(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 618 807**

(51) Int. CI.:

H04L 29/08 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **10.11.2010 PCT/EP2010/067210**

(87) Fecha y número de publicación internacional: **19.05.2011 WO2011058057**

(96) Fecha de presentación y número de la solicitud europea: **10.11.2010 E 10788263 (1)**

(97) Fecha y número de publicación de la concesión europea: **22.02.2017 EP 2486719**

(54) Título: **Elementos de información concordantes**

(30) Prioridad:

10.11.2009 GB 0919675

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
22.06.2017

(73) Titular/es:

SKYPE (100.0%)
70 Sir John Rogerson's Quay
Dublin 2, IE

(72) Inventor/es:

KAAL, MADIS

(74) Agente/Representante:

UNGRÍA LÓPEZ, Javier

DESCRIPCIÓN

Elementos de información concordantes

Campo de la invención

Esta invención se refiere a hallar elementos de información concordantes en una red.

Antecedentes

Una red incluye típicamente una pluralidad de nodos que pueden comunicar uno con otro, de tal manera que cada nodo en la red sea capaz de comunicar con al menos otro nodo en la red. La red puede ser por ejemplo Internet, pero se puede usar alternativa o adicionalmente otras redes. Los nodos pueden comunicar en la red usando enlaces que pueden ser enlaces directos entre los nodos, o alternativamente los enlaces pueden ser enlaces indirectos a través de la red, de tal manera que los nodos comuniquen uno con otro mediante al menos otro nodo en la red.

La red puede emplear un sistema de comunicaciones a base de paquetes. Los sistemas de comunicaciones a base de paquetes permiten al usuario de un dispositivo, tal como un ordenador personal, comunicar a través de la red. Un tipo de sistema de comunicaciones a base de paquetes usa una topología entre iguales ("P2P") incorporada en protocolos de propiedad. Para permitir el acceso a un sistema entre iguales, el usuario debe ejecutar software de cliente P2P proporcionado por un proveedor de software P2P en su ordenador, y registrarse en el sistema P2P. Cuando el usuario se registra en el sistema P2P, un servidor le suministra el software de cliente con un certificado digital. Una vez que el software de cliente ha sido provisto del certificado, se puede establecer posteriormente y dirigir la comunicación entre usuarios del sistema P2P sin el uso adicional de un servidor. En particular, los usuarios pueden establecer sus propias rutas de comunicación a través del sistema P2P en base al intercambio de uno o más certificados digitales (o certificados de identidad de usuario, "UIC"), que permiten el acceso al sistema P2P. El intercambio de los certificados digitales entre usuarios proporciona prueba de las identidades de los usuarios y de que están adecuadamente autorizados y autenticados en el sistema P2P para entrar en comunicación. Otros detalles acerca de tal sistema P2P se describen en WO 2005/009019.

Un primer nodo en la red puede almacenar detalles de contactos de un primer usuario del primer nodo. El primer nodo, o el primer usuario, puede estar interesado en determinar si algunos de los contactos del primer usuario también son contactos de un segundo usuario en un segundo nodo en la red. Un método para determinar si hay contactos comunes entre los usuarios primero y segundo es que el primer nodo envíe una lista de los contactos del primer usuario al segundo nodo. El segundo nodo puede comparar entonces los contactos de los usuarios primero y segundo para hallar contactos comunes y puede devolver los resultados al primer nodo.

En otro escenario, puede usarse un sitio de red local en la red para hallar en la red usuarios que comparten intereses comunes. Los usuarios pueden transmitir identificadores de sus intereses a un nodo del sitio de red local, de modo que el sitio de red local pueda comparar los intereses de usuarios diferentes para identificar usuarios con intereses comunes. Entonces se puede enviar una recomendación a al menos uno de los usuarios identificados para indicar que los usuarios identificados comparten intereses comunes.

En otro escenario en el que la red es una red de compartición de archivos, algunos nodos en la red son nodos de compartición de archivos, por lo que los archivos almacenados en los nodos de compartición de archivos pueden ser recuperados por otros nodos en la red. Cada nodo de compartición de archivos envía identificadores, tal como nombres de archivo, de los archivos almacenados en el nodo de compartición a un nodo de índice central. Un usuario que desee hallar un archivo deseado en la red de compartición de archivos puede enviar una petición al nodo de índice central identificando el archivo deseado. El nodo de índice central puede comparar la petición con los identificadores recibidos de los nodos de compartición para hallar las concordancias. Si el nodo de índice central determina que el archivo deseado está almacenado en un nodo de compartición, entonces el nodo de índice central puede informar al usuario que busca el archivo deseado acerca de la posición del nodo de compartición en la red. El usuario puede contactar entonces con el nodo de compartición para pedir el archivo deseado.

