Quantitative Assessment and Comparison of Cloud Service Providers' Privacy Practices

J. M. del Alamo, R. Trapero, Y. S. Martín, J. C. Yelmo and N. Suri

***Abstract*— The economic and technical advantages of cloud computing are widely recognized by the industry. However, the lack of knowledge on the privacy features offered by cloud service providers remains as one of the barriers for the adoption of cloud services. In this paper we describe a mechanism for the quantitative assessment of the privacy practices of different cloud service providers, so that cloud service clients can compare among them and choose the one that better fits their needs. Our contributions have been validated in three different scenarios.**

***Keywords*—Cloud Computing, Privacy, Privacy Metrics, Privacy Quantification, Privacy Level Agreement**

I. INTRODUCCIÓN

El modelo de computación en la nube (*Cloud Computing*) permite que los Proveedores de Servicios en la Nube (PSNs) ofrezcan sus servicios mediante protocolos de Internet, para que distintos Clientes de Servicios en la Nube (CSNs) puedan acceder a ellos de forma sencilla y rápida, sin necesidad de conocer ni de controlar la infraestructura en la que se apoyan. Este modelo tiene otras ventajas añadidas, como la facilidad de implementar modelos de cómputo bajo demanda o resolver problemas de escalabilidad, rendimiento y disponibilidad de forma transparente para los PSNs, que pueden delegar la gestión de la infraestructura en terceros y ocuparse únicamente de los detalles del negocio. Las bondades de este modelo han sido reconocidas por la industria, que en su mayoría está inmersa en un proceso de migración de servicios hacia la nube: según estudios recientes [1] un 69% de las empresas utilizan este modelo, y la tendencia es que esta cifra aumente en los próximos años.

Según aumenta el despliegue de servicios en la nube, crece también la preocupación entre los CSNs por la calidad de los servicios que usan y el desconocimiento de las prácticas de los PSNs que los ofrecen. Esta preocupación se dispara cuando se desvelan escándalos de filtración de datos por parte del proveedor del servicio, se conocen ataques sufridos por la infraestructura usada, etc. Por ello, en la actualidad, los dos principales factores que los clientes consideran a la hora de decidirse a utilizar los servicios ofrecidos por un PSN son la seguridad (82%) y la privacidad (81%) [2].

En algunos dominios, este problema ya ha sido resuelto mediante la firma de acuerdos de nivel de servicio (Service Level Agreement - SLA). Un SLA representa un compromiso entre un PSN y CSN mediante el cual a) se describen los servicios proporcionados, b) se documentan los objetivos del nivel de servicio que alcanzar, c) se incluyen las responsabilidades o penalizaciones de las partes en caso de que no se cumplan los objetivos a los que se han comprometido. Sin embargo, en el dominio de la privacidad, todavía no se ha abordado este tipo de soluciones, y apenas existen propuestas para evaluar y comparar las prácticas de privacidad de distintos proveedores.

Esta ponencia presenta una solución a los problemas planteados, mediante el diseño de un mecanismo de declaración de los niveles de privacidad ofrecidos por distintos proveedores y su comparación con los requisitos expresados por los clientes potenciales. En particular, las contribuciones de la ponencia son:

- Un análisis de los requisitos de privacidad aplicables a un PSN y su descomposición en distintos niveles de detalle.
- Un mecanismo para cuantificar el nivel de privacidad ofrecido por un PSN para cada requisito de manera individual, y para agrupaciones de requisitos.
- Un mecanismo de comparación entre los niveles de privacidad ofrecidos por varios PSNs y los requeridos por sus clientes.

Estas contribuciones han sido validadas en tres casos de estudio donde un cliente debe elegir, entre un conjunto de PSNs, aquel que mejor cumple sus requisitos de privacidad. Los clientes de cada caso expresan sus requisitos con distinto nivel de granularidad, lo que permite validar las características de los mecanismos de cuantificación y comparación desarrollados.

El resto de la ponencia se organiza de la siguiente forma. En la siguiente sección se ofrece una panorámica del estado del arte en métodos de cuantificación y comparación de los niveles de privacidad ofrecidos por proveedores de servicios. A continuación, la sección 3 describe los resultados del análisis de los requisitos de privacidad aplicables a PSNs, usando como referencia la legislación vigente en la Unión Europea. La sección 4 detalla el mecanismo diseñado para cuantificar requisitos de privacidad, y la sección 5 el mecanismo para evaluar del nivel de privacidad ofrecido por un PSN, en vista de los requisitos expresados por un cliente. La sección 6 describe la validación de las contribuciones presentadas. Finalmente se presenta un resumen de conclusiones y las líneas de trabajo futuro.

II. TRABAJO RELACIONADO

Los métodos cuantitativos para la evaluación y comparación

J.M. del Álamo, Universidad Politécnica de Madrid, Madrid, España, jmdela@dit.upm.es

R. Trapero, Technische Universität Darmstadt, Darmstadt, Alemania, rtrapero@cs.tu-darmstadt.de

Y.S. Martín, Universidad Politécnica de Madrid, Madrid, España, samuelm@dit.upm.es

J.C. Yelmo, Universidad Politécnica de Madrid, Madrid, España, jcyelmo@dit.upm.es

N. Suri, Technische Universität Darmstadt, Darmstadt, Alemania, suri@cs.tu-darmstadt.de

de las capacidades de distintos PSNs representan un área de investigación en auge, que hasta la fecha se ha centrado fundamentalmente en aspectos funcionales y de seguridad, y sólo recientemente ha comenzado a ocuparse de aspectos relacionados con la privacidad. Por ejemplo, ENISA (*European Network and Information Security Agency*) en la Unión Europea, NIST (*National Institute of Standards and Technology*) en EE.UU., o ISO (*International Organization for Standardization*) y CSA (*Cloud Security Alliance*) a nivel global, entre otros, llevan tiempo trabajando en la identificación de los objetivos de seguridad de los SLAs utilizados por los PSNs, conocidos como *Security Level Agreements* (SecLAs) [3] (se puede consultar un resumen y una comparativa de los resultados de las distintas organizaciones, por ejemplo, en [4]). Todos ellos definen los objetivos de seguridad en forma de descripciones textuales, similares a requisitos técnicos detallados. Estas descripciones para el dominio de la seguridad se han extendido al dominio de la privacidad con, por ejemplo, el catálogo de objetivos de privacidad recopilado por NIST [5].

