

Universidad Paraguay Aleman



**UNIVERSIDAD PARAGUAYO ALEMANA
HEIDELBERG - ASUNCIÓN**



Seguridad TICs

Prof.: Chrystian Ruiz Diaz

Contenido

Nota de Uso Académico.....	3
Seguridad Multilateral.....	4
9.1 Introducción	4
9.2 Compartimentación, el Muro Chino y el Modelo de la BMA	6
9.2.2 El Muro Chino	9
9.2.3 El Modelo de la BMA	10
9.2.3.1 El Modelo de Amenaza	12
9.2.3.2 La Política de Seguridad	15
Implementaciones Piloto	17
9.2.4 Cuestiones Actuales de Privacidad	18
9.3 Control de Inferencias	20
9.3.1 Problemas Básicos del Control de Inferencias en Medicina.....	21
Otras Aplicaciones del Control de Inferencias.....	23
9.3.3 La Teoría del Control de Inferencias	24
9.3.3.1 Control del Tamaño del Conjunto de Consultas	25
9.3.3.2 Rastreadores	25
9.3.3.3 Controles de Consultas Más Sofisticados	26
9.3.3.4 Supresión de Celdas	26
Control de Máximo Orden y el Modelo de Lattice	27
9.3.3.6 Control Basado en Auditoría.....	28
9.3.3.7 Aleatorización	29
9.3.4 Limitaciones de los Enfoques Genéricos	30
Active Attacks	30
9.3.5 El Valor de la Protección Imperfecta	32
9.4 El Problema Residual	33
9.5 Resumen	36
Problemas de Investigación.....	36
Lecturas Adicionales	37

Nota de Uso Académico

Este documento ha sido preparado exclusivamente para el uso académico del docente. Este documento no puede ser distribuido a ninguna otra parte ni utilizado para ningún otro propósito, diferente al establecido como alcance en el proyecto académico de la **UNIVERSIDAD PARAGUAYO ALEMANA**. El uso indebido del material fuera del ámbito académico no representa ninguna responsabilidad del docente.

Seguridad Multilateral

9.1 Introducción

A menudo, nuestro objetivo no es prevenir que la información fluya "hacia abajo" en una jerarquía, sino evitar que fluya "a través" de departamentos. Las aplicaciones relevantes abarcan desde la atención médica hasta la inteligencia nacional, e incluyen la mayoría de las aplicaciones donde está en juego la privacidad de los datos de clientes individuales, ciudadanos o pacientes. Representan una proporción significativa de los sistemas de procesamiento de información, pero su protección suele estar mal diseñada e implementada, lo que ha llevado a una serie de fracasos costosos.

El problema básico es que si centralizas sistemas que contienen información sensible, corres el riesgo de crear un activo más valioso y, al mismo tiempo, dar acceso a más personas. Este es ahora un problema urgente en el mundo de 'Web 2.0', ya que las aplicaciones en línea acumulan petabytes de información privada de las personas. Y no solo se trata de Google Documents; varias organizaciones planean almacenar tus registros médicos en línea. Microsoft ha anunciado HealthVault, que permitirá a tus médicos almacenar tus registros médicos en un centro de datos y darte cierto control sobre el acceso; otras empresas de TI tienen planes similares. Sin embargo, los activistas por la privacidad señalan que, por muy conveniente que esto sea en una emergencia, da acceso a compañías de seguros, agencias gubernamentales y cualquier otra persona que se presente con una orden judicial [1332]. Entonces, ¿cuáles son los problemas reales con tales sistemas? ¿Deberían construirse? Si es así, ¿cómo deberíamos protegerlos? ¿Y hay precedentes de los cuales podamos aprender?

Una lección proviene de la banca. En los viejos tiempos, un investigador privado que quería copias de tus extractos bancarios tenía que subvertir a alguien en la sucursal donde se mantenía tu cuenta. Pero después de que los bancos conectaron todas sus sucursales en línea en los años 80, generalmente permitieron que cualquier cajero consultara la cuenta de cualquier cliente. Esto trajo la conveniencia de poder cobrar un cheque cuando estás fuera de la ciudad; pero también significó que los detectives privados compren y vendan tus extractos bancarios por unos pocos cientos de dólares. Solo tienen que corromper a un empleado en cada banco, en lugar de uno en cada sucursal. Otro ejemplo proviene de la UK Inland Revenue, la oficina de recaudación de impuestos; se descubrió que el personal accedía indebidamente a los registros de celebridades, vendía datos a terceros y filtraba detalles de ingresos en casos de pensión alimenticia [129].

En tales sistemas, un requisito típico será detener a los usuarios de mirar registros que pertenecen a una sucursal diferente, una región geográfica diferente o un socio diferente en la empresa, excepto bajo controles estrictos. Por lo tanto, en lugar de que los límites de control del flujo de información sean horizontales como vimos en el modelo de Bell-LaPadula en la Figura 9.1, en su lugar necesitamos que los límites sean principalmente verticales, como se muestra en la Figura 9.2.

Estos controles de flujo de información lateral pueden ser organizacionales, como en una organización de inteligencia que desea mantener los nombres de los agentes que trabajan en un país extranjero en secreto del departamento responsable de espiar en otro.

Pueden estar basados en privilegios, como en un bufete de abogados donde los asuntos de diferentes clientes, y los clientes de diferentes socios, deben mantenerse separados. Incluso pueden ser una mezcla de ambos, como en la medicina, donde la confidencialidad del paciente

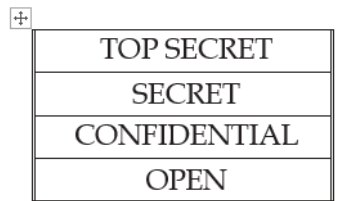


Figure 9.1: Multilevel security

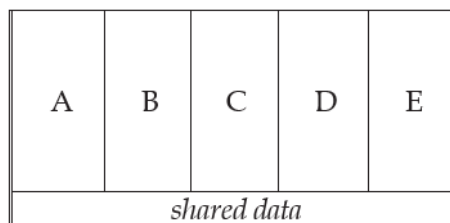


Figure 9.2: Multilateral security

La base legal se fundamenta en los derechos del paciente, pero generalmente se aplica limitando el acceso a los registros médicos a un departamento hospitalario en particular.

El control de los flujos de información lateral es un problema muy general, del cual usaremos la medicina como un ejemplo claro y bien estudiado. Los problemas de los sistemas médicos son fácilmente comprensibles para los no especialistas y tienen una considerable importancia económica y social. Gran parte de lo que decimos sobre ellos se aplica con poco o ningún cambio a la práctica de otras profesiones y a aplicaciones gubernamentales donde el acceso a ciertos tipos de datos clasificados está restringido a equipos o departamentos específicos.

Un problema menor que enfrentamos es el de la terminología. Los controles de flujo de información del tipo que nos interesa son conocidos por varios nombres diferentes; en la comunidad de inteligencia de EE.UU., por ejemplo, se conocen como seguridad compartimentada o compartimentación. Usaremos el término europeo seguridad multilateral, ya que la aplicación en el sector salud es mayor que en inteligencia, y porque el término también abarca el uso de técnicas como el anonimato, siendo el caso clásico las bases de datos de registros médicos desidentificados para la investigación. Esto es una parte importante de la seguridad multilateral. Además de prevenir los flujos de información evidentes, también debemos evitar la fuga de información a través, por ejemplo, de datos estadísticos y de facturación que se liberan.

El uso de datos desidentificados tiene una aplicación más amplia. Otro ejemplo es el procesamiento de datos censales. En general, las técnicas de protección relevantes se conocen como control de inferencias. A pesar de las ocasionales diferencias en la terminología, los problemas que enfrentan los operadores de bases de datos censales y de investigación médica son muy similares.

9.2 Compartimentación, el Muro Chino y el Modelo de la BMA

Existen (al menos) tres modelos diferentes de cómo implementar controles de acceso y controles de flujo de información en un modelo de seguridad multilateral. Estos son la compartimentación, utilizada por la comunidad de inteligencia; el modelo del Muro Chino, que describe los mecanismos utilizados para prevenir conflictos de interés en la práctica profesional; y el modelo de la BMA, desarrollado por la Asociación Médica Británica para describir los flujos de información permitidos por la ética médica. Cada uno de estos tiene aplicaciones potenciales fuera de su campo inicial.

9.2.1 Compartimentación y el Modelo de Lattice

Durante muchos años, ha sido práctica estándar en los Estados Unidos y gobiernos aliados restringir el acceso a la información mediante el uso de palabras clave además de clasificaciones. El ejemplo mejor documentado es la palabra clave Ultra en la Segunda Guerra Mundial, que se refería a las descifraciones británicas y americanas de mensajes alemanes cifrados con la máquina Enigma. El hecho de que se hubiera roto el código de Enigma era tan importante que valía la pena protegerlo a casi cualquier costo. Así, las autorizaciones Ultra se dieron a un pequeño número de personas: además de los criptanalistas y su personal de apoyo, la lista incluía a los líderes aliados, sus generales principales y analistas seleccionados. Nadie que hubiera tenido una autorización Ultra podría estar en riesgo de captura, y la inteligencia nunca podría utilizarse de una manera que dejara a Hitler sospechar que su principal cifrado había sido roto. Así, cuando Ultra informaba de un objetivo, como un convoy italiano hacia el norte de África, los aliados enviaban un avión para "detectarlo" y reportar su posición por radio una hora antes del ataque. Esta política se aplicaba mediante reglas de manejo especial; por ejemplo, Churchill recibía sus resúmenes Ultra en una caja de despacho especial a la que él tenía una llave pero su personal no. Debido a que tales reglas especiales pueden aplicarse, el acceso a una palabra clave a veces se denomina una indoctrinación en lugar de simplemente una autorización. (La seguridad de Ultra se describe en Kahn [677] y en Welchman [1336]).

Hoy en día, se toman precauciones similares para proteger la información cuya filtración podría exponer fuentes o métodos de inteligencia, como nombres de agentes, éxitos criptanalíticos, capacidades de equipos utilizados para espionaje electrónico y el rendimiento de satélites de vigilancia. La proliferación de palabras clave resulta en una gran cantidad de compartimentos, especialmente en niveles de clasificación por encima de Ultra Secreto.

Una razón para esto es que las clasificaciones se heredan por trabajo derivado; por lo tanto, un informe escrito utilizando fuentes de 'Secreto Tormenta del Desierto' y 'Ultra Secreto Umbral' solo puede ser leído en teoría por alguien con una autorización de 'Ultra Secreto' y pertenencia a los grupos 'Umbral' y 'Tormenta del Desierto'. Cada combinación de palabras clave forma un compartimento, y algunas agencias de inteligencia tienen más de un millón de compartimentos activos. Gestionarlos es un problema significativo. Otras agencias permiten que las personas con autorizaciones de alto nivel tengan un acceso relativamente amplio. Pero cuando los mecanismos de control fallan, el resultado puede ser desastroso. Aldrich Ames, un oficial de la CIA que había acumulado acceso a un gran número de compartimentos debido a su largo servicio

y senioridad, y porque trabajaba en contrainteligencia, pudo traicionar casi toda la red de agentes de EE.UU. en Rusia.

Las palabras clave son en efecto una forma pre-computacional de expresar grupos de control de acceso, y pueden ser manejadas usando una variante de Bell-LaPadula, llamada el modelo de lattice. Las clasificaciones junto con las palabras clave forman un lattice, una estructura matemática en la que cualquier dos objetos A y B pueden estar en una relación de dominancia $A > B$ o $B > A$. No tienen que estarlo: A y B podrían ser simplemente incomparables (pero en este caso, para que la estructura sea un lattice, tendrán un límite superior mínimo y un límite inferior máximo). Como ilustración, supongamos que tenemos una palabra clave, digamos 'Crypto'. Entonces, alguien autorizado para 'Ultra Secreto' estaría autorizado a leer archivos clasificados 'Ultra Secreto' y 'Secreto', pero no tendría acceso a archivos clasificados 'Secreto Crypto' a menos que también tuviera una autorización criptográfica. Esto puede expresarse como se muestra en la Figura 9.3.

Para que los sistemas de información soporten esto, necesitamos destilar la esencia de las clasificaciones, autorizaciones y etiquetas en una política de seguridad que podamos utilizar para dirigir objetivos de seguridad, implementación y evaluación. Por coincidencia, el modelo Bell-LaPadula se mantiene más o menos sin cambios. Todavía tenemos flujos de información entre Alto y Bajo como antes, donde Alto es un compartimento que domina a Bajo. Si dos nodos en un lattice son incompatibles, como con 'Ultra Secreto' y 'Secreto Crypto' en el diagrama anterior, entonces no debe haber flujo de información entre ellos en absoluto.

De hecho, los modelos de lattice y Bell-LaPadula son esencialmente equivalentes, y se desarrollaron al mismo tiempo.

- Roger Schell, Peter Downey y Gerald Popek de la Fuerza Aérea de EE.UU. produjeron un modelo de lattice temprano en 1972 [1119].
- Una tesis doctoral de Cambridge por Jeffrey Fenton incluyó una representación en la que las etiquetas se manejaban usando una matriz [464].
- Alrededor de este tiempo, el Sistema de Comando y Control Militar Mundial del Pentágono (WWMCCS) utilizó un modelo de lattice primitivo, pero sin la propiedad *. La demostración de que un sistema crítico en funcionamiento que manejaba datos Ultra Secretos era vulnerable a ataques por troyanos causó cierta consternación [1118]. Significaba que todos los usuarios tenían que estar autorizados al nivel más alto de datos en la máquina.

