *A Self-Protecting Container for Information Commerce*. Proceedings of the First USENIX Workshop on Electronic Commerce, New York, New York, July 1995. http://www.usenix.org/publications/library/proceedings/ec95/full_papers/sibert.txt

[FHITS] Office of the National Coordinator. *Federal Health Information Technology Strategic Plan, 2011-2015*. http://healthit.hhs.gov/portal/server.pt/community/federal_health_it_strategic_plan_-_overview/1211

[CURRY] Andrew Curry. *Rescue of Old Data Offers Lesson for Particle Physicists*. Science, vol. 331, pp. 694–695, 11 Feb. 2011.

[SVC] Schwarz, Marpe, and Wiegand. *Overview of the Scalable Video Coding Extension of the H.264/AVC Standard.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[OCTO] Boccon-Gibod, Boeuf, and Lacy. *Octopus: an application independent DRM toolkit*. Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference, pp. 1148–1154, Jan. 2009.

[NGM] Omar S. Khalil. *Non-Invasive Glucose Measurement Technologies: An Update from 1999 to the Dawn of the New Millennium.* Diabetes Technology & Therapeutics, vol. 5, no. 5, pp. 660-697., Oct. 2004.

[TRAV] Kate Travis. *Sharing Data in Biomedical and Clinical Research*. Clinical and Translational Science Network, 11 Feb. 2011. http://community.sciencecareers.org/ctscinet/articles/2011/02/sharing-data-in-biomedical-and-clinical-research.php