El problema con estos objetivos de privacidad y seguridad, tal y como están enunciados, es que no es fácil medirlos de acuerdo con una métrica establecida, para comparar el nivel de protección ofrecido por distintas implementaciones. Por eso, varios grupos de trabajo, como el *Cloud Select Industry Group on Service-Level Agreements* [6] en la Unión Europea o el *Cloud Computing Reference Architecture and Taxonomy* [7] en EE.UU., han propuesto enfoques para transformar estas descripciones en otras evaluables de forma automática. Mientras que esos trabajos avanzan en el dominio de la seguridad, hasta donde conocemos los autores, este artículo es la primera publicación en la que se describe una metodología de cuantificación y comparación de los niveles de protección de la privacidad que ofrecen distintos proveedores.

Otro aspecto que se ha empezado a explorar recientemente en el dominio de la seguridad es cómo analizar la información contenida en los SLAs. Algunos trabajos se centran en el análisis puramente cualitativo de los niveles de seguridad que un proveedor es capaz de proporcionar [8]. Otros enfoques introducen una metodología cuantitativa para evaluar el nivel de seguridad proporcionado por un proveedor, mediante la asignación de pesos para ponderar la importancia de los objetivos de seguridad que se quieren evaluar [9]. Esto no sólo permite analizar el nivel de seguridad de un proveedor, sino también compararlo con otros, o con los requisitos de un usuario, y así decidir qué proveedores se acercan más o satisfacen completamente sus expectativas. Más recientemente, Taha *et al.* han mejorado esta última metodología para considerar requisitos de seguridad definidos con distintos niveles de abstracción [10]. Esto permite a los usuarios no técnicos definir niveles de importancia a nivel de agrupación de requisitos en lugar de tener que especificar valores concretos para cada requisito particular. Precisamente esta flexibilidad a la hora de definir requisitos de usuario determina que hayamos escogido esta metodología, adaptándola para poder aplicarla al dominio de la privacidad.

III. REQUISITOS DE PRIVACIDAD PARA PROVEEDORES DE SERVICIOS EN LA NUBE

Para poder describir el nivel de privacidad ofrecido por un PSN es necesario 1) identificar los requisitos de privacidad específicos que considerar, 2) decidir el nivel de privacidad comprometido para cada uno de ellos, 3) utilizar un formato de especificación que detalle los requisitos de privacidad seleccionados y el nivel comprometido.

Los requisitos de privacidad que se deben considerar dependen en gran medida de la legislación y/o normativa sectorial de aplicación. Por ejemplo, en la Unión Europea se han analizado los factores que los PSNs deben tener en cuenta cuando actúan como controladores de datos. A tal efecto, son aplicables los principios generales descritos en la Directiva de Protección de Datos [11] y la Directiva de Protección de Datos en el Sector de las Comunicaciones Electrónicas [12]. Fruto de este análisis se obtiene un conjunto de requisitos de alto nivel (principios de privacidad) necesarios para la protección de la privacidad en servicios en la nube [13][14]:

Confidencialidad: El PSN debe asegurar el acceso limitado a los datos personales.

Integridad: El PSN debe asegurar que los datos personales son auténticos y no han sido modificados de forma maliciosa ni alterados accidentalmente.

Disponibilidad: El PSN debe asegurar el acceso a los datos personales en tiempo y forma.

Calidad de los datos: El PSN debe asegurar la calidad de los datos que almacena, velando porque los mismos sean auténticos, completos, actualizados y veraces.

Especificación de propósito: El PSN debe especificar el propósito de los datos recogidos con antelación a la recogida.

Minimización de datos: El PSN debe asegurar que los datos recogidos son los mínimos imprescindibles para el cumplimiento del propósito establecido.

Limitación de uso: El PSN debe asegurar que los datos personales no serán usados para otro propósito que el acordado inicialmente con el sujeto de los datos, y sólo en la cantidad autorizada y por tiempo limitado.

Derechos de los usuarios: El PSN debe asegurar que los sujetos de los datos pueden ejercer sus derechos de acceso, rectificación, cancelación y oposición (derechos ARCO).

Portabilidad: El PSN debe asegurar que los CSNs podrán mover, copiar o transferir sus datos personales.

Transparencia: El PSN debe asegurar que todo el procesamiento de datos personales es entendido por los usuarios, y que puede ser reconstruido en cualquier momento.

Responsabilidad: El PSN debe poder demostrar que se han tomado las medidas adecuadas para garantizar que se respetan los principios anteriores.

Los principios de confidencialidad, integridad y disponibilidad (CID) son ampliamente conocidos en el mundo de la seguridad. De hecho, la aplicación exclusivamente de los principios CID a datos personales es una condición necesaria pero no suficiente para una adecuada gestión de la privacidad, que debe incluir también el resto de principios.

Los principios de privacidad derivados de la legislación describen requisitos de muy alto nivel, y por lo tanto alejados

de implementaciones particulares que puedan ser evaluables. Sin embargo, los principios generales se pueden refinar en distintos niveles jerárquicos mediante un proceso de operativización [15]. Así, un principio da lugar a una o varias pautas, que proporcionan objetivos más específicos que perseguir para satisfacer un principio, y éstas a su vez se descomponen en objetivos de privacidad específicos. Por ejemplo, el principio general de minimización de datos se puede refinar en pautas más específicas como: minimizar la información de identidad desvelada por un usuario del sistema, minimizar los datos de localización recopilados, etc. A su vez, la pauta de minimización de información de identidad se puede refinar en objetivos particulares como: garantizar el anonimato del usuario, impedir la trazabilidad de sus acciones, etc. Esta estructura jerárquica será útil para evaluar las prácticas de privacidad de distintos proveedores.

Una vez seleccionados el conjunto de objetivos de privacidad específicos que satisfacer, el PSN puede recurrir a distintas técnicas para lograrlo. En el campo de la privacidad, se ha desarrollado una intensa actividad investigadora para el desarrollo de herramientas, aplicaciones y mecanismos que contribuyen a resolver problemas específicos de privacidad, y que se agrupan bajo el acrónimo PET (*Privacy Enhancing Technology*). Por ejemplo, existen distintas técnicas que permiten que un usuario se identifique frente a un sistema, a la vez que minimizan la información desvelada a terceros (y por lo tanto reducen la trazabilidad), como Idemix [16] o U-Prove [17]. Hoy, existen varios catálogos de objetivos de privacidad y de técnicas para ayudar a satisfacerlos [5] (apéndice J) [18].