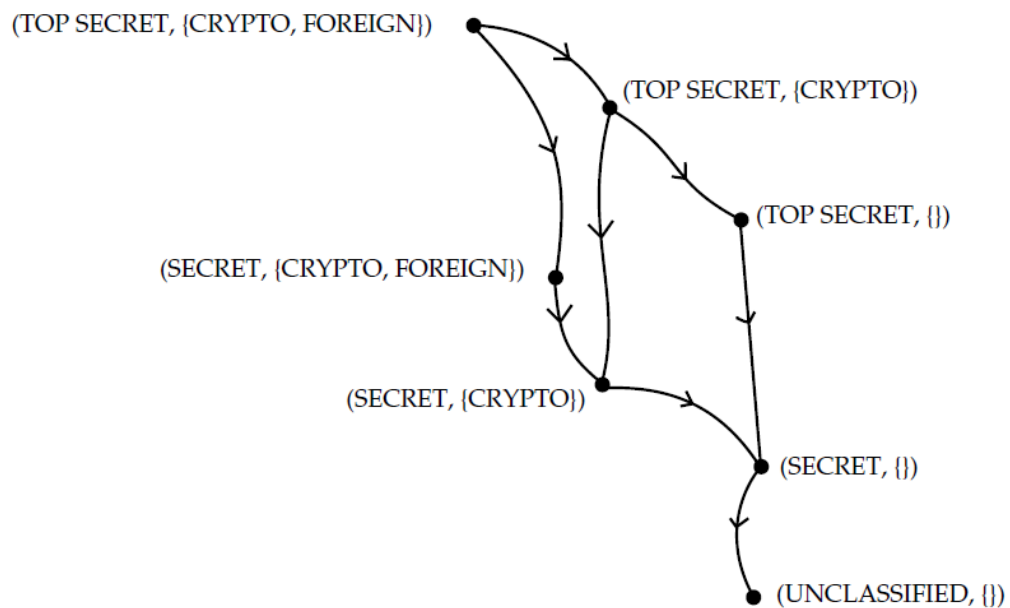


Figure 9.3: A lattice of security labels

Kenneth Walter, Walter Ogden, William Rounds, Frank Bradshaw, Stan Ames y David Shumway de la Universidad Case Western produjeron un modelo de lattice más avanzado, así como la resolución de muchos problemas con los atributos de archivos y directorios, los cuales transmitieron a Bell y LaPadula [1312, 1313]. Finalmente, el modelo de lattice fue sistematizado y popularizado por Denning [368].

La mayoría de los productos construidos para el mercado de seguridad multinivel se pueden reutilizar en modo compartimentado. Sin embargo, en la práctica, estos productos no son tan efectivos como se podría desear. Es fácil usar un sistema operativo multinivel para mantener los datos en diferentes compartimentos separados — simplemente dándoles etiquetas incompatibles (‘Secreto Tulipán’, ‘Secreto Narciso’, ‘Secreto Crocus’, etc.). Pero ahora el sistema operativo se ha convertido en un mecanismo de aislamiento, en lugar de un mecanismo de compartición; el verdadero problema es cómo controlar el intercambio de información.

Una solución es imponer límites superiores mínimos en el lattice usando algún algoritmo. Un ejemplo proviene del sistema utilizado por el gobierno de Arabia Saudita para gestionar el Haj, la peregrinación anual a La Meca [606]. Aunque la mayoría de los compartimentos son por defecto Confidenciales, la combinación de datos de diferentes compartimentos es Secreta. Así, ‘Visas-Haj’ y ‘Invitados-Gubernamentales’ son confidenciales, pero su combinación es Secreta.

En muchos sistemas de inteligencia, donde los usuarios ya operan al nivel más alto de autorización, los propietarios de los datos no desean un nivel de clasificación adicional en el que todo sea visible. Así, los datos derivados de dos compartimentos efectivamente crean un tercer compartimento utilizando el modelo de lattice. La proliferación de millones de compartimentos es compleja de gestionar y puede estar entrelazada con las aplicaciones. Por lo tanto, una solución más común es usar un producto multinivel estándar, como un guardián de correo, para asegurar que los correos

electrónicos ‘no confiables’ vayan a filtros. Pero ahora el núcleo de la base de computación confiable consiste en los filtros en lugar del guardián.

Peor aún, el guardián puede perder algunas de las funcionalidades más importantes del sistema operativo subyacente. Por ejemplo, el Guardián de Correo Estándar [1193] se construyó sobre un sistema operativo llamado LOCK cuyo mecanismo básico es la aplicación de tipos, como se describió en el capítulo anterior. Las versiones posteriores de LOCK soportan el control de acceso basado en roles, lo cual sería un mecanismo más apropiado para gestionar las relaciones entre compartimentos directamente [612]. Utilizarlo simplemente como una plataforma para soportar BLP puede haber sido un desperdicio.

En general, los problemas reales que enfrentan los usuarios de sistemas de inteligencia tienen que ver con la combinación de datos en diferentes compartimentos, y su degradación después saneamiento. Los modelos de seguridad multinivel y lattice ofrecen poca ayuda aquí. De hecho, uno de los mayores problemas que enfrenta la comunidad de inteligencia de EE. UU. desde el 9/11 es cómo manejar la búsqueda en sistemas con muchos compartimentos. Una búsqueda realizada en las bases de datos de muchas agencias puede arrojar resultados con muchas palabras clave adjuntas; si esto se agregara en un solo lugar, ese lugar tendría efectivamente todas las autorizaciones. Lo que hacen los nuevos sistemas es enviar consultas de búsqueda vinculadas con la autorización del usuario: "Muéstrame todo lo que coincida con Uzbekistán, Peshawar, armas y motocicleta, y que pueda ser visto por alguien con una autorización de Ultra Secreto Umbral". Aquí, las etiquetas locales solo entorpecen; pero sin ellas, ¿cómo se evita un futuro Aldrich Ames?

También hay un precedente preocupante en el caso del espía Walker. Allí, un intento de mantener los buques navales en compartimentos simplemente no funcionó, ya que un barco podría ser enviado a cualquier parte sin previo aviso, y para un barco estar aislado sin material de clave local era inaceptable operativamente. Así, los 800 barcos de la Marina de EE. UU. terminaron con el mismo conjunto de claves de cifrado, que fueron vendidas a los rusos [587].

9.2.2 El Muro Chino

El segundo modelo de seguridad multilateral es el modelo del Muro Chino, desarrollado por Brewer y Nash [224]. Su nombre proviene del hecho de que las firmas de servicios financieros, desde bancos de inversión hasta contadores, tienen reglas internas diseñadas para prevenir conflictos de interés, a las que llaman Muros Chinos. El alcance del modelo es más amplio que solo las finanzas. Hay muchas firmas profesionales y de servicios cuyos clientes pueden estar en competencia entre sí: los vendedores de software y las agencias de publicidad son otros ejemplos. Una regla típica es que "un socio que haya trabajado recientemente para una empresa en un sector comercial no puede ver los documentos de ninguna otra empresa en ese sector". Así, una vez que un redactor publicitario ha trabajado en (digamos) la cuenta de Shell, no se le permitirá trabajar en la cuenta de ninguna otra compañía petrolera durante un período de tiempo determinado. El modelo del Muro Chino presenta una combinación de elección libre y control de acceso obligatorio: un socio puede elegir para qué compañía petrolera trabajar, pero una vez que toma esa decisión, sus acciones en ese sector están

completamente restringidas. También introduce el concepto de separación de funciones en el control de acceso; un usuario dado puede realizar la transacción A o la transacción B, pero no ambas. Parte de la atracción del modelo del Muro Chino para la comunidad de investigación en seguridad proviene del hecho de que puede expresarse de una manera bastante similar a Bell-LaPadula. Si escribimos, para cada objeto c , $y(c)$ para la compañía de c y $x(c)$ para la clase de conflicto de intereses de c , entonces, como BLP, puede expresarse en dos propiedades:

- La propiedad de seguridad simple: un sujeto s tiene acceso a c si y solo si, para todos c , que s puede leer, ya sea $y(c) \notin x(c,)$ o $y(c) = y(c,)$
- La propiedad *: un sujeto s puede escribir en c solo si s no puede leer ningún c , con $x(c,) \neq 0$ y $y(c) \neq y(c,)$. El modelo del Muro Chino hizo una contribución seminal a la teoría del control de acceso. También provocó un debate sobre la medida en que es consistente con las propiedades de tranquilidad de BLP, y algo de trabajo sobre la semántica formal de tales sistemas (ver, por ejemplo, Foley [480] sobre la relación con la no interferencia). También hay algunas nuevas preguntas interesantes sobre canales encubiertos. Por ejemplo, ¿podría una compañía petrolera descubrir si un competidor que usa el mismo banco de inversión estaba planeando una oferta por una tercera compañía petrolera, preguntando qué especialistas estaban disponibles para consulta y notando que su número había disminuido repentinamente? En la práctica, sin embargo, los Muros Chinos todavía se implementan utilizando métodos manuales. Una gran consultora de software hace que cada uno de sus empleados mantenga un currículum vitae 'no clasificado' con entradas que han sido saneadas y acordadas con el cliente. Una entrada típica podría ser:
- Sep 97 — Abr 98: consultor sobre requisitos de seguridad para un nuevo sistema de contabilidad de sucursales para un importante banco minorista de EE. UU. Este no es el único control. El gerente de un consultor debe estar al tanto de los posibles conflictos y no enviar el CV al cliente si tiene dudas; si esto falla, el cliente puede detectar posibles conflictos él mismo a partir del CV; y si esto también falla, entonces el consultor tiene la obligación de informar cualquier posible conflicto tan pronto como aparezca.

9.2.3 El Modelo de la BMA

Quizás el ejemplo más importante, interesante e instructivo de seguridad multilateral se encuentra en los sistemas de información médica. El sector sanitario gasta una parte mucho mayor del ingreso nacional que el militar en los países desarrollados, y aunque los hospitales están menos automatizados, se están poniendo al día rápidamente. Un estudio de 2006 para el Departamento de Salud y Servicios Humanos de EE. UU. (DHHS) mostró que las inversiones en TI de salud se recuperaban en tres a trece años, y podían hacer que la atención médica fuera más segura y eficiente [1160]. La seguridad en la atención médica y (especialmente) la privacidad se han convertido en temas candentes en muchos países. En EE. UU., la Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA) fue aprobada por el Congreso en 1996 tras una serie de fallos de privacidad. En un caso notorio, Mark Farley, un violador de niños condenado que trabajaba como técnico ortopédico en el Hospital Newton-Wellesley en Newton, Massachusetts, fue atrapado usando la contraseña de un ex empleado para revisar los registros de 954 pacientes (principalmente mujeres jóvenes) para obtener los números

de teléfono de chicas a las que luego hizo llamadas obscenas [223]. Terminó en la cárcel, y el senador de Massachusetts Edward Kennedy fue uno de los patrocinadores de HIPAA.

Hay muchos más incidentes de naturaleza menos dramática. También en 1995-96, el gobierno del Reino Unido intentó centralizar todos los registros médicos, lo que llevó a una confrontación con la Asociación Médica Británica (BMA). La BMA me contrató para diseñar una política de seguridad y privacidad de la información clínica, que discutiré a continuación. La controversia continuó. A fines de la década de 1990, un proyecto en Islandia para construir una base de datos médica nacional que incorporara no solo registros médicos sino también datos genéticos y genealógicos, para que las enfermedades hereditarias pudieran rastrearse a través de generaciones, causó un alboroto. El once por ciento de la población optó por no participar; finalmente, la Corte Suprema de Islandia decidió que la base de datos debía ser de inclusión voluntaria en lugar de exclusión voluntaria, y ahora aproximadamente la mitad de la población participa. En 2002, el presidente Bush reescribió y relajó las regulaciones de HIPAA, conocidas como la 'Regla de Privacidad'; esto fue seguido por una mayor 'simplificación administrativa' en 2006. La situación en EE. UU. ahora es que, aunque los datos médicos aún deben protegerse en hospitales, clínicas y aseguradoras, su uso fuera del entorno de atención inmediata (por ejemplo, por investigadores, empleadores y agencias de bienestar) está fuera de las regulaciones y, por lo tanto, mucho menos controlado. Nadie está completamente feliz: los defensores de la privacidad de la salud consideran que el régimen es bastante inadecuado; los hospitales se quejan de que añade costos innecesarios; y los defensores de los pacientes señalan que HIPAA a menudo es utilizada por el personal del hospital como una excusa para ser poco serviciales [560]. Al momento de escribir esto (2007), el Hospital Piedmont de Atlanta acaba de convertirse en la primera institución en EE. UU. en ser auditada para el cumplimiento de las regulaciones de seguridad y privacidad, que entraron en vigor en 2005. Esta auditoría abarcó temas desde el acceso físico y lógico a sistemas y datos hasta el uso de Internet y violaciones de reglas de seguridad por parte de empleados, y ayudó a muchos otros proveedores de atención médica a decidir invertir en tecnologías de cifrado y otras protecciones [1295]. Además, la Oficina de Responsabilidad Gubernamental (GAO) acaba de informar que el DHHS necesita hacer mucho más para asegurar la privacidad del paciente, particularmente al definir una estrategia general para la privacidad y adoptar hitos para abordar el intercambio nacional de datos de salud (que no es solo una cuestión de protección técnica inadecuada, sino también de leyes estatales variables) [735]. En varios países europeos, ha habido debates sobre las compensaciones entre seguridad y privacidad involucradas con la información médica de emergencia. Los alemanes colocan datos como las recetas actuales y las alergias en la tarjeta de seguro médico que llevan los residentes; otros países han retenido esto, razonando que si los datos actualmente en un brazalete MedAlert legible por humanos, como las alergias, se trasladan a un dispositivo legible por máquina como una tarjeta inteligente, entonces hay un riesgo para los pacientes que se enferman en lugares donde no hay un lector disponible, como en un avión o en unas vacaciones en el extranjero. En el Reino Unido, el gobierno está creando un 'registro de atención resumido' de recetas y alergias que se mantendrá en una base de datos central y estará disponible para muchos trabajadores de la salud, desde clínicos de urgencias hasta paramédicos y operadores de servicios de línea de ayuda médica fuera del horario. Un problema es que los medicamentos actuales de un paciente a menudo revelan información altamente sensible, como el tratamiento

para el VIH, la depresión o el alcoholismo, y hacer que dicha información esté disponible para cientos de miles de personas conlleva riesgos sustanciales de abuso. Se ha ofrecido a los pacientes el derecho a optar por no participar en este sistema.