El autor de la invención ha observado que hay dos problemas comunes a los escenarios separados descritos anteriormente. En primer lugar, la transmisión de los detalles de contactos, los identificadores de intereses o los identificadores de archivos por la red requiere que se transmita una gran cantidad de datos a través de la red. Esto reduce los recursos de red que están disponibles para otros fines. En segundo lugar, puede haber un problema de seguridad o privacidad en los escenarios descritos anteriormente. Por ejemplo, es posible que un primer usuario no desee revelar sus contactos a un segundo usuario. Como otro ejemplo, los usuarios podrían no querer revelar sus intereses a un sitio de red local. Además, un nodo de compartición puede no desear revelar los archivos almacenados en el nodo de compartición a un nodo de índice central y también un usuario que busca un archivo deseado puede no querer revelar el archivo deseado al nodo de índice central.

SHIN-YAN CHIOU y colaboradores: "Common friends discovery with privacy and authenticity", INFORMATION ASSURANCE AND SECURITY, 2009. IAS '09. FIFTH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY,

NJ, Estados Unidos de América, 18 agosto 2009 (2009-08-18), páginas 337-340, XP031544547, ISBN: 978-0-7695-3744-3 describe un algoritmo para una aplicación de red no centralizada que permite a usuarios de una red social ver amigos que son comunes a cada usuario.

WO2007106791 describe una configuración P2P para un centro de llamadas entrantes usando una tabla hash distribuida (DHT) para enrutar llamadas a nodos en la red del centro de llamadas.

WO2008041173 describe incorporar a la eficiencia de DHT el más amplio alcance de un algoritmo de inundación de modo que una petición de contenido por parte del usuario pueda ser compartida en una red social.

US20040148506 describe un algoritmo hashing unidireccional para evitar que las compañías envíen mensajes electrónicos no deseados a abonados.

MARCO VON ARB y colaboradores: “VENETA: Serverles friend-of-friend detection in mobile social networking”, NETWORKING AND COMMUNICATIONS, 2008. WIMOB ’08. IEEE INTERNATIONAL CONFERENCE ON WIRELSS AND MOBILES COMPUTING, IEEE, PISCATAWAY, NJ, Estados Unidos de América, 12 octubre 2008 (2008-10-12), páginas 184-189,XP031350713,ISBN: 978-0-7695-3393-3 describe una aplicación para redes móviles que asigna valores hash a listas de contactos del usuario de modo que los contactos comunes puedan coincidir y los usuarios pueden ser presentados uno a otro como amigos mutuos de un contacto común.

Resumen

En un primer aspecto de la invención se facilita un método de identificar la presencia de elementos de información concordantes en una red, incluyendo la red un primer nodo y un segundo nodo, incluyendo el método: usar un esquema hashing para generar un conjunto de primeros valores hash a partir de un conjunto respectivo de primeros elementos de información almacenados en el primer nodo; transmitir el conjunto de primeros valores hash por la red al segundo nodo; comparar el conjunto de primeros valores hash en el segundo nodo con un conjunto de segundos valores hash generados, usando el esquema hashing, de un conjunto respectivo de segundos elementos de información almacenados en la red, para determinar por ello al menos un valor hash concordante entre el conjunto de primeros valores hash y el conjunto de segundos valores hash; usar el al menos valor hash único concordante determinado para identificar la presencia de al menos un elemento de información concordante entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información, donde el esquema hashing se elige de modo que un valor hash único en el esquema hashing indique un número suficiente de elementos de información para evitar que el valor hash único sea usado como un identificador de un único elemento de información, de tal manera que la transmisión del conjunto de primeros valores hash al segundo nodo no describa el conjunto de primeros elementos de información al segundo nodo.