La correspondencia entre objetivos y técnicas de privacidad no es siempre unívoca, ya que distintas técnicas pueden satisfacer un objetivo; ni recíproca, puesto que se puede requerir combinar varias técnicas para satisfacer un objetivo. Evidentemente, la elección de una técnica u otra dependerá del contexto tecnológico y de negocio del PSN, e influirá en el nivel de privacidad alcanzado para cada uno de los objetivos de privacidad. La cuantificación de objetivos de privacidad se desarrolla en la siguiente sección.

Por otra parte, el uso de técnicas de diseño no específicas del dominio de la privacidad también pueden impactar en el nivel de privacidad de un objetivo. Efectivamente, si bien el mecanismo de obtención de la localización de un usuario no es una técnica específica de privacidad, la elección de un mecanismo u otro impacta en el anonimato de los usuarios del sistema. Por ejemplo, la obtención de la localización mediante GPS proporciona una posición muy precisa (en un rango del orden de metros cuadrados, prácticamente vinculada a un único usuario) mientras que si sólo se conoce el identificador del operador de red que ofrece conexión al usuario la precisión será mucho menor (el territorio de un país, en el orden de los cientos de miles de kilómetros cuadrados, compartida con millones de individuos). La tabla 1 muestra varios ejemplos de principios, pautas y objetivos de privacidad, así como ejemplos de técnicas que permiten satisfacerlos. Hay que destacar que, hasta donde conocemos los autores, esta clasificación de controles de privacidad es una de las primeras iniciativas realizadas hasta este momento sobre identificación y cuantificación de controles de privacidad.

Tabla 1. Ejemplo de principios, pautas y objetivos de privacidad; escalas de medida para cuantificar el nivel de privacidad de cada objetivo y rango de valores de la escala, y técnicas de diseño que se pueden utilizar para satisfacer los objetivos junto con el nivel de privacidad que aportan según la escala elegida.

Principio	Pauta	Objetivo	Descripción del objetivo	Escala de medida y rango de valores	Ejemplos de técnicas y niveles
Minimización de datos (MD)	Minimización de identidad (MI)	No trazabilidad (MI-01)	Impedir la vinculación de las acciones del usuario en distintos dominios	**Nivel de trazabilidad** 3) No trazable 2) Trazable por IdP y PSN en connivencia 1) Trazable por IdP 0) Trazable por cualquier IdP y PSN	Idemix (3) U-Prove (2) Certificados PKI (1) OAuth/OpenID (0)
		Anonimato (MI-02)	Garantizar que un usuario permanece anónimo dentro de un conjunto de usuarios	**Nivel de anonimato** Valor k-anonymity: 1 — 10^{10} personas	MNC de Operador: 10^6 pers. LCID de estación base: 10^3 pers. SSID de WiFi: 20 pers. Localización GPS: 1 pers.
	Minimización de localización (ML)	Granularidad de localización (ML-01)	Proteger la localización del usuario	**Nivel de granularidad** Valor l-diversity: $1m^2$—$10^{15}m^2$	MCC de Operador: 10^{11} m^2 LCID de estación base: 10^6 m^2 SSID de WiFi: 400 m^2 Localización GPS: 1 m^2
Transparencia (TR)	Claridad de información proporcionada (CI)	Facilitar la comprensión de las políticas (CI-01)	Ofrecer políticas de privacidad rápidas de comprender por el usuario	**Nivel de expresividad visual** 1) Alto 0) Bajo	Uso de iconos de privacidad (1) Uso de texto plano (0)
		Detallar el uso de la información personal (CI-02)	Ofrecer el nivel de detalle necesario en cada momento	**Nivel de detalle de las políticas** Número de capas: 1 — ∞	Políticas de privacidad estratificadas en varias capas: (>4) > 3 > 2 > 1
	Obtención de consentimiento (OC)	Utilización de mecanismos para permitir al usuario decidir (OC-01)	El usuario debe consentir la recolección, uso y procesamiento de sus datos	**Nivel de intervención del usuario** 2) Alto 1) Medio 0) Nulo	Aceptación explícita (2) Rechazo explícito (1) Aceptación implícita (0)
Derechos de los usuarios (DU)	Facilidad de ejercicio de derechos (FE)	Permitir al usuario un control centralizado (FE-01)	El usuario debe poder gestionar sus datos desde un único punto de control	**Centralización de control** 1) Si 0) No	Panel de control de privacidad (1) Controles dispersos (0)
Calidad de los datos (CD)	Exactitud de los datos personales (ED)	Comprobar la exactitud de los datos (ED-01)	La exactitud de los datos debe comprobarse periódicamente	**Periodo de comprobación** Número de meses: 1 — ∞ (Nunca)	Comprobación periódica de la exactitud de los datos: 1 mes > 3 m. > 6 m. > 9 m. (>12 m.)

Finalmente, el formato de especificación de los acuerdos y objetivos de privacidad debe facilitar su análisis y evaluación automáticos, de forma que se puedan comparar los niveles ofrecidos por distintos PSNs, o enfrentar el ofrecido por un PSN con el requerido por un CSN. Esta información se puede especificar en SLAs especializados, llamados acuerdos de nivel de privacidad (*Privacy Level Agreement* - PLA). Para ampliar información sobre formatos de descripción de PLAs el lector puede consultar, por ejemplo, [14].

La siguiente sección describe el proceso para cuantificar los objetivos de privacidad declarados por un PSN. Este es el paso previo necesario para poder evaluar el nivel de privacidad global que ofrece.

IV.CUANTIFICACIÓN DE OBJETIVOS DE PRIVACIDAD

Para poder comparar los niveles de privacidad ofrecidos por distintos PSNs, necesitamos crear un modelo que nos permita trasladar los niveles de privacidad ofrecidos por cada PSN a valores normalizados para usarlos como parámetros del algoritmo de evaluación. Dado que los objetivos de privacidad son requisitos descritos en formato textual, es necesario traducirlos a métricas operativas que sigan escalas de medida cuantificables.