También ha habido debates sobre cuestiones de privacidad y ética relacionadas con los usos secundarios de la información médica, como en la investigación. En primer lugar, existen preocupaciones sobre fallos de privacidad, por ejemplo, cuando un profesor investigador pierde una computadora portátil que contiene los registros de millones de pacientes. Aunque los registros utilizados en la investigación a menudo tienen nombres y direcciones eliminados, es un trabajo muy difícil desidentificar los registros adecuadamente; discutiré esto en detalle a continuación. En segundo lugar, hay cuestiones éticas relacionadas con el consentimiento. Por ejemplo, una mujer católica devota podría oponerse a que sus datos ginecológicos se utilicen para desarrollar una mejor píldora del día después. En tercer lugar, hay cuestiones económicas; si mis datos se utilizan para desarrollar un medicamento del cual una empresa gana miles de millones de dólares, ¿no debería recibir una parte? La protección de la información médica es, por lo tanto, un caso histórico interesante para el ingeniero de seguridad. Tiene muchas compensaciones ricas y complejas; es importante para todos nosotros; y está frecuentemente en las noticias. La privacidad médica también es un modelo para proteger otros tipos de información personal, como la información que tienen las empresas y agencias gubernamentales sobre clientes individuales. En todos los países europeos (y en muchos otros, como Canadá y Australia) hay leyes de protección de datos que restringen la difusión de dichos datos. Discutiré la ley de protección de datos en la Parte III; para los propósitos actuales, es suficiente notar que algunas clases de datos (afectando a la salud, el comportamiento sexual, la actividad política y la creencia religiosa) el sujeto de los datos debe dar su consentimiento para compartir la información, o tener derecho de veto, o debe haber una ley específica que permita compartirla por el interés público en circunstancias que estén lo suficientemente definidas como para que el sujeto de los datos pueda preverlas. Esto plantea la cuestión de cómo se puede construir una política de seguridad en la que las decisiones de control de acceso no las tome una autoridad central (como en Bell-LaPadula) o los usuarios del sistema (como en el control de acceso discrecional), sino los sujetos de los datos. Primero veamos los aspectos del control de acceso.

9.2.3.1 El Modelo de Amenaza

La principal amenaza a la privacidad médica es el abuso del acceso autorizado por parte de los internos, y el vector de amenaza más común es la ingeniería social. El ataque típico proviene de un detective privado que llama a la oficina de un médico o aseguradora de salud con una historia plausible:

"Hola, soy el Dr. Burnett del departamento de cardiología del Hospital Conquest en Hastings. Su paciente Sam Simmonds acaba de ser admitido aquí en coma, y tiene una arritmia ventricular de aspecto extraño. ¿Puede decirme si hay algo relevante en su expediente?"

Este tipo de ataque suele ser tan exitoso que tanto en los EE. UU. como en el Reino Unido hay personas que se ganan la vida haciéndolo [411]. (No está restringido a los registros médicos; en junio de 2000, el recaudador de fondos de Tony Blair, Lord Levy,

quedó muy avergonzado después de que alguien llamara a la oficina de impuestos fingiendo ser él y descubriera que solo había pagado £5000 en impuestos el año anterior [1064]. Pero el contexto médico es bueno para discutirlo).

Como mencioné brevemente en el Capítulo 2, en 1996 se realizó un experimento en el Reino Unido en el que el personal de una autoridad sanitaria (una aseguradora gubernamental que compra atención médica para un distrito de varios cientos de miles de personas) fue capacitado para filtrar llamadas telefónicas falsas. El consejo que se les dio se describe en [36], pero el elemento más importante era que siempre debían devolver la llamada, y no a un número dado por el llamador, sino al número en la guía telefónica del hospital u otra institución donde el llamador afirmaba trabajar. Resultó que unas treinta consultas telefónicas a la semana eran falsas.

Estas medidas de seguridad operativas son mucho más importantes que la mayoría de las medidas de protección técnica, pero son difíciles. Si todos fueran tan poco serviciales como el personal de las agencias de inteligencia está entrenado para ser, el mundo se detendría. Y la mejor capacitación del personal del mundo no protegerá un sistema donde demasiadas personas ven demasiados datos. Siempre habrá personal que sea descuidado o incluso deshonesto; y cuántos más registros puedan obtener, más daño pueden hacer. Además, las organizaciones tienen culturas establecidas; simplemente no hemos podido incorporar medidas de seguridad operativa, incluso ligeras, a gran escala en la atención médica, simplemente porque así no es como trabaja la gente. El personal está enfocado en brindar atención en lugar de cuestionarse entre sí. Las pocas mejoras operativas reales en los últimos años han seguido a sustos; por ejemplo, las unidades de maternidad en Gran Bretaña ahora tienen controles de entrada razonables, después de incidentes en los que se robaron bebés de las nurserías. Además, las salas geriátricas a menudo están cerradas para evitar que los pacientes dementes se escapen. Sin embargo, la mayoría de las salas de los hospitales están completamente abiertas; cualquiera puede entrar desde la calle para visitar a sus familiares, y los beneficios clínicos de las visitas frecuentes superan los incidentes violentos ocasionales. Las PC se dejan desatendidas y con sesión iniciada en la red del hospital. Recientemente, un programa de inversión en TI de salud en el Reino Unido ha intentado estandarizar el control de acceso y emitir tarjetas inteligentes al personal clínico para iniciar sesión en los sistemas hospitalarios; pero dado que cerrar sesión como la Enfermera Jones e iniciar sesión nuevamente como la Enfermera Smith lleva varios segundos, el personal no se molesta.

Un problema más general es que incluso donde el personal se comporta éticamente, la falta de comprensión técnica, o lo que podríamos describir más propiamente como la mala usabilidad de la seguridad, provoca filtraciones de información personal. Las PC antiguas vendidas en el mercado de segunda mano o donadas a escuelas a menudo tienen datos recuperables en el disco duro; la mayoría de las personas no son conscientes de que el comando de 'eliminar' habitual no elimina el archivo, sino que simplemente marca el espacio que ocupa como reutilizable. Una PC vendida en el mercado de segunda mano por el banco de inversión Morgan Grenfell Asset Management tenía archivos recuperables que contenían las transacciones financieras del ex-Beatle Paul McCartney [254]; ha habido problemas similares con registros de salud antiguos. También se roba equipo: aproximadamente el 11% de los médicos de familia del Reino Unido han experimentado el robo de una PC de la consulta, y en un caso dos prominentes damas de la sociedad fueron chantajeadas por interrupciones de embarazo

después de tal robo [37]. La respuesta del gobierno del Reino Unido a esta amenaza es tratar de persuadir a los médicos de familia para que se trasladen a sistemas 'alojados', donde los datos de la consulta se mantienen en granjas de servidores regionales; pero no está claro que haya una ganancia neta en privacidad. El robo de datos puede ser más difícil, pero una vez centralizados los datos, se puede esperar un aumento en el acceso; cada vez más agencias públicas argumentarán por qué necesitan acceso a los datos. Incluso si todos los casos de acceso son individualmente sólidos, el efecto neto con el tiempo puede ser bastante destructivo para la privacidad.

El problema fundamental es este. La probabilidad de que un recurso sea abusado depende de su valor y del número de personas que tienen acceso a él. Agregar información personal en grandes bases de datos aumenta ambos factores de riesgo al mismo tiempo. En términos simples, podemos vivir con una situación en la que el recepcionista de un médico tiene acceso a los registros de 2,000 pacientes: habrá abusos de vez en cuando, pero a un nivel tolerablemente bajo. Sin embargo, si los recepcionistas de los 5,000 médicos de familia que podrían trabajar con una gran HMO estadounidense, o en una de las cinco regiones del Servicio Nacional de Salud de Inglaterra, tienen acceso a los registros de tal vez diez millones de pacientes, entonces el abuso se vuelve probable. Solo se necesita un interno que aprenda a acercarse a una PC que esté registrada con la tarjeta inteligente de otra persona, leer un archivo y pasar la información a un detective privado a cambio de dinero. No solo son los médicos; en Inglaterra, cada región tiene decenas de miles de personas con acceso, desde enfermeras y programadores y recepcionistas hasta conductores, proveedores de alimentos y limpiadores. Muchos del personal son temporales, muchos son extranjeros y muchos ganan cerca del salario mínimo. Y los problemas de privacidad no se limitan a las organizaciones que tratan directamente a los pacientes: algunas de las mayores colecciones de información de salud personal están en manos de aseguradoras de salud y organizaciones de investigación. Discutiré sus problemas especiales a continuación en la sección 9.3.

En un entorno como este, se requieren controles de flujo de información lateral. Un buen ejemplo de lo que puede salir mal sin ellos proviene de un sistema hospitalario temprano en el Reino Unido cuyos diseñadores creían que por razones de seguridad, todo el personal debería tener acceso a todos los registros. Esta decisión fue influenciada por el cabildeo de geriatras y pediatras, cuyos pacientes a menudo son tratados por varios departamentos especializados en el hospital. Se frustraron por las incompatibilidades entre los diferentes sistemas departamentales. El sistema se implementó en 1995 en Hampshire, donde el entonces ministro de salud Gerry Malone tenía su escaño parlamentario. El sistema hacía que todas las pruebas de laboratorio realizadas para los médicos locales en el laboratorio de patología del hospital fueran visibles para la mayor parte del personal del hospital. Una enfermera que había realizado una prueba con su médico de familia se quejó con él después de encontrar el resultado en el sistema hospitalario de Basingstoke donde trabajaba; esto causó indignación entre los médicos locales, y Malone perdió su escaño en el Parlamento en las elecciones de 1997 (por dos votos) [46].

Entonces, ¿cómo podemos evitar que todos vean todos los registros? Hay muchas cosas ad hoc que se pueden hacer: una medida bastante efectiva es mantener los registros de los pacientes anteriores en un archivo separado y dar a solo un pequeño número de

personal de admisiones el poder de mover registros de allí al sistema principal. Otra es introducir una trampa: un hospital de Boston tiene en su sistema algunos 'registros médicos' falsos con los nombres de los miembros de la familia Kennedy, para poder identificar y disciplinar al personal que los consulta. Una propuesta particularmente ingeniosa, debida a Gus Simmons, es investigar a todo el personal que consulta un registro de paciente pero no presenta una reclamación de pago al asegurador dentro de los treinta días; esto alinea el interés del paciente en la privacidad con el interés del hospital en maximizar sus ingresos.

Sin embargo, un mosaico de medidas ad hoc no es una buena manera de asegurar un sistema. Necesitamos una política de control de acceso adecuada, pensada desde los primeros principios y basada en un modelo realista de las amenazas. ¿Qué política es apropiada para la atención médica?

9.2.3.2 La Política de Seguridad

Esta pregunta enfrentó la BMA en 1995. El gobierno del Reino Unido había introducido una estrategia de TI para el Servicio Nacional de Salud que implicaba centralizar muchos datos en servidores centrales y cuya política de seguridad era multinivel: la idea era que las bases de datos de SIDA estarían a un nivel correspondiente a Secreto, los registros de pacientes normales a Confidencial y los datos administrativos como recetas de medicamentos y facturas de tratamiento a Restringido. Pronto se dieron cuenta de que esto no iba a funcionar. Por ejemplo, ¿cómo debería clasificarse una receta de AZT? Como es una receta de medicamento, debería ser Restringida; pero como identifica a una persona como VIH positiva, debe ser Secreta. Así que todas las recetas de AZT 'Secretas' deben ser eliminadas del archivo 'Restringido' de recetas de medicamentos. Pero entonces casi todas las demás recetas también, ya que identifican tratamientos para individuos nombrados y por lo tanto deberían ser 'Confidenciales'. Pero entonces, ¿de qué servirá el archivo de recetas para alguien?

Un segundo problema es que la estrategia se basaba en la idea de un único registro electrónico de pacientes (EPR) que seguiría al paciente desde la concepción hasta la autopsia, en lugar del sistema tradicional de tener diferentes registros sobre el mismo paciente en diferentes hospitales y consultorios médicos, con información que fluye entre ellos en forma de cartas de referencia y de alta. Un intento de diseñar una política de seguridad para el EPR, que observara las normas éticas existentes, rápidamente se volvió incontrolablemente complejo [558]. En un proyecto del cual fui responsable, la BMA desarrolló una política de seguridad para llenar el vacío. La innovación crítica fue definir el registro médico no como el total de todos los hechos clínicos relacionados con un paciente, sino como el conjunto máximo de hechos relacionados con un paciente y al cual el mismo personal tenía acceso. Así, un paciente individual tendrá más de un registro, y esto ofendió a los defensores 'puristas' del EPR. Pero múltiples registros son dictados de todos modos por la ley y la práctica. Dependiendo del país (e incluso del estado) en el que te encuentres, es posible que tengas que mantener registros médicos separados para la fertilización humana, enfermedades de transmisión sexual, servicios médicos en prisiones e incluso registros de nacimientos (ya que se refieren a la salud de la madre así como del niño, y no pueden ser simplemente entregados al niño más tarde sin violar la confidencialidad de la madre). Esta situación probablemente se volverá aún

más compleja a medida que los datos genéticos comiencen a ser utilizados más ampliamente.