En un segundo aspecto de la invención se facilita una red incluyendo: un primer nodo incluyendo medios para usar un esquema hashing para generar un conjunto de primeros valores hash a partir de un conjunto respectivo de primeros elementos de información almacenados en el primer nodo; y un segundo nodo incluyendo: medios para recibir el conjunto de primeros valores hash por la red; medios para comparar el conjunto de primeros valores hash con un conjunto de segundos valores hash generados, usando el esquema hashing, a partir de un conjunto respectivo de segundos elementos de información almacenados en la red, para determinar por ello al menos un valor hash concordante entre el conjunto de primeros valores hash y el conjunto de segundos valores hash, donde el al menos único valor hash concordante determinado se usa para identificar la presencia de al menos un elemento de información concordante entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información, y donde el esquema hashing se elige de modo que un valor hash único en el esquema hashing indique un número suficiente de elementos de información para evitar que el valor hash único sea usado como un identificador de un único elemento de información, de tal manera que la transmisión del conjunto de primeros valores hash al segundo nodo no describa el conjunto de primeros elementos de información al segundo nodo.

Realizaciones de la presente invención proporcionan un método de identificar elementos de información mantenidos en común en un sistema entre iguales sin describir los elementos de información mantenidos a otro igual en la red. Esto se logra comparando valores hash generados a partir de los elementos de información asegurando al mismo tiempo que el número de posibles valores hash en el esquema hashing sea menor, y preferiblemente significativamente menor, que el número total de elementos de información en el sistema entre iguales. Los valores hash colisionan así fuertemente a través de todo el conjunto de elementos de información en el sistema. Esto asegura que no se puede establecer una correspondencia de 1:1 de valores hash y elementos de información, de tal manera que describir un valor hash a un nodo no describa el elemento de información usado para generar el valor hash al nodo. De esta forma, la invención permite identificar elementos de información mantenidos en común sin transmitir innecesariamente los elementos de información por la red (y por ello usar innecesariamente recursos de red) a un nodo, y sin describir o identificar de forma única los elementos de información al nodo.

Se conocen esquemas hashing que pueden ser usados para generar valores hash a partir de elementos de información, de tal manera que los valores hash puedan ser comparados más bien que los elementos de

información con el fin de hallar elementos de información concordantes. Hashing es una técnica para compactar información a identificadores (valores hash), de tal forma que tanto el contenido de los elementos de información como el orden de los elementos de información en la lista de elementos de información se tomen en cuenta al generar valores hash a partir de una lista de elementos de información. Típicamente, los valores hash son más pequeños que los elementos de información a partir de los que se generan. La situación en la que un solo valor hash corresponde a más de un elemento de información se denomina “colisión hash”. Donde se usa un esquema hashing en el que no hay colisiones hash, la identificación de valores hash concordantes es igual a una identificación de elementos de información concordantes identificados por los valores hash.

La utilización de valores hash más bien que los elementos de información propiamente dichos reduce la cantidad de datos transmitidos por la red y oculta el contenido de los elementos de información. Se puede generar valores hash usando un algoritmo hashing unidireccional. Las funciones hash son por definición irreversibles, lo que quiere decir que el contenido original a partir del que se calculó el hash no pueden ser creados de nuevo a partir del valor hash, protegiendo por ello el contenido original. Esto asume que el contenido es suficientemente grande para hacer inviable un ataque de fuerza bruta (que implica generar valores hash de todas las variaciones posibles del contenido). Por lo tanto, se puede ver que el uso de los valores hash proporciona un nivel de seguridad para proteger contra la descripción de elementos de información.

Dado que los valores hash son generalmente más pequeños que los elementos de información a partir de los que se generan, es posible que más de un elemento de información genere el mismo valor hash usando una función hash particular, es decir, hay una colisión hash. Típicamente, las funciones hash se eligen con esmero para minimizar el número de colisiones hash en la red. Minimizando el número de colisiones hash, los valores hash pueden ser usados para indexar grandes elementos de datos para ahorrar espacio de índice. Los elementos de datos hallados por índice hash pueden ser comparados adicionalmente para eliminar registros no deseados que se recuperaron a causa de una colisión hash. Dado que se minimiza el número de colisiones hash en el esquema hashing, se minimiza el número de comparaciones de los elementos de datos grandes que se requieren. También se puede usar valores hash para detectar duplicados en grandes conjuntos de datos donde comparar directamente elementos de datos con otros muchos elementos de datos sería prohibitivamente complicado. Los elementos de datos son sometidos primero a hash, y luego se comparan los valores hash resultantes. En todos estos sistemas, el esquema hashing se elige para minimizar, o si es posible eliminar, las apariciones de colisiones hash. Se puede afirmar que el esquema hashing tiene ’unicidad’ si no hay colisiones hash, y que tiene casi unicidad si hay un pequeño número de colisiones hash.