En el dominio de la privacidad existen un conjunto de métricas que pueden ser de ayuda. Por ejemplo, *k-anonimity* [19] mide el nivel de anonimato de un individuo dentro de una población, representando *k* el tamaño de la población entre la que no es posible distinguir al individuo. Por su parte, *l-diversity* [20] es otra medida de anonimato que, en este caso, se fija en la granularidad de un dato personal conocido.

Frecuentemente, la implementación de una u otra técnica de privacidad, o, en general, de diseño, para satisfacer un objetivo de privacidad implica obtener un valor particular en alguna de estas métricas. Por ello, en muchas ocasiones es útil conocer qué técnicas ha implementado el PSN, para utilizar el valor típico asociado a esa técnica como una aproximación heurística de la medida, cuando ésta no es posible.

La tabla 1 incluye ejemplos de escalas de medida que se pueden usar para cuantificar el nivel de privacidad alcanzado en cada objetivo, junto con el rango de valores posibles; y ejemplos de técnicas que se emplean habitualmente para satisfacer cada objetivo de privacidad, junto con los valores típicos de privacidad asociados al uso de cada técnica.

El proceso de cuantificación de cada objetivo de privacidad es distinto en función del tipo de escala de medida que se utiliza para evaluarlo.

Dicotómica: La escala de medida sólo evalúa la presencia o ausencia de una característica, normalmente, una técnica particular que satisface el objetivo. Por ejemplo, para facilitar el ejercicio de derechos ARCO a los usuarios (pauta de privacidad FE) el PSN debería disponer de un punto de control centralizado (objetivo de privacidad FE-01), normalmente implementado como un panel de control de privacidad. La cuantificación del objetivo FE-01 en función de si existe o no el punto de control centralizado da lugar a los valores *true* o *false*, o directamente 1 y 0. Algo análogo ocurre cuando se usan iconos de privacidad en las políticas de privacidad declaradas por un PSN, ya que facilitan la rápida comprensión de la información proporcionada (objetivo CI-01).

Ordinal: La escala de medida ofrece un rango finito de valores ordenados, entre los que no existe un concepto de distancia, y estando normalmente cada valor asociado al uso de una técnica particular para satisfacer el objetivo de privacidad. Por ejemplo, para el objetivo de OC-01 se puede medir el nivel de intervención del usuario durante el proceso. La literatura describe un conjunto de técnicas utilizadas para ello, concluyendo que la aceptación explícita del consentimiento (por ejemplo, mediante la pulsación de un botón) siempre es mejor que requerir el rechazo explícito a la recogida y uso, y ésta a su vez mejor que la asunción automática del consentimiento sin preguntar siquiera al usuario. Por tanto, esta métrica tiene tres niveles posibles, que pueden normalizarse como {2, 1, 0}, y ordenarse como 2>1>0. Lo mismo ocurre si se cuantifica el objetivo de no trazabilidad de los usuarios (MI-01), ya que existen en el estado del arte distintas técnicas que permiten reducir la trazabilidad del usuario, ordenables en función de su bondad.

Continua: La escala de medida ofrece un rango de valores entre los que tiene sentido formular el concepto de distancia. Por ejemplo, podemos obtener el valor *k* usando *k-anonimity*, obteniendo valores en el rango 1 y 10^{10} (la población mundial) donde k=1 indica que existe una correspondencia unívoca entre usuario y localización, y k=10^{10} indica que es imposible distinguir a un usuario de cualquier otro. Algo similar ocurre con la métrica *l-diversity*, donde obtenemos una medida de la granularidad de la localización, desde l=1 m^2 hasta l=10^{15} m^2 (la superficie de la Tierra).

Para poder manejar con facilidad este tipo de métricas, las transformamos en ordinales mediante la definición de un conjunto limitado de valores ordenados que, de nuevo, se pueden asociar con el uso de técnicas particulares. De esta forma, el nivel de anonimato (objetivo MI-02) y de granularidad de localización (objetivo ML-01) ofrecidos por un PSN dependerán de si, para localizar a un usuario, utiliza la identificación de la red (MNC - *Mobile Network Code*) o del país (MCC - *Mobile Country Code*) del operador que da servicio al usuario (k=10^6 o l=10^5 km^2), el identificador de celda (LCID - *UMTS Cell ID*) de la estación base que da servicio al usuario (k=1000 o l=1 km^2), el identificador de la red WiFi (SSID - *Service Set Identifier*) a la que el usuario está conectado (k=20 o l=400 m^2), o las coordenadas GPS del usuario (k=1 o l=1 m^2). Por tanto, estos objetivos de privacidad pueden adoptar cuatro niveles posibles que pueden modelarse como {3, 2, 1, 0}, de forma que 3>2>1>0.

Otros casos de métricas continuas son el número de niveles de la política de privacidad (aplicado para medir el nivel del objetivo CI-02), o la frecuencia de comprobación de datos personales (aplicada para cuantificar la comprobación periódica de la exactitud de los datos, para el objetivo ED-01).

Como resultado del proceso de cuantificación, cada PSN obtiene valores concretos a nivel de objetivos de privacidad. El proceso se puede repetir para todos los PSNs, de modo que al final se disponga de un catálogo de PLAs con valores distintos por cada PSN. Este tipo de catálogos ya existen en el

dominio de la seguridad [21], aunque aún no en el dominio de la privacidad.

V. EVALUACIÓN DE PRIVACIDAD

El proceso de evaluación del nivel de privacidad ofrecido por un PSN parte de los valores resultantes de la cuantificación de los objetivos de privacidad que implementa el PSN y de la estructura jerárquica de los principios, pautas y objetivos descrita anteriormente. A partir de esta información, realizamos un análisis jerárquico mediante el empleo del algoritmo *Quantitative Hierarchy Process* (QHP) [10].

El empleo de QHP como algoritmo de evaluación nos permite beneficiarnos de muchas de sus principales características, entre otras, la posibilidad de manejar tanto objetivos de privacidad cualitativos como cuantitativos –para realizar un análisis más o menos granular en función de los requisitos de los CSNs–, o la flexibilidad a la hora de identificar inconsistencias entre objetivos de privacidad.