En muchos países, incluidos todos los signatarios de la Convención Europea de Derechos Humanos, se otorga un estatus especial al consentimiento del paciente tanto en la ley como en la ética médica. Los registros solo pueden ser compartidos con terceros si el paciente lo aprueba, o en un rango limitado de excepciones legales, como rastrear contactos de personas con enfermedades infecciosas como la tuberculosis. Las definiciones son ligeramente fluidas; en algunos países, la infección por VIH es notificable, en otros no, y en otros los datos se recopilan sigilosamente.

Por lo tanto, los objetivos de la política de seguridad de la BMA eran hacer cumplir el principio del consentimiento del paciente y evitar que demasiadas personas tuvieran acceso a demasiados registros identificables. No intentó hacer nada nuevo, sino simplemente codificar las mejores prácticas existentes. También buscó expresar otras características de seguridad de la gestión de registros médicos, como la seguridad y la responsabilidad. Por ejemplo, debe ser posible reconstruir el contenido del registro en cualquier momento en el pasado, para que, por ejemplo, si se presenta una demanda por negligencia, el tribunal pueda determinar qué información estaba disponible para el médico en ese momento. Los detalles del análisis de requisitos están en [37].

La política consta de nueve principios:

1. Control de acceso: cada registro clínico identificable deberá estar marcado con una lista de control de acceso que nombre a las personas o grupos de personas que pueden leerlo y agregar datos a él. El sistema deberá impedir que cualquier persona que no esté en la lista de control de acceso acceda al registro de cualquier manera.
2. Apertura de registro: un médico puede abrir un registro con ella misma y el paciente en la lista de control de acceso. Cuando un paciente ha sido derivado, puede abrir un registro con ella misma, el paciente y el(los) médico(s) remitente(s) en la lista de control de acceso.
3. Control: Uno de los médicos en la lista de control de acceso debe ser marcado como responsable. Solo ella puede alterar la lista de control de acceso, y solo puede agregar a otros profesionales de la salud a ella.
4. Consentimiento y notificación: la médico responsable debe notificar al paciente los nombres en la lista de control de acceso de su registro cuando se abre, de todas las adiciones posteriores y cada vez que se transfiera la responsabilidad. También debe obtener su consentimiento, excepto en casos de emergencia o en el caso de excepciones legales.
5. Persistencia: nadie tendrá la capacidad de eliminar información clínica hasta que haya expirado el período de tiempo apropiado.
6. Atribución: todos los accesos a registros clínicos deben estar marcados en el registro con el nombre del sujeto, así como la fecha y hora. También debe mantenerse un registro de auditoría de todas las eliminaciones.
7. Flujo de información: la información derivada del registro A puede agregarse al registro B si y solo si la lista de control de acceso de B está contenida en la de A.
8. Control de agregación: deberán existir medidas efectivas para prevenir la agregación de información de salud personal. En particular, los pacientes deben

recibir una notificación especial si se propone agregar a una persona a su lista de control de acceso que ya tiene acceso a información de salud personal de un gran número de personas.

9. Base informática confiable: los sistemas informáticos que manejen información de salud personal deberán tener un subsistema que haga cumplir los principios anteriores de manera efectiva. Su efectividad deberá ser evaluada por expertos independientes.

Esta política puede parecer solo sentido común, pero es sorprendentemente completa y radical en términos técnicos. Por ejemplo, es estrictamente más expresiva que el modelo Bell-LaPadula del capítulo anterior; contiene un mecanismo de control de flujo de información tipo BLP en el principio 7, pero también contiene estado. (Una discusión más completa desde el punto de vista del control de acceso, y para una audiencia técnica, se puede encontrar en [38]).

Políticas similares fueron desarrolladas por otros cuerpos médicos, incluidas las asociaciones médicas suecas y alemanas; la Asociación de Informática de Salud de Canadá y un proyecto de la UE (estos se describen en [732]). Sin embargo, el modelo de la BMA es el más detallado y ha sido sometido a la revisión más rigurosa; fue adoptado por la Unión de Organizaciones Médicas Europeas (UEMO) en 1996. Los comentarios de la consulta pública sobre la política se pueden encontrar en [39].

Implementaciones Piloto

En un enfoque de arriba hacia abajo para la ingeniería de seguridad, primero se debe determinar el modelo de amenaza, luego escribir la política, y finalmente probar la política observando si funciona en la vida real.

Los sistemas compatibles con la BMA ahora se han implementado tanto en la práctica general [585] como en un sistema hospitalario desarrollado en Hastings, Inglaterra, que aplica reglas de acceso similares utilizando una combinación de roles y capacidades. Tiene reglas como "una enfermera de sala puede ver los registros de todos los pacientes que hayan estado en su sala en los últimos 90 días", "un médico puede ver los registros de todos los pacientes que hayan sido tratados en su departamento" y "un médico senior puede ver los registros de todos los pacientes, pero si accede al registro de un paciente que nunca ha sido tratado en su departamento, el médico senior responsable del cuidado de ese paciente será notificado". (El sistema hospitalario fue inicialmente diseñado independientemente del proyecto de la BMA. Cuando nos enteramos el uno del otro, nos sorprendió lo mucho que coincidían nuestros enfoques, y nos tranquilizó saber que habíamos capturado las expectativas de la profesión de una manera razonablemente precisa).

Las lecciones aprendidas se discuten en [366, 367, 585]. Una de ellas fue la dificultad de construir una base informática confiable pequeña. El sistema de registros hospitalarios tiene que depender del sistema administrativo de pacientes para decirle qué pacientes y qué enfermeras están en qué sala. Un sistema prototipo diferente en un hospital en Cambridge, Inglaterra, proporcionaba al personal certificados en tarjetas inteligentes que utilizaban para iniciar sesión.

9.2.4 Cuestiones Actuales de Privacidad

En 2002, el Primer Ministro Tony Blair fue persuadido para asignar £6 mil millones para modernizar la informática del servicio de salud en Inglaterra. Esto llevó a una competencia por los contratos con la seguridad como algo secundario. La visión original era de comunicaciones mucho mejores en cada comunidad de salud local; de modo que si un paciente diabético estaba siendo atendido por un médico de familia, un diabetólogo hospitalario, una enfermera comunitaria y un optometrista, todos podrían ver las notas y los resultados de las pruebas de los demás. La propia paciente también podría subir datos como niveles de glucosa en sangre, ver sus notas médicas y participar en su atención. Esta visión se había desarrollado en el Wirral cerca de Liverpool.

Cuando se asentó el polvo del proceso de contratación, la visión de empoderamiento local había sido reemplazada por un enfoque mucho más centralizado. Se adjudicaron contratos para cinco regiones, cada una con unos 10 millones de personas, llamando a reemplazar todos los sistemas hospitalarios durante 2004-2010 con sistemas estándar. El número de proveedores de sistemas se redujo a dos: Cerner e iSoft, y la política de seguridad ha sido objeto de mucho debate. La política actual es para tres mecanismos principales.

1. El caballo de batalla del control de acceso será el control de acceso basado en roles, similar a los pioneros en Hastings, pero mucho más complejo; en lugar de una docena de roles, ahora el plan es que haya más de trescientos.
2. Para acceder a los datos del paciente, un miembro del personal también necesitará una relación legítima. Esta es una abstracción de la idea de Hastings de "su departamento".
3. Por defecto, cada paciente tiene un único registro electrónico de paciente. Sin embargo, los pacientes también podrán declarar que ciertas partes de sus registros están "selladas" o "selladas y bloqueadas". En el último caso, los registros solo serán visibles para un equipo de atención en particular. En el primero, su existencia será visible para otro personal que consulte el registro del paciente, y podrán romper el sello en una emergencia.

Las implementaciones iniciales han planteado una serie de problemas detallados. Por ejemplo, los pacientes que reciben atención psiquiátrica ambulatoria en un hospital solían tener sus notas en papel en el archivo del psiquiatra; todo lo que la recepcionista sabía era que la Sra. Smith era vista una vez al mes por el Dr. Jones. Ahora, sin embargo, la recepcionista puede ver las notas también. Su rol tenía que tener acceso a los registros de los pacientes para que pudiera ver y modificar datos administrativos como los horarios de las citas; y si está trabajando en la recepción en el ala del hospital donde el Dr. Jones tiene su consultorio, entonces tiene una relación legítima. El sellado y bloqueo de registros aún no se ha implementado. Por lo tanto, tiene acceso a todo. Este es un buen ejemplo de por qué la doctrina del "EPR" de un registro por paciente fue una mala idea, y la visión de la BMA de múltiples registros vinculados era mejor; ahora parece que todos los registros en psiquiatría, salud sexual, etc., pueden tener que ser sellados (o incluso sellados y bloqueados) por defecto. Entonces, la atención de estos pacientes a través de diferentes departamentos comenzará a causar problemas. Al igual que con los sistemas de seguridad multinivel, lo difícil no es tanto separar sistemas, sino gestionar los flujos de información entre niveles o compartimentos.

Quizás los problemas más difíciles con los nuevos sistemas ingleses, sin embargo, se refieren al consentimiento del paciente. El servicio de salud está permitiendo que las personas opten por no participar en el registro de atención resumida, la base de datos central de información médica de emergencia, que contiene cosas como medicamentos, alergias e historial médico importante. Esto no es un gran problema; la mayoría de la gente no tiene nada estigmatizante allí. (De hecho, la mayoría de las personas menores de la edad de jubilación no tienen condiciones crónicas significativas y podrían hacerlo perfectamente bien sin un registro resumido). El problema mayor es que los nuevos sistemas hospitalarios harán que los registros detallados estén disponibles para terceros como nunca antes, para la investigación, la gestión del servicio de salud e incluso la aplicación de la ley.

Anteriormente, su privacidad médica estaba protegida por el hecho de que un hospital podría haber tenido más de setenta sistemas de registros departamentales diferentes, mientras que sus registros en su médico de familia estaban protegidos al estar en parte en papel y en parte en una PC que se apagaba a las seis cada noche y a la que los forasteros no tenían acceso. Una vez que todo está en sistemas estándar en una granja de servidores de salud regional, el juego cambia. Anteriormente, un policía que quería ver sus registros médicos necesitaba persuadir a un juez de que tenía motivos razonables para creer que encontraría evidencia real de un crimen; luego tenía que llevar la orden a su médico de familia o al director médico de su hospital. Los costos de este procedimiento aseguraban que se invocara solo raramente, y en casos como terrorismo, asesinato o violación. Sin embargo, una granja de servidores es un objetivo mucho más fácil, y si contiene datos de todos los que han confesado el uso ilegal de drogas a su médico, es un objetivo tentador. De hecho, desde junio de 2007, se supone que todos los médicos del Reino Unido deben completar un "perfil de resultados de tratamiento" para los usuarios de drogas, preguntándoles si han cometido algún delito en las últimas cuatro semanas, incluidos robos, asaltos y venta de drogas. Es difícil creer que esta información no acabará eventualmente en manos de la policía. Pero, ¿cuáles son las consecuencias para la salud pública cuando las personas ya no pueden confiar en sus médicos, especialmente los miembros más vulnerables y marginados de la sociedad? Ya tenemos casos de inmigrantes con tuberculosis que se fugan, ya que los datos demográficos del servicio de salud comenzaron a ser utilizados para encontrar inmigrantes ilegales.

Así que, incluso si la política de seguridad en sistemas centralizados equivale a una implementación fiel de la política de la BMA, con la excepción del octavo principio de no agregación, podemos esperar problemas. Hay algunos aspectos de la política de seguridad que simplemente no escalan. Crear grandes bases de datos de información personal sensible es intrínsecamente peligroso. Aumenta la motivación para el abuso y la oportunidad de abuso al mismo tiempo. Y, aunque los controles funcionen perfectamente para prevenir el abuso ilegal (ya sea por forasteros o internos), la existencia de tales bases de datos puede llevar al abuso legal, intereses poderosos en la sociedad presionan por y logran el acceso a datos en una escala y de un tipo que las personas sensatas no permitirían.

Hay algunas ventajas en los sistemas centrales estándar. En los EE. UU., la Administración de Veteranos opera tales sistemas para su red hospitalaria; después del huracán Katrina, los veteranos de Luisiana que terminaron como refugiados en Texas o

Florida, o incluso Minnesota, podían ir directamente a los hospitales locales de VA y encontrar sus notas allí en la punta de los dedos del médico. Los pacientes de muchos otros hospitales y clínicas en Nueva Orleans perdieron sus notas por completo. Pero la centralización definitivamente puede dañar la privacidad. En mayo de 2006, la información personal de los 26.5 millones de veteranos de EE. UU., incluidos nombres, números de seguro social y, en algunos casos, discapacidades, fue robada de la residencia de un empleado del Departamento de Asuntos de Veteranos que había llevado los datos a casa sin autorización. Y no es suficiente simplemente compartimentar los propios registros médicos: en los Países Bajos, que han evitado cuidadosamente la centralización de registros, todavía hay una base de datos "Vecozo" que contiene detalles del seguro médico de los ciudadanos, y casi 80,000 personas tenían acceso a ella, desde médicos y farmacéuticos hasta curanderos alternativos e incluso empresas de taxis. Hubo un escándalo cuando los periodistas encontraron que era fácil obtener las direcciones privadas y los números de teléfono no listados de varios políticos, criminales y personalidades famosos [126]. (Después de que estalló el escándalo, los aseguradores y su operador de base de datos intentaron culparse mutuamente: ninguno aceptaría la responsabilidad del hecho de que se hacía disponible demasiada información a demasiadas personas).