En algunas aplicaciones, las funciones hash se eligen de tal forma que elementos de información similares produzcan valores hash que también sean similares. Esta propiedad puede ser usada para facilitar la distribución de datos en ‘intervalos’ separados pero conservando todavía una propiedad de unicidad hash.

En algunos sistemas, tal como un sistema de comunicaciones P2P como el descrito anteriormente, los problemas de privacidad hacen indeseable revelar información mantenida en un nodo a otro nodo de tal forma que la información se pueda ver, o incluso ser identificada de forma única por el otro nodo. Por ejemplo, en un sistema P2P, aunque la presentación de certificados digitales proporciona suficiente confianza acerca de la identidad de un usuario para establecer comunicación con dicho usuario a través de la red, los certificados digitales podrían no proporcionar confianza suficiente para describir elementos de información a dicho usuario. Aunque el uso de un esquema hashing proporciona un grado de privacidad con respecto a los elementos de información, la propiedad de unicidad generalmente deseable (o casi unicidad) de las funciones hash hace posible la identificación de elementos de información que generaron los valores hash creando un mapeado precalculado entre elementos de información y sus valores hash. Esto se conoce como un “ataque de diccionario” en el que todos los valores hash posibles son precalculados y almacenados en un “diccionario” en asociación con los elementos de información correspondientes a partir de los se generan. Si se conoce un valor hash, el elemento de información que genera el valor hash puede ser determinado entonces usando el “diccionario”.

Sin embargo, en realizaciones de la presente invención, el esquema hashing se elige intencionadamente de tal manera que genere valores hash que colisionen tan frecuentemente que un valor hash no pueda ser usado fiablemente como un identificador único de un elemento de información. Esto es contrario a la práctica hashing en la mayoría de las aplicaciones anteriores en las que el esquema hashing se elige para minimizar la aparición de colisiones hash.

Las realizaciones de la presente invención son especialmente útiles en sistemas donde diferentes elementos de información son mantenidos en nodos diferentes mientras que los nodos o usuarios de los nodos implicados no confían plenamente uno en otro. Tales sistemas pueden ser redes P2P para uso en compartición de archivos o mensajería instantánea. En una realización, el método puede ser usado para hallar información potencialmente interesante, tal como contactos compartidos en común entre usuarios en un sistema de mensajería instantánea P2P. En otra realización, el método puede ser usado para localizar fuentes potenciales de elementos de información sin revelar exactamente qué información se está buscando, tal como localizar un archivo en un sistema de compartición de archivo P2P. En otra realización, el método puede ser usado para identificar usuarios en la red con intereses comunes sin revelar los intereses propiamente dichos.

Breve descripción de los dibujos

Para una mejor comprensión de la presente invención y para mostrar cómo la misma se puede poner en práctica, ahora se hará referencia, a modo de ejemplo, a los dibujos siguientes en los que:

La figura 1 es un diagrama esquemático de un sistema de comunicaciones según una realización preferida.

La figura 2 representa una interfaz de usuario según una realización preferida.

La figura 3 representa la red de una primera realización.

La figura 4 representa un diagrama de flujo de un proceso para determinar contactos comunes en la primera realización.

La figura 5 representa la red de una segunda realización.

La figura 6 representa un diagrama de flujo de un proceso para localizar archivos en la segunda realización.

La figura 7 representa la red de una tercera realización.

Y la figura 8 representa un diagrama de flujo de un proceso para identificar usuarios con intereses comunes en la tercera realización.

Descripción detallada de realizaciones preferidas

En realizaciones de la invención un sistema de comunicaciones P2P opera en una red. La figura 1 ilustra un sistema de comunicaciones P2P a base de paquetes 100. Un primer usuario del sistema de comunicaciones (usuario A 112) opera un terminal de usuario 102, que se representa conectado al resto de la red 111. El terminal de usuario 102 puede ser, por ejemplo, un teléfono móvil, un asistente digital personal ("PDA"), un ordenador personal ("PC") (incluyendo, por ejemplo, PCs Windows™, Mac OS™ y Linux™), un dispositivo de juegos de azar u otro dispositivo embebido capaz de conectar con la red 111. La red 111 puede ser, por ejemplo, Internet. El dispositivo de usuario 102 está dispuesto para recibir y enviar información de/a un usuario 112 del dispositivo. En una realización preferida, el dispositivo de usuario 102 incluye una pantalla tal como una pantalla y un dispositivo de entrada tal como un teclado, joystick, pantalla táctil, teclado y/o ratón. El dispositivo de usuario 102 está conectado a la red 112 mediante el enlace 106.