En efecto, la evaluación del cumplimiento del nivel de privacidad de un PSN puede realizarse a distintos niveles de la estructura jerárquica, de forma que es posible seleccionar a qué ámbito se quiere restringir el análisis. El nivel más alto en la jerarquía es la raíz: Un análisis a ese nivel proporciona un nivel de privacidad global y permite, entre otras cosas, establecer un ranking de PSNs.

Por otra parte, la estructura jerárquica proporciona un modelo abstracto sobre el cual un CSN puede expresar sus necesidades de privacidad para que se puedan comparar con las características ofertadas por distintos proveedores y elegir el más adecuado. En nuestro caso consideramos tres mecanismos de expresión de necesidades de privacidad.

CSNs expertos en privacidad. Pueden ser usuarios finales u otros PSNs consumidores de sus servicios que podrán especificar los niveles concretos que esperan tener a nivel de objetivo de privacidad. Estos CSNs deben tener un conocimiento profundo para poder adoptar soluciones de compromiso entre objetivos, a la vez que considerar otro tipo de requisitos funcionales y no funcionales. Por ejemplo, si un CSN quiere que sus acciones en distintos dominios no puedan ser vinculadas entre sí exigirá que el PSN ofrezca un nivel 1 para el objetivo MI-01, aunque esta decisión es incompatible con requerir al sistema la trazabilidad de las acciones de otros usuarios por motivos de seguridad.

CSNs con pocos conocimientos en privacidad. Este caso está pensado para usuarios finales que no disponen de un gran conocimiento técnico, y por tanto no son capaces de especificar valores concretos para los objetivos de privacidad. En este caso, el enfoque es cualitativo, definiéndose un nivel de importancia determinado a cualquier nivel de la jerarquía. Por ejemplo, un usuario podría determinar que el objetivo de privacidad CI-01 es *Muy Importante* para él. El número de posibles niveles de importancia es configurable, de modo que se puede aumentar o disminuir el nivel de granularidad del análisis: cuanto más niveles de importancia se tengan, más preciso será el análisis, pero más difícil será para el usuario estimar cómo de importante es un objetivo. Los niveles de importancia pueden determinarse usando tanto etiquetas textuales (*Muy Importante*, *Importante*, *No Importante*) como valores numéricos en una escala (por ejemplo, una importancia de 2 sobre 10).

Igual que se hizo al cuantificar los objetivos de privacidad, los niveles de importancia se cuantificarán también según una escala numérica, para que el algoritmo pueda procesarlos. Usaremos la escala *Muy Importante* > *Importante* > *No Importante*, que se cuantifica como {3, 2, 1}, donde 3 es el valor para *Muy Importante* y 1 el valor para *No Importante*.

Especificación de requisitos a distintos niveles. El algoritmo de evaluación permite combinar los dos enfoques anteriores. Esto permite que un CSN pueda detallar valores concretos para un subconjunto de objetivos de privacidad que es relevante para él, mientras que otro subconjunto de objetivos puede no ser tan relevante y le baste con especificar valores cualitativos a nivel de principio o pauta de privacidad.

Una vez que se dispone de los valores para los objetivos de privacidad ofrecidos por los PSNs y de los requisitos de privacidad expresados por los CSNs, se comparan los valores por pares. Por cada PSN, se compara el valor cuantificado de cada objetivo con el valor declarado por otro PSN o con el valor solicitado por el CSN (dependiendo de si se quiere comparar varios PSNs entre sí o si se quiere comparar un PSN respecto a los requisitos de un CSN). El resultado es una matriz de comparación (MC) por pares, de forma que para un objetivo C, PSN_1/PSN_2 se corresponde con el valor relativo de la métrica. Esto da lugar a una matriz de tamaño $n \times n$, donde n es el número total de PSNs que se comparan.

$$MC = \begin{pmatrix} PSN_1/PSN_1 & PSN_1/PSN_2 & \dots & PSN_1/PSN_n \\ PSN_2/PSN_1 & PSN_2/PSN_2 & \dots & PSN_2/PSN_n \\ \dots & \dots & \dots & \dots \\ PSN_n/PSN_1 & PSN_n/PSN_2 & \dots & PSN_n/PSN_n \end{pmatrix}$$

Para cada objetivo se puede extraer un ranking de todos los PSNs partiendo de la matriz de comparación. Para ello, se calcula el autovector principal normalizado de dicha matriz, o Vector de Prioridad (VP), que contiene los valores que representan la puntuación de cada PSN con respecto a cada objetivo. Para más detalles sobre el cálculo de VPs, remitimos a la fuente de la metodología QHP [10] o a la metodología para la toma de decisiones basada en análisis jerárquico [22].

Una vez que se tienen los VPs para cada objetivo se realiza una ponderación con los valores de importancia definidos por los CSNs. Como los CSNs han podido definir valores de importancia a cualquier nivel de la jerarquía, la agregación comienza desde el nivel más bajo hacia arriba, obteniendo puntuaciones parciales a medida que se van aplicando los niveles de importancia definidos por el CSN (cuantificados según las reglas definidas anteriormente).

De esta manera se obtiene el vector de prioridad agregado, $\mathrm{VP}_{agregado} = (\mathrm{VP}_1, \mathrm{VP}_2 \dots \mathrm{VP}_n)\,(\mathrm{w}_i)$, siendo w_i el nivel de importancia asignado por el usuario para el objetivo i y VP_n el vector de prioridad para cada PSN_n, siendo n el número de proveedores que se comparan. El vector agregado contiene las puntuaciones de cada PSN con respecto a las preferencias del CSN, pudiendo así decidir qué PSN cumple mejor.

Todo el proceso anteriormente descrito puede utilizarse para obtener resultados analíticos a varios niveles en forma de

comparaciones, tanto globales como a cualquier nivel de la jerarquía, o para un grupo concreto de objetivos. La siguiente sección detalla la validación de este modelo, aplicándolo a tres escenarios, cada uno con tres PSNs distintos.

VI. VALIDACIÓN

Para validar la metodología propuesta, disponemos de una configuración con tres PSNs que declaran las características de privacidad que ofrecen. Hemos aplicado la metodología en tres casos distintos: en cada uno, un CSN expresa sus requisitos de privacidad con distinto nivel de granularidad, y se evalúan todos los PSNs para encontrar el que mejor cumple con sus requisitos. Describimos a continuación los datos utilizados, los resultados obtenidos y su análisis.