Por lo tanto, si se toma una decisión política para tener una gran base de datos centralizada, el problema de la agregación perseguirá el diseño detallado y la operación continua: incluso si algunas personas (o aplicaciones) tienen permiso para ver todo, es una muy mala idea no controlar a los principales que realmente lo hacen. Si encuentra que la mayoría de los médicos de su hospital ven unos pocos miles de los varios millones de registros en la base de datos, y uno los ve todos, ¿qué le dice eso? Es mejor que lo averigüe². Pero muchos sistemas implementados no tienen controles de tasa o alarmas efectivas, y, aunque existan alarmas, a menudo no se actúa sobre ellas. Nuevamente, en el Reino Unido, más de 50 empleados del hospital revisaron los registros de una personalidad del fútbol en el hospital, a pesar de no estar involucrados en su atención, y ninguno de ellos fue disciplinado.

Y, aparte de los usos controvertidos de los registros médicos, como el acceso policial, hay serios problemas para proteger usos relativamente no controvertidos, como la investigación. A eso me referiré a continuación.

9.3 Control de Inferencias

El control de acceso en los sistemas de registros médicos es bastante difícil en hospitales y clínicas que atienden directamente a los pacientes. Es mucho más difícil asegurar la privacidad del paciente en aplicaciones secundarias como bases de datos para investigación, control de costos y auditoría clínica. Este es un aspecto en el que los médicos tienen más dificultades para proteger sus datos que los abogados; los abogados pueden bloquear sus archivos confidenciales de clientes y nunca dejar que ningún extraño los vea, mientras que los médicos están bajo todo tipo de presiones para compartir datos con terceros.

9.3.1 Problemas Básicos del Control de Inferencias en Medicina

La forma estándar de proteger los registros médicos utilizados en la investigación es eliminar los nombres y direcciones de los pacientes y, por lo tanto, hacerlos anónimos. De hecho, los defensores de la privacidad a menudo hablan de "Tecnologías de Mejora de la Privacidad" (PET) y la desidentificación es un ejemplo frecuentemente citado. Pero esto rara vez es infalible. Si una base de datos permite consultas detalladas, los individuos aún pueden ser identificados generalmente, y esto es especialmente cierto si se puede vincular información sobre diferentes episodios clínicos. Por ejemplo, si estoy tratando de averiguar si un político nacido el 2 de junio de 1946 y tratado por una fractura de clavícula después de un partido de fútbol universitario el 8 de mayo de 1967, ha sido tratado desde entonces por problemas de drogas o alcohol, y puedo hacer una consulta en esas dos fechas, entonces probablemente podría sacar su registro de una base de datos nacional. Incluso si la fecha de nacimiento se reemplaza por un año de nacimiento, aún es probable que pueda comprometer la privacidad del paciente si los registros son detallados o si se pueden vincular los registros de diferentes individuos. Por ejemplo, una consulta como "muéstrame los registros de todas las mujeres de 36 años con hijas de 14 y 16 años, de modo que la madre y exactamente una hija tengan psoriasis" también es probable que encuentre a un individuo entre millones. Y las consultas complejas con muchas condiciones son precisamente las que los investigadores quieren hacer.

Por esta razón, la Administración de Financiamiento de Atención Médica de EE. UU. (HCFA), que paga a médicos y hospitales por los tratamientos proporcionados bajo el programa Medicare, mantiene tres conjuntos de registros. Hay registros completos, utilizados para facturación. Hay registros cifrados para beneficiarios, con solo los nombres y números de seguro social de los pacientes oscurecidos. Estos aún se consideran datos personales (ya que aún tienen fechas de nacimiento, códigos postales, etc.) y solo son utilizables por investigadores confiables. Finalmente, hay registros de acceso público que han sido despojados de identificadores hasta el nivel en que los pacientes solo se identifican en términos generales como "una mujer blanca de 70-74 años que vive en Vermont". No obstante, los investigadores han encontrado que muchos pacientes aún pueden ser identificados mediante la correlación cruzada de los registros de acceso público con bases de datos comerciales, y tras las quejas de los defensores de la privacidad, un informe de la Oficina General de Contabilidad criticó a HCFA por laxitud en la seguridad [520].

La ley de EE. UU., que se enmarca en la regla de privacidad HIPAA, ahora reconoce la información desidentificada como datos médicos que han sido desidentificados "adecuadamente". Esto significa que se han eliminado 18 identificadores específicos y el operador de la base de datos no tiene conocimiento real de que la información restante pueda usarse sola o en combinación con otros datos para identificar al sujeto; o que un estadístico calificado concluya que el riesgo es sustancialmente limitado. Donde dichos datos son inadecuados para la investigación, también reconoce conjuntos de datos limitados que contienen más información, pero donde los usuarios están obligados por medidas contractuales y técnicas a proteger la información y no intentar reidentificar a los sujetos.

Muchos otros países tienen sistemas de monitoreo de atención médica que utilizan enfoques similares. Alemania tiene leyes de privacidad muy estrictas y sigue la ruta de la "información desidentificada"; la caída del Muro de Berlín obligó a los registros de cáncer de la antigua Alemania Oriental a instalar mecanismos de protección rápidamente [192]. Nueva Zelanda sigue el enfoque de "conjuntos de datos limitados" con una base de datos nacional de registros médicos cifrados para beneficiarios; el acceso está restringido a un pequeño número de estadísticos médicos especialmente autorizados, y no se responde a ninguna consulta con respecto a menos de seis registros [955]. En Suiza, algunos sistemas de investigación fueron reemplazados a insistencia de los reguladores de privacidad [1137].

En otros países, la protección ha sido menos adecuada. El Servicio Nacional de Salud de Gran Bretaña construyó una serie de bases de datos centralizadas en la década de 1990 que hacen que la información de salud personal esté ampliamente disponible dentro del gobierno y que llevaron a una confrontación con los médicos. El gobierno estableció un comité para investigar bajo Dame Fiona Caldicott; su informe identificó más de sesenta flujos de información ilegales dentro del servicio de salud [46, 252]. Algunos conjuntos de datos de investigación fueron desidentificados; otros (incluidos los datos sobre personas con VIH/SIDA) fueron reidentificados posteriormente, de modo que personas y organizaciones benéficas contra el VIH cuyos datos se habían recopilado bajo una promesa de anonimato fueron engañados. Luego, el Parlamento aprobó una ley que otorgaba a los ministros el poder de regular los usos secundarios de los datos médicos. Los datos mantenidos para usos secundarios se mantienen con código postal más fecha de nacimiento, y como los códigos postales del Reino Unido son compartidos por un máximo de unas pocas docenas de casas, esto significa que la mayoría de los registros son fácilmente identificables. Esto sigue siendo motivo de controversia. En 2007, el Comité Selecto de Salud del Parlamento llevó a cabo una investigación sobre el Registro Electrónico de Pacientes y escuchó testimonios de una amplia gama de puntos de vista: desde investigadores que creían que la ley debería obligar a compartir información para la investigación, hasta médicos, abogados de derechos humanos y defensores de la privacidad que argumentaban que solo debería haber las excepciones más estrechas a la privacidad médica.³ El Comité hizo muchas recomendaciones, incluida la de que se debería permitir a los pacientes evitar el uso de sus datos en la investigación [624]. El Gobierno rechazó esto.

Lo más controvertido de todo fue una base de datos genética en Islandia, que discutiré con más detalle a continuación.

Eliminar información personal es importante en muchos otros campos. Bajo el rubro de la Tecnología de Mejora de la Privacidad (PET) se ha promovido recientemente por reguladores en Europa y Canadá como un mecanismo general de privacidad [447]. Pero, como muestran los ejemplos médicos, puede haber una seria tensión entre el deseo de los investigadores de datos detallados y el derecho de los pacientes (u otros sujetos de datos) a la privacidad. La anonimización es mucho más frágil de lo que parece; y cuando falla, las empresas y las personas que confiaron en ella pueden sufrir consecuencias graves.

AOL enfrentó una tormenta de protestas en 2006 cuando lanzó los registros supuestamente anónimos de 20 millones de consultas de búsqueda realizadas durante

tres meses por 657,000 personas. Los nombres y direcciones IP de los buscadores fueron reemplazados por números, pero eso no ayudó. Los periodistas de investigación revisaron las búsquedas e identificaron rápidamente a algunos de los buscadores, que se sorprendieron por la violación de la privacidad [116]. Estos datos fueron publicados "con fines de investigación": la filtración llevó a la presentación de quejas ante la FTC, tras lo cual el CTO de la empresa renunció y la empresa despidió tanto al empleado que liberó los datos como al supervisor del empleado.

Otro ejemplo es en la privacidad de las películas. La empresa de alquiler de DVD Netflix envía más de un millón de DVD al día a más de 6 millones de clientes en EE. UU., tiene un sistema de calificación para combinar películas con clientes, y publicó las calificaciones de los espectadores de 500,000 suscriptores sin sus nombres. (Ofrecieron un premio de \$1 millón por un mejor algoritmo de recomendación). En noviembre de 2007, Arvind Narayanan y Vitaly Shmatikov mostraron que muchos suscriptores podían ser reidentificados al comparar los registros anónimos con preferencias expresadas públicamente en la Base de Datos de Películas en Internet [928]. Esto se debe en parte al efecto de "larga cola": una vez que se descartan las 100 o más películas que todos ven, las preferencias de visualización de las personas son bastante únicas. De todos modos, la ley de EE. UU. protege la privacidad de los alquileres de películas, y el ataque fue una vergüenza seria para Netflix.

Por lo tanto, es importante entender qué se puede y qué no se puede lograr con esta tecnología.

Otras Aplicaciones del Control de Inferencias

El problema del control de inferencias fue estudiado seriamente por primera vez en el contexto de los datos del censo. Un censo recopila una gran cantidad de información sensible sobre individuos y luego hace resúmenes estadísticos disponibles por unidades geográficas (y gubernamentales) como regiones, distritos y barrios. Esta información se utiliza para determinar distritos electorales, establecer niveles de financiación gubernamental para servicios públicos y como insumos para todo tipo de decisiones políticas. El problema del censo es algo más simple que el problema de los registros médicos, ya que los datos son bastante restringidos y están en un formato estándar (edad, sexo, raza, ingresos, número de hijos, nivel educativo más alto alcanzado, etc.).

Hay dos enfoques generales, dependiendo de si los datos se desidentifican antes o durante el procesamiento, o, en otras palabras, si el software que procesará los datos es no confiable o confiable.

Un ejemplo del primer tipo de procesamiento proviene del tratamiento de los datos del censo de EE. UU. hasta la década de 1960. El procedimiento entonces era que un registro de cada mil estaba disponible en cinta, menos los nombres, direcciones exactas y otros datos sensibles. También se agregaba ruido a los datos para evitar que las personas con algún conocimiento adicional (como los salarios pagados por el empleador en una ciudad de compañía) rastrearán a los individuos. Además de los registros de muestra, también se proporcionaban promedios locales para las personas seleccionadas por varios atributos. Pero se suprimían los registros con valores extremos, como

ingresos muy altos. La razón de esto es que una familia adinerada que vive en un pequeño pueblo podría hacer una diferencia significativa en el ingreso per cápita del pueblo. Por lo tanto, su ingreso podría deducirse comparando el ingreso promedio del pueblo con el de otros pueblos cercanos.

En el segundo tipo de procesamiento, se retienen datos identificables en una base de datos y la protección de la privacidad proviene del control del tipo de consultas que se pueden realizar. Los primeros intentos en esto no tuvieron mucho éxito y se propusieron varios ataques al procesamiento utilizado en ese momento por el censo de EE. UU. La pregunta era si era posible construir una serie de consultas sobre muestras que contuvieran a un individuo objetivo y trabajar para obtener información supuestamente confidencial sobre ese individuo.

Si nuestro sistema de censo permite una amplia gama de consultas estadísticas, como "dime el número de hogares encabezados por un hombre que gane entre \$50,000 y \$55,000", "dime la proporción de hogares encabezados por un hombre de 40 a 45 años que gane entre \$50,000 y \$55,000", "dime la proporción de hogares encabezados por un hombre que gane entre \$50,000 y \$55,000 cuyos hijos hayan crecido y se hayan ido de casa", y así sucesivamente, entonces un atacante puede rápidamente centrarse en un individuo. Tales consultas, en las que agregamos información circunstancial adicional para derrotar el promedio y otros controles, se conocen como rastreadores. Generalmente son fáciles de construir.

Un problema relacionado con la inferencia es que un oponente que obtiene varios archivos no clasificados podría deducir información sensible de ellos. Por ejemplo, un periodista neozelandés dedujo las identidades de muchos oficiales en el GCSB (el equivalente en ese país de la NSA) al examinar listas de personal de servicio y buscar patrones de asignaciones a lo largo del tiempo [576]. Las asignaciones de cobertura de los oficiales de inteligencia también podrían ser descubiertas si un oponente obtiene el directorio telefónico interno de la unidad donde se supone que está asignado el oficial y no encuentra su nombre allí. La lista del ejército podría ser pública y el directorio telefónico "Restringido"; pero el hecho de que un oficial dado esté involucrado en el trabajo de inteligencia podría ser "Secreto". Combinar fuentes de bajo nivel para sacar una conclusión de alto nivel se conoce como un ataque de agregación. Está relacionado con el mayor riesgo para la información personal que surge cuando las bases de datos se agregan juntas, lo que hace que más contexto esté disponible para el atacante y facilita los ataques de rastreo y otros. Las técnicas que se pueden utilizar para contrarrestar las amenazas de agregación son similares a las utilizadas para ataques de inferencia general en bases de datos, aunque hay algunos problemas particularmente difíciles cuando tenemos una política de seguridad multinivel y las amenazas de inferencia o agregación tienen el potencial de subvertirla.