A. Características de proveedores y requisitos de clientes

La tabla 2 muestra los requisitos de los tres CSNs. Consideramos que los usuarios tienen distintos niveles de conocimientos sobre privacidad. El CSN_1 es un cliente experto que especifica requisitos a nivel de objetivo de privacidad. El CSN_2 es un cliente con requisitos específicos para el principio DU, y por ello detalla valores concretos para el objetivo FE-01. Para el resto de requisitos no tiene especial interés o conocimientos, fijando como *Nada Importante* (-) el principio CD, *Importante* (~) el principio TR y la pauta MI, y *Muy Importante* (+) la pauta ML. Por último, el CSN3 es un cliente sin necesidades estrictas de privacidad, y expresa requisitos cualitativos a varios niveles (a nivel de principios MD, DU y CD, y detallando las pautas de privacidad CI y OC).

Tabla 2. Requisitos de privacidad de tres clientes.

Principio	Pauta	Objetivo	CSN_1	CSN_2	CSN_3
MD	MI	MI-01	*U-Prove*	~	~
		MI-02	*20 pers.*		
	ML	ML-01	*1km²*	+	
TR	CI	CI-01	*Alto*	~	+
		CI-02	*3*		
	OC	OC-01	*Medio*		~
DU	FE	FE-01	*Sí*	*Sí*	-
CD	ED	ED-01	*3 meses*	-	-

Por otra parte, la tabla 3 muestra el conjunto específico de valores de cada PSN para cada objetivo de privacidad. Cada conjunto de valores para cada PSN se corresponde con el PLA que ese proveedor es capaz de proporcionar.

Tabla 3. Niveles de privacidad garantizados por tres proveedores.

Objetivo	PSN_1	PSN_2	PSN_3
MI-01	*Trazable por IdP*	*No trazable*	*Trazable por IdP-PSN*
MI-02	*20 pers.*	*1000 pers.*	*20 pers*
ML-01	*1km2*	*400m2*	*1km2*
CI-01	*Alto*	*Bajo*	*Alto*
CI-02	*1*	*2*	*2*
OC-01	*Nulo*	*Medio*	*Nulo*
FE-01	*No*	*Si*	*Si*
ED-01	*3 meses*	*1 mes*	*3 meses*

B. Resultados cuantitativos

Una vez obtenida la descripción de las características de privacidad ofrecidas por cada PSN y de los requisitos de los CSNs, se cuantifican los valores según el procedimiento descrito más arriba, se comparan los valores y se obtienen los vectores de prioridad para los tres casos analizados. A continuación detallamos los resultados obtenidos (se omiten los cálculos detallados por razones de espacio y simplicidad).

1) Caso 1: CSN_1

Primero, obtenemos los VPs para cada uno de los objetivos de privacidad descritos, es decir, en el nivel más bajo de la jerarquía (tabla 4). Podemos ver como, por ejemplo, para la propiedad $VP_{ML\text{-}01}$ el CSN_1 ve cumplido su requisito por parte del PSN_1 y PSN_3, pero no por el PSN_2.

Tabla 4. Vectores de prioridad para los objetivos de privacidad del CSN1.

	PSN_1	PSN_2	PSN_3	CSN_1
$VP_{MI\text{-}01}$	0.1818	0.2727	0.2727	0.2727
$VP_{MI\text{-}02}$	0.2727	0.1818	0.2727	0.2727
$VP_{ML\text{-}01}$	0.2727	0.1818	0.2727	0.2727
$VP_{CI\text{-}01}$	0.3077	0.0769	0.3077	0.3077
$VP_{CI\text{-}02}$	0.2500	0.2500	0.2500	0.2500
$VP_{OC\text{-}01}$	0.2500	0.2500	0.2500	0.2500
$VP_{FE\text{-}01}$	0.0000	0.3333	0.3333	0.3333
$VP_{ED\text{-}01}$	0.2500	0.2500	0.2500	0.2500

Se puede hacer el mismo cálculo para los niveles de principio y pauta y obtener valores similares. Omitimos dichos valores por motivos de espacio, y detallamos simplemente las puntuaciones globales en la raíz de la jerarquía. Este VP nos proporciona una puntuación total de cómo cada PSN cumple con los requisitos globales del CSN.

$$\begin{matrix} & PSN_1 & PSN_2 & PSN_3 & CSN_1 \end{matrix}$$
$$VP_{raíz} = (0.1911 \quad 0.2487 \quad 0.2801 \quad 0.2801)$$

El resultado es que el proveedor que mejor cumple con los requisitos de este cliente es el PSN_3.

2) Caso 2: CSN_2

Aplicando el algoritmo, de forma similar, obtenemos los VPs en aquellos niveles en donde el usuario ha especificado requisitos (recordemos que este CSN ha especificado requisitos cualitativos en algunas partes de la jerarquía de requisitos de privacidad).

Tabla 5. Vectores de prioridad para el CSN_2 a distintos niveles.

	PSN_1	PSN_2	PSN_3	CSN_2	*Nivel de evaluación*
VP_{MI}	0.2368	0.2368	0.2632	0.2632	Pauta
VP_{ML}	0.2500	0.1667	0.2500	0.3333	Pauta
VP_{TR}	0.2676	0.1972	0.2676	0.2676	Principio
$VP_{FE\text{-}01}$	0.0000	0.3333	0.3333	0.3333	Objetivo
VP_{CD}	0.2500	0.2500	0.2500	0.2500	Principio

Para este caso, el vector de prioridad global resulta:

$$\begin{matrix} & PSN_1 & PSN_2 & PSN_3 & CSN_2 \end{matrix}$$
$$VP_{raíz} = (0.1762 \quad 0.2438 \quad 0.2856 \quad 0.2994)$$

De nuevo, el PSN_3 es el que mejor cumple con los requisitos de este CSN.

3) Caso 3: CSN_3

Aquí, los VPs reflejan los requisitos del CSN, expresados de forma cualitativa en todos los niveles de la jerarquía.

$$\begin{matrix} & PSN_1 & PSN_2 & PSN_3 & CSN_3 \end{matrix}$$
$$VP_{raíz} = (0.2650 \quad 0.2009 \quad 0.2626 \quad 0{,}2715)$$

Ahora, el proveedor que mejor cumple con los requisitos del cliente es el PSN_1 seguido muy de cerca por el PSN_3.