9.3.3 La Teoría del Control de Inferencias

Una teoría del control de inferencias fue desarrollada por Denning y otros a fines de la década de 1970 y principios de la de 1980, en gran parte en respuesta a los problemas de las oficinas de censo [369]. Los desarrolladores de muchos sistemas modernos de privacidad a menudo desconocen este trabajo y repiten muchos de los errores de la década de 1960. (El control de inferencias no es el único problema en la seguridad

informática donde esto sucede). A continuación se presenta una visión general de las ideas más importantes.

Una fórmula característica es la expresión (en algún lenguaje de consulta de bases de datos) que selecciona un conjunto, conocido como el conjunto de consulta, de registros. Un ejemplo podría ser "todas las empleadas del Laboratorio de Computación en el grado de profesora". Los conjuntos de consulta más pequeños, obtenidos por el AND lógico de todos los atributos (o sus negaciones), se conocen como conjuntos elementales o celdas. Las estadísticas correspondientes a los conjuntos de consulta pueden ser estadísticas sensibles si cumplen con los criterios que discutiré a continuación (como que el tamaño del conjunto sea demasiado pequeño). El objetivo del control de inferencias es evitar la divulgación de estadísticas sensibles.

Si dejamos que D sea el conjunto de estadísticas que se divulgan y P el conjunto que es sensible y debe ser protegido, entonces necesitamos $D \cap P = \emptyset$, para la privacidad, donde P es el complemento de P . Si $D \subseteq P$, entonces la protección se dice que es precisa. La protección que no es precisa generalmente llevará algún costo en términos del rango de consultas que la base de datos puede responder y, por lo tanto, puede degradar su utilidad para su propietario.

9.3.3.1 Control del Tamaño del Conjunto de Consultas

El mecanismo de protección más simple es especificar un tamaño mínimo de consulta. Como mencioné, las bases de datos del Sistema Nacional de Información de Salud de Nueva Zelanda rechazarán las consultas estadísticas cuyas respuestas se basarían en menos de seis registros de pacientes. Pero esto no es suficiente en sí mismo. Un ataque rastreador obvio es hacer una consulta sobre los registros de seis pacientes y luego sobre esos registros más el objetivo. En lugar de reducir la efectividad de la base de datos incorporando controles de consulta más restrictivos, los diseñadores de este sistema optaron por restringir el acceso a un pequeño número de estadísticos médicos especialmente autorizados.

Aun así, se necesita un control adicional y a menudo se olvida. Debes evitar que el atacante consulte todos menos uno de los registros en la base de datos. En general, si hay N registros, el control del tamaño del conjunto de consultas con un umbral de t significa que entre t y $N - t$ de ellos deben ser el sujeto de una consulta para que se permita.

9.3.3.2 Rastreadores

Probablemente los ataques más importantes a las bases de datos estadísticas provienen de rastreadores. Hay muchos ejemplos simples. En nuestro laboratorio, solo una de las profesoras titulares es mujer. Así que podemos encontrar su salario con solo dos consultas: "¿Salario promedio de los profesores?" y "¿Salario promedio de los profesores hombres?".

Este es un ejemplo de un rastreador individual, una fórmula personalizada que nos permite calcular la respuesta a una consulta prohibida indirectamente. También hay rastreadores generales: conjuntos de fórmulas que permitirán revelar cualquier

estadística sensible. Un descubrimiento algo deprimente realizado a fines de la década de 1970 fue que los rastreadores generales son generalmente fáciles de encontrar. Siempre que el tamaño mínimo del conjunto de consultas n sea menos de una cuarta parte del número total de estadísticas N y no haya más restricciones en el tipo de consultas que se permiten, entonces podemos encontrar fórmulas que proporcionen rastreadores generales [372]. Así que los ataques de rastreo son fáciles, a menos que coloquemos restricciones severas en el tamaño del conjunto de consultas o controlemos las consultas permitidas de alguna otra manera. (De hecho, resultados como este hicieron que la comunidad de investigación perdiera en gran medida el interés en la seguridad de inferencias por ser "demasiado difícil", y esta es una de las razones por las que muchos diseñadores de sistemas no son conscientes de los problemas y construyen bases de datos vulnerables a rastreadores y otros ataques).

9.3.3.3 Controles de Consultas Más Sofisticados

Hay varias alternativas al control simple del tamaño del conjunto de consultas. El censo de EE. UU., por ejemplo, utiliza la "regla de dominancia n -respondent, $k\%$ ": no divulgará una estadística de la cual el $k\%$ o más sea contribuido por n valores o menos. Otras técnicas incluyen, como mencioné, suprimir datos con valores extremos. Una oficina de censo puede tratar con individuos de alto patrimonio en estadísticas nacionales, pero no en las cifras locales, mientras que algunas bases de datos médicas hacen lo mismo para enfermedades menos comunes. Por ejemplo, un sistema de estadísticas de prescripción del Reino Unido suprime las ventas del medicamento contra el SIDA AZT de las estadísticas locales [847]. Cuando se diseñó a fines de la década de 1990, había condados con solo un paciente recibiendo este medicamento.

9.3.3.4 Supresión de Celdas

La siguiente pregunta es cómo lidiar con los efectos secundarios de suprimir ciertas estadísticas. Las reglas del Reino Unido, por ejemplo, requieren que sea "improbable que cualquier unidad estadística, habiéndose identificado a sí misma, pueda usar ese conocimiento, por deducción, para identificar otras unidades estadísticas en los resultados de las Estadísticas Nacionales" [953]. Para concretar esto, supongamos que una universidad quiere publicar las notas promedio para varias combinaciones de cursos, de modo que la gente pueda verificar que la calificación es justa en todos los cursos. Supongamos ahora que la tabla en la Figura 9.4 contiene el número de estudiantes que estudian dos materias de ciencias, una como su materia principal y otra como su materia secundaria.

Las reglas del Reino Unido implican que nuestro tamaño mínimo del conjunto de consultas es 3 (si lo establecemos en 2, entonces cualquiera de los dos estudiantes que estudiaron "geología con química" podría trivialmente averiguar la nota del otro). Entonces no podemos publicar la nota promedio para "geología con química". Pero si se conoce la nota promedio de química, entonces esta nota puede reconstruirse fácilmente a partir de los promedios para "biología con química" y "física con química". Así que tenemos que suprimir al menos una otra nota en la fila de química, y por razones similares necesitamos suprimir una en la columna de geología. Pero si suprimimos "geología con biología" y "física con química", entonces también deberíamos suprimir

"física con biología" para evitar que estos valores se descubran a su vez. Nuestra tabla ahora se verá como la Figura 9.5.

Este proceso se llama supresión de celdas complementarias. Si hay más atributos en el esquema de la base de datos, por ejemplo, si las cifras también se desglosan por raza y sexo, para mostrar el cumplimiento con las leyes antidiscriminación, entonces se puede perder aún más información. Donde un esquema de base de datos contiene m-tuplas, borrar una sola celda generalmente significa suprimir $2^m - 1$ celdas en total.

Major:	Biology	Physics	Chemistry	Geology
Minor:				
Biology	–	16	17	11
Physics	7	–	32	18
Chemistry	33	41	–	2
Geology	9	13	6	–

Figure 9.4: Table containing data before cell suppression

Major:	Biology	Physics	Chemistry	Geology
Minor:				
Biology	–	blanked	17	blanked
Physics	7	–	32	18
Chemistry	33	blanked	–	blanked
Geology	9	13	6	–

Figure 9.5: Table after cell suppression

Control de Máximo Orden y el Modelo de Lattice

Lo siguiente que podríamos intentar para dificultar la construcción de rastreadores es limitar el tipo de consultas que se pueden hacer. El control de máximo orden limita la cantidad de atributos que puede tener cualquier consulta. Sin embargo, para ser efectivo, el límite puede tener que ser severo. Un estudio encontró que de 1000 registros médicos, tres atributos eran seguros mientras que con cuatro atributos, se podía encontrar un registro individual y con 10 atributos se podían aislar la mayoría de los registros. Un enfoque más completo (donde sea factible) es rechazar consultas que dividirían la población de muestra en demasiados conjuntos.

Vimos cómo se pueden usar los lattice en la seguridad compartimentada para definir un orden parcial para controlar los flujos de información permitidos entre compartimentos con combinaciones de palabras clave. También se pueden usar de una manera ligeramente diferente para sistematizar los controles de consultas en algunas bases de datos. Si tenemos, por ejemplo, tres atributos A, B y C (digamos, área de residencia, año de nacimiento y condición médica), podemos encontrar que, aunque las consultas sobre cualquiera de estos atributos no son sensibles, al igual que las consultas sobre A y B y sobre B y C, la combinación de A y C podría ser sensible. Se sigue que una consulta sobre los tres no sería permisible tampoco. Así, el lattice se divide naturalmente en una

'mitad superior' de consultas prohibidas y una 'mitad inferior' de consultas permitidas, como se muestra en la Figura 9.6.

9.3.3.6 Control Basado en Auditoría

Como se mencionó, algunos sistemas intentan sortear los límites impuestos por el control de consultas estático llevando un registro de quién accedió a qué. Conocido como control de superposición de consultas, este método permite a los administradores de bases de datos monitorear las consultas y detectar patrones de acceso que podrían sugerir intentos de inferencia.

Este enfoque implica registrar y analizar todas las consultas realizadas para identificar patrones que podrían permitir a un atacante inferir información sensible. Por ejemplo, si un usuario hace una serie de consultas que, en conjunto, pueden reducir un conjunto de datos a un tamaño pequeño y, por lo tanto, permitir la identificación de individuos, se podrían tomar medidas para prevenir más consultas de ese tipo o para alertar a los administradores de la base de datos sobre el posible abuso.

El control basado en auditoría tiene la ventaja de ser dinámico y adaptativo, ya que puede responder a patrones de consulta en tiempo real, en lugar de depender únicamente de reglas estáticas predefinidas. Sin embargo, también puede ser más complejo de implementar y requiere una vigilancia continua para ser efectivo.

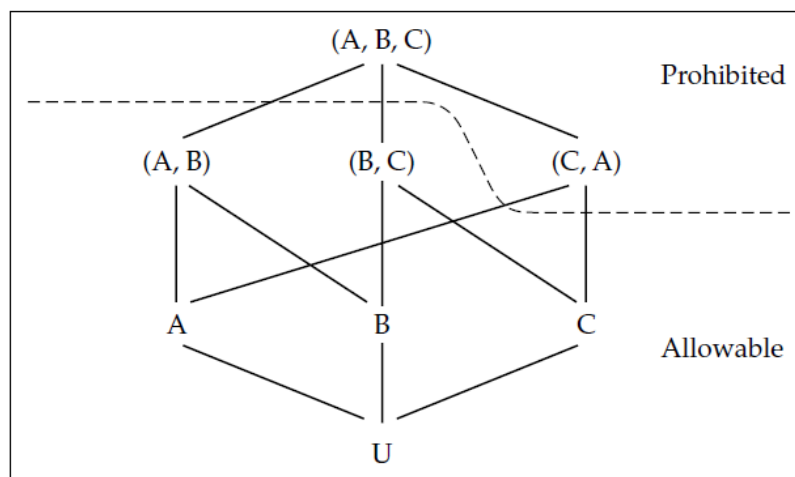


Figure 9.6: Table lattice for a database with three attributes

Este método implica rechazar cualquier consulta de un usuario que, combinada con lo que el usuario ya sabe, revelaría una estadística sensible. Esto puede sonar perfecto en teoría, pero en la práctica sufre de dos inconvenientes generalmente insuperables. Primero, la complejidad del procesamiento involucrado aumenta con el tiempo, y a menudo exponencialmente. Segundo, es extremadamente difícil asegurarse de que sus usuarios no están en colusión, o que un usuario no se ha registrado con dos nombres diferentes. Incluso si todos sus usuarios son personas honestas y distintas hoy, siempre

es posible que uno de ellos se apodere de otro, o sea tomado por un depredador, mañana.

9.3.3.7 Aleatorización

Nuestro ejemplo de supresión de celdas muestra que si los diversos tipos de control de consultas son los únicos mecanismos de protección utilizados en una base de datos estadística, a menudo pueden tener una penalización de rendimiento inaceptable. Por lo tanto, el control de consultas se usa a menudo junto con varios tipos de aleatorización, diseñados para degradar la relación señal-ruido desde el punto de vista del atacante, mientras se afecta lo menos posible al usuario legítimo.

La técnica más simple de este tipo es la perturbación, o agregar ruido con media cero y una varianza conocida a los datos. Una forma de hacerlo es redondear o truncar los datos por alguna regla determinista; otra es intercambiar algunos registros. La perturbación a menudo no es tan efectiva como se desearía, ya que tiende a dañar los resultados del usuario legítimo precisamente cuando los tamaños de los conjuntos de muestra son pequeños, y dejarlos intactos cuando los conjuntos de muestra son grandes (donde podríamos haber podido usar controles de consultas simples de todos modos). También existe la preocupación de que se puedan usar técnicas de promediado adecuadas para eliminar parte del ruido agregado. Una variante moderna y sofisticada en el mismo tema es el ajuste tabular controlado, donde se identifican las celdas sensibles y se reemplazan sus valores por otros "seguros" (suficientemente diferentes), y luego se ajustan otros valores en la tabla para restaurar las relaciones aditivas [330].

A menudo, una buena técnica de aleatorización es usar consultas de muestra aleatoria. Este es otro de los métodos utilizados por las oficinas de censo. La idea es que hacemos todos los conjuntos de consulta del mismo tamaño, seleccionándolos al azar de las estadísticas relevantes disponibles. Así, todos los datos liberados se calculan a partir de muestras pequeñas en lugar de toda la base de datos. Si esta selección aleatoria se hace usando un generador de números pseudoaleatorios claveado a la consulta de entrada, los resultados tendrán la virtud de la repetibilidad. Las consultas de muestra aleatoria son un mecanismo de protección natural para grandes bases de datos médicas, donde las correlaciones que se investigan a menudo son tales que una muestra de unos pocos cientos es suficiente. Por ejemplo, al investigar la correlación entre una enfermedad dada y algún aspecto del estilo de vida, la correlación debe ser fuerte antes de que los médicos aconsejen a los pacientes hacer cambios radicales en su forma de vida, o tomar otras acciones que podrían tener efectos secundarios indeseables. Si un hospital de enseñanza tiene registros de cinco millones de pacientes, y cinco mil tienen la enfermedad que se investiga, entonces una muestra seleccionada al azar de doscientos pacientes podría ser todo lo que el investigador pueda usar.

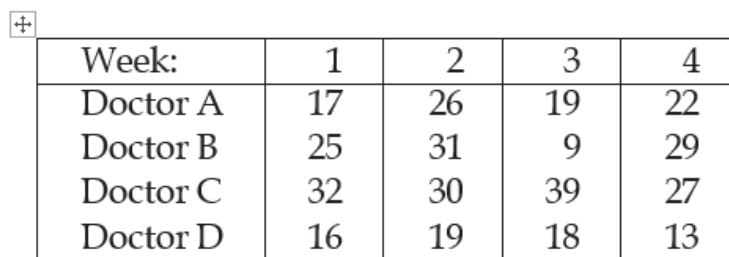
Esto no funciona tan bien cuando la enfermedad es rara, o donde por otras razones hay solo un pequeño número de estadísticas relevantes. Una posible estrategia aquí es la respuesta aleatoria, donde aleatoriamente restringimos los datos que recopilamos (las respuestas de los sujetos). Por ejemplo, si las tres variables bajo investigación son obesidad, fumar y SIDA, podríamos pedir a cada sujeto con infección por VIH que registre si fuma o si tiene sobrepeso, pero no ambos. Por supuesto, esto puede limitar el valor de los datos.

9.3.4 Limitaciones de los Enfoques Genéricos

Como ocurre con cualquier tecnología de protección, la seguridad estadística solo puede evaluarse en un entorno particular y contra un modelo de amenaza particular. Si es adecuada o no, depende en mayor medida de lo habitual de los detalles de la aplicación.

Un ejemplo instructivo es un sistema utilizado para analizar tendencias en la prescripción de medicamentos. Aquí, las recetas se recopilan (sin nombres de pacientes) de las farmacias. Una etapa adicional de desidentificación elimina las identidades de los médicos, y la información se vende luego a los departamentos de marketing de las compañías farmacéuticas. El sistema debe proteger la privacidad de los médicos así como de los pacientes: lo último que un médico de familia ocupado quiere es ser molestado por un representante farmacéutico por recetar marcas de la competencia.

Un prototipo temprano de este sistema simplemente reemplazaba los nombres de los médicos en una celda de cuatro o cinco prácticas con "doctor A", "doctor B", y así sucesivamente, como se muestra en la Figura 9.7. Nos dimos cuenta de que un representante farmacéutico alerta podría identificar a los médicos a partir de los patrones de prescripción, notando, por ejemplo, "Bueno, el doctor B debe ser Susan".



Week:	1	2	3	4
Doctor A	17	26	19	22
Doctor B	25	31	9	29
Doctor C	32	30	39	27
Doctor D	16	19	18	13

Figure 9.7: Sample of de-identified drug prescribing data

Active Attacks

Los ataques activos son particularmente poderosos. Estos son aquellos donde los usuarios tienen la capacidad de insertar o eliminar registros en la base de datos. Un usuario podría agregar registros para crear un grupo que contenga el registro del objetivo más los de varios sujetos inexistentes creados por él mismo. Una contramedida (imperfecta) es agregar o eliminar nuevos registros en lotes. Llevado al extremo, esto da lugar a la partición, en la cual los registros se agregan en grupos y cualquier consulta debe responderse con respecto a todos ellos o ninguno. Sin embargo, esto equivale una vez más a publicar tablas de microestadísticas.

Los ataques activos no se limitan a los datos, sino que también pueden dirigirse a los metadatos. Un buen ejemplo, debido a Whit Diffie, es el ataque de fármacos elegidos. Supongamos que una empresa farmacéutica tiene acceso a través de un sistema estadístico a las cantidades de dinero gastadas en nombre de varios grupos de pacientes y desea averiguar qué pacientes están recibiendo qué medicamento, con el fin de dirigir mejor su marketing (hubo un escándalo en Quebec sobre un ataque de inferencia de este tipo). Un truco posible es establecer los precios de los medicamentos de tal manera que las ecuaciones resultantes sean fáciles de resolver. Un caso prominente a finales del

siglo XX fue una base de datos de investigación médica en Islandia. El plan era tener tres bases de datos vinculadas: una con los registros médicos de la nación, una segunda con la genealogía de toda la población y una tercera con datos genéticos adquiridos mediante secuenciación. La justificación era que, dado que la población de Islandia descende en gran medida de unas pocas familias fundadoras que se establecieron allí hace unos mil años, hay mucha menos variación genética que en la población humana en general, por lo que los genes de enfermedades hereditarias deberían ser mucho más fáciles de encontrar. Una empresa farmacéutica suiza financió la construcción de la base de datos, y el gobierno de Reikiavik la abrazó como un medio para modernizar la infraestructura de TI de salud del país y al mismo tiempo crear unos pocos cientos de empleos de alta tecnología en investigación médica. Sin embargo, la mayoría de los médicos islandeses reaccionaron negativamente, viendo el sistema como una amenaza tanto para la privacidad del paciente como para la autonomía profesional.

El problema de privacidad en la base de datos islandesa era más agudo que en el caso general. Por ejemplo, al vincular los registros médicos con genealogías, que en cualquier caso son públicas (la genealogía es un pasatiempo común en Islandia), se puede identificar a los pacientes por factores como el número de sus tíos, tías, tíos abuelos, tías abuelas y así sucesivamente, en efecto, por la forma de sus árboles genealógicos. Hubo mucho debate sobre si el diseño podría incluso teóricamente cumplir con los requisitos legales de privacidad [47], y los funcionarios de privacidad europeos expresaron grave preocupación por las posibles consecuencias para el sistema de leyes de privacidad de Europa [349]. El gobierno islandés siguió adelante de todos modos, con una opción de exclusión para los pacientes. Muchos médicos aconsejaron a los pacientes que se excluyeran, y el 11% de la población lo hizo. Finalmente, el Tribunal Supremo de Islandia determinó que la ley de privacidad europea requería que la base de datos fuera de inclusión voluntaria en lugar de exclusión voluntaria. Además, muchos islandeses habían invertido en la empresa de la base de datos y perdieron dinero cuando el valor de sus acciones se desplomó al final del auge de las puntocom. Hoy en día, alrededor de la mitad de la población ha optado por participar en el sistema y la controversia se ha desactivado.

Mi propia opinión, por lo que vale, es que el consentimiento del paciente es la clave para una investigación médica efectiva. Esto no solo permite el acceso completo a los datos, sino los problemas que hemos estado discutiendo en esta sección, sino que proporciona sujetos motivados y una información clínica de mucha mayor calidad que la que se puede recolectar simplemente como un subproducto de las actividades clínicas normales. Por ejemplo, una red de investigadores sobre la ELA (la enfermedad de las neuronas motoras que sufre el astrónomo de Cambridge Stephen Hawking) comparte información totalmente identificable entre médicos e investigadores de más de una docena de países con el consentimiento total de los pacientes y sus familias. Esta red permite compartir datos entre Alemania, con leyes de privacidad muy estrictas, y Japón, con casi ninguna; y los datos continuaron compartiéndose entre investigadores en EE. UU. y Serbia incluso cuando la Fuerza Aérea de los EE. UU. estaba bombardeando Serbia. El modelo de consentimiento se está extendiendo. La mayor organización benéfica médica de Gran Bretaña está financiando una base de datos "Biobank" en la que se pedirá a varios cientos de miles de voluntarios que proporcionen a los investigadores no solo respuestas a un extenso cuestionario y acceso completo a sus registros de por vida, sino también muestras de sangre para que aquellos que desarrollen

enfermedades interesantes más adelante en la vida puedan tener su composición genética y proteómica analizada.

9.3.5 El Valor de la Protección Imperfecta

Hacer la desidentificación correctamente es difícil, y los problemas pueden ser políticamente conflictivos. La mejor manera de resolver el problema del control de inferencias es evitarlo, por ejemplo, reclutando voluntarios para su investigación médica en lugar de reciclar datos recopilados para otros propósitos. Pero hay aplicaciones donde se usa, y aplicaciones donde es todo lo que está disponible. Un ejemplo fue la epidemia de VIH/SIDA; en las décadas de 1980 y 1990, los investigadores que luchaban por entender lo que estaba sucediendo tenían pocas opciones más que usar datos médicos que se habían recopilado originalmente para otros fines. Otro ejemplo, por supuesto, es el censo. En tales aplicaciones, la protección que puede proporcionar será imperfecta. ¿Cómo se maneja eso?

Algunos tipos de mecanismos de seguridad pueden ser peores que inútiles si se pueden comprometer. Un buen ejemplo es el cifrado débil. El principal problema que enfrentan las agencias de inteligencia de señales del mundo es la selección de tráfico: cómo filtrar los fragmentos interesantes del gran volumen de tráfico internacional de teléfono, fax, correo electrónico y otros. Un terrorista que cifra útilmente su tráfico importante hace esta parte del trabajo de la policía por ellos. Si el algoritmo de cifrado utilizado es vulnerable, o si los sistemas finales pueden ser hackeados, entonces el resultado neto es peor que si el tráfico se hubiera enviado en claro.

La seguridad estadística no es generalmente así. La principal amenaza para las bases de datos de información personal suele ser la expansión de la misión. Una vez que una organización tiene acceso a datos potencialmente valiosos, se desarrollarán todo tipo de formas de explotar ese valor. Algunas de estas probablemente serán altamente objetables; un ejemplo reciente en EE. UU. es la reventa de registros médicos a bancos para su uso en la filtración de solicitudes de préstamos. Sin embargo, incluso un sistema de desidentificación imperfecto puede destruir el valor de los datos médicos para el departamento de préstamos de un banco. Si solo se puede identificar al 5% de los pacientes, y solo con esfuerzo, el banco puede decidir que es más simple decirles a los solicitantes de préstamos que contraten su propio seguro y dejar que las compañías de seguros envíen cuestionarios médicos si lo desean. Por lo tanto, la desidentificación puede ayudar a prevenir la expansión de la misión, incluso si el efecto principal es la profilaxis contra daños futuros en lugar del tratamiento de defectos existentes.

Además de dañar la privacidad, la expansión de la misión puede tener implicaciones de seguridad. En el Reino Unido, se establecieron registros de diabéticos en la década de 1990 para monitorear la calidad de la atención de la diabetes; eran bases de datos a las que los médicos generales, consultores hospitalarios, enfermeras y oftalmólogos podían cargar resultados de pruebas, para que no se pasaran por alto indicadores importantes. Como los hospitales no tenían un sistema de correo electrónico funcional, se abusó de ellos rápidamente para proporcionar un sistema rudimentario de mensajería entre hospitales y la práctica general. Pero como los registros de diabetes nunca fueron diseñados como sistemas de comunicación, carecían de los mecanismos de seguridad y otros que dichos sistemas deberían haber tenido si se iban a usar para datos clínicos.

Incluso una desidentificación rudimentaria habría prevenido este abuso y motivado a los diabetólogos a poner en funcionamiento el correo electrónico en su lugar.

Entonces, en la seguridad estadística, la pregunta de si uno debería dejar que lo mejor sea enemigo de lo bueno puede requerir una evaluación más detallada que en otros lugares.

9.4 El Problema Residual

Las secciones anteriores pueden haberlo convencido de que el problema de gestionar la privacidad de los registros médicos en el contexto de la atención inmediata (como en un hospital) es razonablemente sencillo, mientras que en el contexto de bases de datos secundarias (como para investigación, auditoría y control de costos) existen técnicas de seguridad estadística que, con cuidado, pueden resolver gran parte del problema. Técnicas algo similares se pueden usar para gestionar datos comerciales altamente sensibles, como detalles de fusiones y adquisiciones inminentes en un banco de inversión, e incluso información de inteligencia. (Hubo mucho interés en el modelo de la BMA por parte de personas que diseñaban sistemas de inteligencia policial). En todos los casos, el concepto subyacente es que el material realmente secreto se restringe a un compartimento de un pequeño número de individuos identificados, y se pueden fabricar versiones menos secretas de los datos para un uso más amplio. Esto implica no solo suprimir los nombres de los pacientes, o espías, o empresas objetivo, sino también la gestión cuidadosa de la información contextual y otra información mediante la cual podrían ser reidentificados.

Pero hacer que tales sistemas funcionen bien en la vida real es mucho más difícil de lo que parece. Primero, determinar el nivel de sensibilidad de la información es extremadamente difícil, y muchas expectativas iniciales resultan ser incorrectas. Podrías esperar, por ejemplo, que el estado de VIH sea la información médica más sensible que existe; sin embargo, muchos portadores de VIH son bastante abiertos sobre su estado. También podrías esperar que las personas prefieran confiar información personal sensible a un profesional de la salud como un médico o farmacéutico en lugar de a una base de datos de marketing. Sin embargo, muchas mujeres son tan sensibles sobre la compra de productos de higiene femenina que, en lugar de ir a una farmacia y comprarlos en efectivo, prefieren usar una instalación de autopago en un supermercado, incluso si esto significa que tienen que usar su tarjeta de la tienda y tarjeta de crédito, para que la compra se vincule a su nombre y permanezca en la base de datos de marketing para siempre. La vergüenza inmediata de ser vista con un paquete de tampones es inmediata y supera la vergüenza futura de recibir cupones de descuento para ropa de bebé seis meses después de la menopausia.