C. *Análisis comparativo*

Los resultados numéricos determinan de forma precisa qué PSN satisface de mejor forma los requisitos fijados por un CSN. Sin embargo, los resultados numéricos pueden ser difíciles de interpretar, sobre todo a la hora de comparar de manera rápida las distintas capacidades de privacidad que ofrecen múltiples proveedores a distintos niveles.

La Fig. 1 muestra el resultado del análisis al nivel más alto de la jerarquía. Como se ha comentado anteriormente, esto nos permite conocer cuánto se aproxima el PLA de cada PSN a los requisitos de privacidad de los CSNs. El resultado es que para el CSN_1 y CSN_2 el mejor proveedor es el PSN_3, mientras que el mejor proveedor para el CSN_3 es el PSN_1.

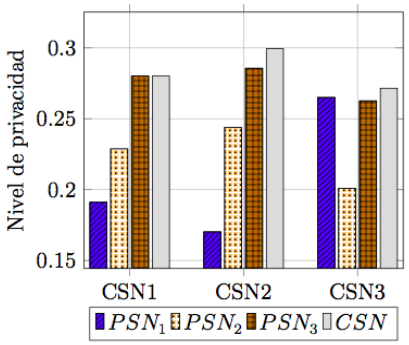

Figura 1. Nivel de privacidad global de cada PSN con respecto a cada CSN.

Es interesante analizar los resultados parciales a distintos niveles de la jerarquía. Por ejemplo, los resultados obtenidos para el CSN_1 a nivel de principio de privacidad aparecen en la Fig. 2. Esto permite conocer de forma agregada qué proveedores se comportan mejor que otros en aspectos concretos de la privacidad. Por ejemplo, cabe destacar que el proveedor PSN_1 se comporta realmente mal en el principio DU. Esto se debe a que, mientras el cliente ha especificado a ese nivel un valor "Alto" para el objetivo FE-01, lo que implica el uso de un panel de control para gestionar sus datos desde un punto centralizado, el proveedor, sin embargo, no lo proporciona y, por tanto, no es capaz de satisfacer el requisito.

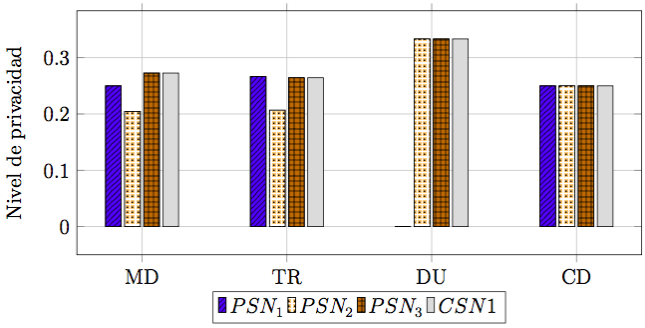

Figura 2. Nivel de privacidad de cada PSN con respecto a los requisitos del CSN_1 a nivel de pautas de privacidad.

Los datos obtenidos con el análisis nos permiten también representar la información, para conocer exactamente qué aspectos de privacidad están por debajo o por encima de las necesidades de los CSNs. La Fig. 3 muestra esta información, al nivel de pautas de privacidad, para el CSN_1. En la parte izquierda se puede ver que el proveedor PSN_1 no está satisfaciendo las necesidades del usuario en la pauta MI (ya que no consigue satisfacer el nivel deseado para el objetivo MI-01). En el caso del PSN_2 (centro) son las pautas CI, ML y MI las que no se están satisfaciendo, mientras que en el caso del PSN_3 (derecha) las características de privacidad ofrecidas cumplen perfectamente con las necesidades del cliente.

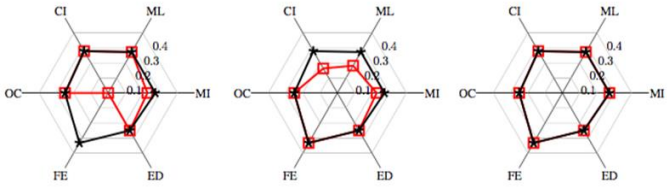

Figura 3. Análisis a nivel de pauta de privacidad: PSN_1 vs CSN_1 (izquierda), PSN_2 vs CSN_1 (centro) y PSN_3 vs CSN_1 (derecha). En rojo los valores proporcionado por el PSN y en negro los valores requeridos por el CSN.

VII. CONCLUSIONES Y TRABAJO FUTURO

En esta ponencia hemos abordado la problemática que supone la falta de información sobre las prácticas de privacidad de distintos proveedores de servicios en la nube. Para ello, hemos analizado primero los requisitos de privacidad que se aplican en la Unión Europea, y cómo descomponer estos requisitos en distintos niveles de refinamiento hasta llegar a requisitos fácilmente observables. Luego, hemos descrito un mecanismo para cuantificar el nivel de privacidad aportado por cada uno de los requisitos, tanto de forma individual como por agrupaciones de requisitos. Finalmente, hemos descrito un mecanismo de evaluación de las prácticas de privacidad de varios proveedores de servicios en la nube mediante la comparación con las necesidades expresadas por un potencial cliente. Todas estas contribuciones han sido validadas en múltiples escenarios.

Nuestros próximos pasos apuntan a la mejora de algunas de las características descritas en esta ponencia. En particular, estamos trabajando en la introducción de mecanismos de selección de requisitos por parte de los potenciales clientes que encuentran dificultades para expresar sus expectativas de privacidad, y en la utilización de lógica difusa para especificar requisitos y poder cuantificarlos teniendo en cuenta su incertidumbre.

AGRADECIMIENTOS

Esta ponencia presenta resultados parciales de los proyectos PRIPARE y SPECS, financiados por el Séptimo Programa Marco de la Unión Europea bajo los acuerdos de financiación número 610613 y 610795, respectivamente.

REFERENCIAS

[1] IDG, "Enterprise Cloud Computing Study", 2014. Disponible en línea en http://www.idgenterprise.com/report/idg-enterprise-cloud-computing-study-2014

[2] KPMG, "Cloud Survey Report: Elevating Business in the Cloud", 2014. Disponible en línea en https://www.kpmg.com/PL/pl/IssuesAndInsights/ArticlesPublications/Documents/2015/2014-KPMG-Cloud-Survey-Report-online-secured.pdf

[3] M. Dekker and G. Hogben, "Survey and analysis of security parameters in cloud SLAs across the European public sector". Technical Report TR-2011-12-19, European Network and Information Security Agency, 2011.

[4] Cloud Security Alliance, “Cloud Control Matrix v3.0.1”, 2014. Disponible en línea en https://cloudsecurityalliance.org/research/ccm/
[5] “Security and Privacy Controls for Federal Information Systems and Organizations”, NIST Special Publication 800-53, revision 4, April 2013
[6] “Cloud Service Level Agreement Standardisation Guidelines,” European Commission, C-SIG SLA, Tech. Rep. C-SIG SLA 2014, 2014.
[7] “(Draft) Cloud Computing: Cloud Service Metrics Description,” National Institute of Standards and Technology, Tech. Rep. NIST Public RATAX WG, 2014.
[8] M. Almorsy, J. Grundy and A. Ibrahim, “Collaboration-Based Cloud Computing Security Management Framework”, in Proc. of IEEE Intl Conference on Cloud Computing, pp. 364–371, 2011.
[9] J. Luna, H. Ghani, T. Vateva and N. Suri, “Quantitative Assessment of Cloud Security Level Agreements: A Case Study”, in Proc. of Int. Conf. on Security and Cryptography, pp. 64-73, 2012.
[10] A. Taha, R. Trapero, J. Luna and N. Suri, “AHP-Based Quantitative Approach for Assessing and Comparing Cloud Security,” in Proc. of the IEEE Conference on Trust, Security and Privacy in Computing and Communications, pp. 284–291, 2014.
[11] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. On the protection of individuals with regard to the processing of personal data and on the free movement of such data. L, 281:0031–0050, November 23 1995.
[12] Directive 2002/58/EC of the European Parliament and the Council of 12 July 2002, Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector. Official J. EU, vol. L 201, July 2002.
[13] Article 29 Data Protection Working Party, "Opinion 05/2012 on Cloud Computing", 01037/12/EN WP 196, 2012.
[14] CSA, "Privacy Level Agreement Outline for the Sale of Cloud Services in the European Union", Cloud Security Alliance, Privacy Level Agreement Working Group, 2013.
[15] N. Notario, A. Crespo, Y.S. Martín, J.M. Del Alamo, D. Le Métayer, T. Antignac, A. Kung, I. Kroener and D. Wright, “PRIPARE: Integrating Privacy Best Practices into a Privacy Engineering Methodology”, in 1st Int. Workshop on Privacy Engineering (IWPE15), 2015 (In press).
[16] J. Camenisch and E. Van Herreweghen, “Design and implementation of the idemix anonymous credential system”, in Proceedings of the 9th ACM conference on Computer and communications security, pp. 21-30, 2012.
[17] C. Paquin, “U-Prove Technology Overview V1.1 (Revision 2)”, April 2013. Disponible en línea en http://research.microsoft.com/apps/pubs/default.aspx?id=166980
[18] W. Fumy, M. De Soete, E. J. Humphreys, T. Chikazawa, J. Amsenga, and K. Rannenberg, “ISO/IEC 27018:2014 - Information technology -- Security techniques -- Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors,” vol. 8, no. attachment 1. 2014.
[19] L. Sweeney, “k-anonymity: A model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557–570, 2002.
[20] A. Machanavajjhala et al. “L-diversity: Privacy beyond k-anonymity”, ACM Trans. Knowl. Discov. Data, 1(1), March 2007.
[21] Cloud Security Alliance, ”Security, Trust & Assurance Registry (STAR),” 2011. Disponible en línea en https: //cloudsecurityalliance.org/star/
[22] K. Garg, S. Versteeg, and R. Buyya, “A framework for ranking of cloud computing services,” Journal of Future Generation Computer Systems, pp. 1012–1023, 2013.

José M. del Álamo es Doctor Ingeniero de Telecomunicación (2009) y profesor (2011) en el Departamento de Ingeniería de Sistemas Telemáticos (DIT) de la Universidad Politécnica de Madrid (UPM). Sus líneas de investigación incluyen la gestión de datos personales, incluyendo gestión de identidad y privacidad, y su introducción en las metodologías de ingeniería de software y sistemas.

Rubén Trapero es doctor por la UPM (2010). EN 2012 se unió como investigador postdoctoral a la Universidad Carlos III de Madrid, y desde 2014 es investigador postdoctoral en la Technische Universität Darmstadt (Alemania). Sus líneas de interés son la gestión de la identidad y privacidad, seguridad en cloud, análisis de riesgos e ingeniería de servicios.

Yod Samuel Martín (S'14) es Ingeniero de Telecomunicación (2004) por la UPM, donde ha trabajado como investigador en el DIT y en el Center for Open Middleware desde 2004. En la actualidad, sus líneas de investigación se centran en la introducción de requisitos no funcionales en los servicios telemáticos, con especial énfasis en la accesibilidad y la privacidad.

Juan C. Yelmo es Doctor Ingeniero de Telecomunicación (1996) y Profesor Titular de Universidad (1998) en el Departamento de Ingeniería de Sistemas Telemáticos de la Universidad Politécnica de Madrid. El profesor Yelmo ha desarrollado una intensa y continuada actividad investigadora caracterizada por la transferencia y publicación de resultados de investigación mediante artículos en revistas de primer nivel mundial en su sector y ponencias en congresos de relevancia internacional y patentes internacionales. Sus líneas de investigación actuales incluyen la ingeniería de servicios, la gestión de la identidad y la privacidad y el modelado de usuario en servicios y redes sociales.

Neeraj Suri es Doctor por la Universidad de Massachusetts en Amherst. Actualmente es catedrático en la Technische Universität Darmstadt, Alemania. También está afiliado con la universidad de Texas-Austin y con Microsoft Research. Sus puestos anteriores incluyen director de la Cátedra Saab y profesor en la Universidad de Boston. Sus intereses incluyen los sistemas distribuidos, computación móvil, y sistemas operativos enfocándose en el diseño, análisis y evaluación de servicios confiables en la Web a gran escala.