Segundo, es extraordinariamente difícil excluir puntos únicos de falla, no importa cuánto intentes construir compartimentos herméticos. Los activos soviéticos de la CIA fueron comprometidos por Aldrich Ames, quien, como hombre senior de contrainteligencia, tenía acceso a demasiados compartimentos. Las operaciones en el extranjero del KGB fueron comprometidas de manera similar por Vassily Mitrokhin, un oficial que se desilusionó con el comunismo después de 1968 y fue enviado a trabajar en los archivos mientras esperaba su pensión [77]. Y en marzo de 2007, los historiadores Margo Anderson y William Seltzer encontraron que, contrariamente a

décadas de negaciones, los datos del censo se utilizaron en 1943 para reunir a japoneses-estadounidenses para su internamiento [1142]. El punto único de falla allí parece haber sido el director de la Oficina del Censo JC Capt, quien liberó ilegalmente los datos al Servicio Secreto tras una solicitud del Secretario del Tesoro HC Morgenthau. La Oficina ha pedido disculpas públicamente desde entonces [893].

En medicina, muchos de los problemas difíciles están en los sistemas que procesan reclamaciones médicas para pago. Cuando un paciente es tratado y se envía una solicitud de pago al asegurador, este no solo tiene detalles completos de la enfermedad, el tratamiento y el costo, sino también el nombre del paciente, el número de seguro y otros detalles como la fecha de nacimiento. Ha habido propuestas para que el pago se efectúe utilizando tarjetas de crédito anónimas [191], pero hasta donde sé, ninguna de ellas se ha implementado. Los aseguradores quieren saber qué pacientes y qué médicos son los más caros. De hecho, durante un debate sobre la privacidad médica en una conferencia del IEEE en 1996, justo cuando HIPAA se estaba aprobando en el Congreso de EE. UU., un representante de una gran empresa de sistemas declaró que los registros médicos de 8 millones de estadounidenses eran uno de los activos estratégicos de su empresa, que nunca abandonarían. Esto se mantiene ya sea que el asegurador sea una compañía de seguros privada (o empleador) o una autoridad de salud gubernamental, como HCFA, VA, o el Servicio Nacional de Salud de Gran Bretaña. Una vez que un asegurador posee grandes cantidades de información de salud personal, se vuelve muy reacio a eliminarla. Su valor potencial futuro, en todo tipo de aplicaciones desde control de costos hasta investigación y marketing, es inmediato y obvio, mientras que las preocupaciones de privacidad de los pacientes no lo son.

En EE. UU., la retención de copias de registros médicos por aseguradores, empleadores y otros se ve ampliamente como un problema grave. Escritores de puntos de vista políticos tan diferentes como el comunitarista Amitai Etzioni [441] y el libertario Simson Garfinkel [515] coinciden en este punto, si en poco más. Como se mencionó, HIPAA solo facultó al DHHS para regular los planes de salud, los centros de intercambio de información de salud y los proveedores de atención médica, dejando fuera de su alcance a muchas organizaciones que procesan datos médicos (como abogados, empleadores y universidades). De hecho, el anuncio reciente de Microsoft de que establecería un "HealthVault" para proteger sus registros médicos fue recibido con una respuesta aguda de los activistas de privacidad que señalaron que, dado que Microsoft no es una "entidad cubierta" según se especifica en HIPAA, poner sus datos médicos allí los dejaría fuera de la protección de HIPAA [81].

¿Qué lecciones se pueden sacar de otros países?

La privacidad médica está fuertemente condicionada por cómo las personas pagan por la atención médica. En Gran Bretaña, el gobierno paga la mayor parte de la atención médica, y los intentos de los sucesivos gobiernos británicos de centralizar los registros médicos para control de costos y propósitos de gestión han llevado a más de una década de conflicto con los médicos y las asociaciones de pacientes. En Alemania, las personas más ricas usan aseguradores privados (que están sujetos a leyes de protección de datos estrictas), mientras que los pobres usan aseguradores estatales de salud que son administrados por médicos, por lo que los no médicos no tienen acceso a los registros. Los residentes de Singapur pagan en cuentas de ahorro obligatorias de sus salarios y las

utilizan para pagar la atención médica; el gobierno interviene para asegurar procedimientos costosos, pero la mayoría de las visitas al médico son pagadas directamente por el paciente. Los pacientes que se mantienen saludables y acumulan un excedente pueden agregar parte de este a su pensión y pasar el resto a sus herederos. La solución más radical está en Japón, donde los costos se controlan regulando las tarifas: los médicos están desalentados de realizar procedimientos costosos como trasplantes de corazón al ponerles un precio por debajo del costo. A mediados de la década de 1990, la atención médica representaba alrededor del 3% del PIB en Japón, frente al 7-9% para el país desarrollado típico y el 15% para EE. UU.; desde entonces, las cifras han aumentado en un porcentaje o más, pero las clasificaciones generales siguen siendo las mismas. Los japoneses (y los singapurenses) pagan menos por la atención médica que los europeos, y los estadounidenses pagan más. Lo curioso es que los japoneses (y los singapurenses) viven más que los europeos, que viven más que los estadounidenses. La esperanza de vida y los costos médicos parecen estar negativamente correlacionados.

En resumen, el problema de la privacidad de los registros de salud no es solo socio-técnico, sino socio-técnico-político. Si grandes cantidades de registros médicos se acumulan en una base de datos depende de cómo esté organizado el sistema de atención médica, y si estos se destruyen o se desidentifican después de procesar el pago tiene más que ver con las estructuras institucionales, los incentivos y la regulación que con la tecnología. En tales debates, una función del ingeniero de seguridad es lograr que los responsables de la formulación de políticas comprendan las posibles consecuencias de sus acciones.

La privacidad es más pobre en los países que no alinean adecuadamente los incentivos, y como resultado, tienen una supervisión detallada de los costos de los tratamientos individuales, ya sea por parte de aseguradores/empleadores, como en EE. UU., o por burócratas, como en Gran Bretaña.

En el Reino Unido, estalló un escándalo en noviembre de 2007 cuando las autoridades fiscales perdieron los registros de 25 millones de personas. Los registros de todos los niños de la nación y sus familias, incluidos nombres, direcciones, números de teléfono y detalles de las cuentas bancarias de los padres, fueron grabados en dos CD para su envío a la Oficina Nacional de Auditoría y se perdieron en el correo. El Primer Ministro tuvo que disculparse ante el Parlamento y prometió compensar cualquier pérdida resultante de "robo de identidad". Tras esto, ha habido una amplia cuestionamiento público sobre el programa de su gobierno para construir bases de datos centrales cada vez más grandes de información personal de los ciudadanos, no solo para impuestos sino para investigación médica, administración del servicio de salud y bienestar infantil. Mientras escribo en diciembre de 2007, la sensación en Londres es que los planes para una tarjeta de identificación nacional están efectivamente muertos, al igual que una propuesta para construir una base de datos de todos los movimientos de vehículos para facilitar el cobro de peajes. El Servicio Nacional de Salud continúa construyendo bases de datos de salud centralizadas contra la creciente resistencia médica, pero el Partido Conservador de la oposición (que ahora tiene una clara ventaja en las encuestas) ha prometido abolir no solo el sistema de tarjetas de identificación sino las propuestas de bases de datos de niños si ganan las próximas elecciones. Otros problemas de privacidad también tienden a tener un serio enredo político.

La privacidad de los clientes bancarios puede estar relacionada con la política interna del banco; el impulso más fuerte para la protección de la privacidad puede provenir de la renuencia de los gerentes de sucursales a permitir que otras sucursales conozcan a sus clientes. El acceso a registros criminales e información de inteligencia depende de cómo las agencias de aplicación de la ley deciden compartir datos entre sí y de las elecciones que hacen internamente sobre si el acceso a información altamente sensible sobre fuentes y métodos debe ser descentralizado (arriesgando pérdidas ocasionales) o centralizado (trayendo exposición de baja probabilidad pero alto costo a un traidor en la oficina central). El mundo desde el 11 de septiembre se ha movido abruptamente hacia la centralización; espere que un traidor de alto perfil como Aldrich Ames aparezca en algún momento pronto.

9.5 Resumen

En este capítulo, vimos el problema de asegurar la privacidad de los registros médicos. Este es típico de varios problemas de seguridad de la información, que van desde la protección de datos de inteligencia nacional a la práctica profesional en general hasta la protección de datos del censo.

Resulta que con los registros médicos hay un problema fácil, un problema más difícil y un problema realmente difícil.

El problema fácil es configurar sistemas de control de acceso para que el acceso a un registro particular esté limitado a un número sensato de empleados. Tales sistemas se pueden diseñar en gran medida automatizando las prácticas laborales existentes, y los controles de acceso basados en roles son actualmente la tecnología de elección. El problema más difícil es la seguridad estadística: cómo se diseñan bases de datos de registros médicos (o respuestas del censo) para permitir que los investigadores realicen consultas estadísticas sin comprometer la privacidad de los individuos. El problema más difícil es cómo gestionar la interfaz entre los dos, y en el caso específico de la medicina, cómo prevenir la difusión de la información de pago. La única solución realista para esto radica en la regulación.

Los sistemas médicos también nos enseñan sobre los límites de algunas tecnologías de mejora de la privacidad, como la desidentificación. Si bien hacer que los registros médicos sean anónimos en bases de datos de investigación puede ayudar a mitigar las consecuencias del acceso no autorizado y prevenir la expansión de la misión, no es a prueba de balas. Los datos ricos sobre personas reales generalmente pueden ser reidentificados. Los mecanismos utilizados en el cuidado de la salud para lidiar con este problema valen la pena estudiarlos.

Problemas de Investigación

En un futuro cercano, gran parte del tratamiento médico puede involucrar información genética. Por lo tanto, sus registros médicos pueden involucrar información de salud personal sobre sus padres, hermanos, primos, etc. ¿Cómo pueden extenderse los modelos de privacidad para tratar con múltiples individuos? Por ejemplo, en muchos países, tiene derecho a no conocer el resultado de una prueba de ADN que un pariente tiene para una enfermedad hereditaria como la corea de Huntington, ya que puede

afectar las probabilidades de que usted tenga la enfermedad también. Su pariente tiene derecho a saber y puede decírselo a otros. Este es un problema no solo para la tecnología, sino también para la ley de privacidad [1231]. ¿Existen formas de vincular las políticas de control de acceso para la privacidad con la seguridad estadística? ¿Puede existir algo como la privacidad perfecta donde todo encaja perfectamente? ¿O terminaría dando a los pacientes un conjunto extremadamente complejo de opciones de control de acceso, como las de Facebook pero peor, en las que cada paciente tendría que revisar docenas de páginas de opciones y aprobar o denegar el permiso para que sus datos se utilicen en cada una de docenas de aplicaciones secundarias y proyectos de investigación? En resumen, ¿existen abstracciones útiles y utilizables?

¿Qué otras formas de redactar políticas de privacidad existen? Por ejemplo, ¿hay formas útiles de combinar BMA y Chinese Wall? ¿Existen formas, ya sean técnicas o económicas, de alinear los intereses del sujeto de los datos con los del operador del sistema y otros interesados?

Lecturas Adicionales

La literatura sobre seguridad en modo compartimentado está algo dispersa: la mayoría de los artículos de dominio público se encuentran en las actas de las conferencias NCSC/NISSC y ACSAC citadas en detalle al final del Capítulo 8. Los libros de texto estándar como los de Amoroso [27] y Gollmann [537] cubren los fundamentos de los modelos de lattice y Chinese Wall.

Para el modelo BMA, consulte el documento de políticas en sí, el Blue Book [37], la versión más corta en [38], y las actas de la conferencia sobre la política [43]. Vea también los artículos sobre el sistema piloto en Hastings [366, 367]. Para más información sobre la atención médica en Japón, consulte [263]. Para un estudio del Consejo Nacional de Investigación sobre los problemas de privacidad médica en los EE. UU., consulte [951]; también hay un informe del HHS sobre el uso de datos desidentificados en la investigación en [816].

En cuanto al control de inferencias, este se ha convertido en un campo de investigación activo nuevamente en los últimos años, con conferencias regulares sobre "Privacidad en Bases de Datos Estadísticas"; consulte las actas de estos eventos para ponerse al día con las fronteras actuales. El libro de Denning [369] es la referencia clásica, y aún vale la pena echarle un vistazo; hay una actualización en [374]. Un libro de texto más moderno sobre seguridad en bases de datos es el de Castano et al [276]. El recurso más completo, desde el punto de vista práctico, con enlaces a una amplia gama de literatura práctica en varias áreas de aplicación, puede ser el sitio web de la American Statistical Association [26]. La referencia estándar para las personas involucradas en trabajos gubernamentales es el 'Informe sobre la Metodología de Limitación de Divulgación Estadística' del Comité Federal sobre Metodología Estadística, que proporciona una buena introducción a las herramientas estándar y describe los métodos utilizados en varios departamentos y agencias de los EE. UU. [455]. Como ejemplo de una aplicación bastante diferente, Mark Allman y Vern Paxson discuten los problemas de anonimizar trazas de paquetes IP para la investigación de sistemas de red en [23].

Finalmente, los artículos de Margo Anderson y William Seltzer sobre los abusos de los datos del censo en los EE. UU., particularmente durante la Segunda Guerra Mundial, se pueden encontrar en [31].

References

Needham, R., & Mcnealy, S. Multilateral Security.