Obsérvese que, en realizaciones alternativas, el terminal de usuario 102 puede conectar con la red de comunicaciones 111 mediante redes intermedias adicionales no representadas en la figura 1. Por ejemplo, si el terminal de usuario 102 es un dispositivo móvil, entonces puede conectar con la red de comunicaciones 111 mediante una red celular móvil (no representada en la figura 1), por ejemplo una red GSM o UMTS.

El terminal de usuario 102 ejecuta un cliente de comunicación 116, proporcionado por el proveedor de software. El cliente de comunicación 116 es un programa de software ejecutado en un procesador local en el terminal de usuario 102.

Un ejemplo de una interfaz de usuario 200 del cliente de comunicación 116 ejecutado en el terminal de usuario 102 del primer usuario 112 se ilustra en la figura 2. Obsérvese que la interfaz de usuario 200 puede ser diferente dependiendo del tipo de terminal de usuario 102. Por ejemplo, la interfaz de usuario puede ser menor o presentar información de forma diferente en un dispositivo móvil, debido al pequeño tamaño de la pantalla. En el ejemplo de la figura 2, la interfaz de usuario cliente 200 visualiza el nombre de usuario 202 de "Usuario_A" 112 en el sistema de comunicaciones. La interfaz de usuario cliente 200 incluye una pestaña 204 etiquetada "contactos", y cuando se selecciona esta pestaña, los contactos del usuario A en el sistema de comunicaciones P2P son visualizados en una hoja 206 debajo de la pestaña 204. En la interfaz de usuario ejemplar de la figura 2, se muestran cuatro contactos de otros usuarios del sistema de comunicaciones enumerados en la hoja 206, teniendo los contactos los nombres de usuario "Usuario B", "Amy", "Rosie" y "Martyn" como se representa en la figura 2. Cada uno de estos contactos ha autorizado al usuario A 112 del cliente 116 que vea sus detalles de contactos.

Volviendo a la figura 1, el nodo 102 incluye un almacenamiento 108 para contener información o datos. La información está típicamente en forma de elementos de información discretos. Los elementos de información pueden ser, por ejemplo, detalles de los contactos del usuario 112 o identificadores de archivos almacenados en el nodo 102 o identificadores de intereses del usuario 112. El enlace 106 conecta el dispositivo de usuario 102 con un segundo dispositivo de usuario 104 por la red 111. El enlace 106 puede ser una conexión directa entre nodos 102 y 104, o alternativamente el enlace 106 puede ser una conexión indirecta entre nodos 102 y 104 mediante otros nodos en el resto de la red 111. El nodo 104 está asociado con el usuario B 114 e incluye un cliente de comunicación 118 y un almacenamiento de información 110 similares a los del nodo 102.

Se puede querer identificar la presencia de elementos de información concordantes entre elementos de información almacenados en los almacenamientos 108 y 110. Sin embargo, debido a cuestiones de seguridad y privacidad, que pueden ser especialmente importantes en sistemas P2P, puede ser indeseable revelar elementos de información a otro nodo, o incluso identificar de forma única los elementos de información para el otro nodo.

Usando un esquema hashing como el descrito anteriormente, el nodo 102 puede generar una lista de valores hash usando una función hashing a partir de los elementos de información en el almacenamiento 108 en el nodo 102 y transmitir los valores hash al nodo 104 mediante el enlace 106. El nodo 104 puede generar valores hash de la lista 110 a partir de elementos de información en el almacenamiento 110 en el nodo 104 usando la misma función hashing y comparar los valores hash con los valores hash recibidos del nodo 102 para hallar valores hash concordantes. Si el esquema hashing no tiene colisiones hash, la identificación de valores hash concordantes es igual a una identificación de elementos de información concordantes identificados por los valores hash.

Las funciones hash con unicidad o casi unicidad permiten la identificación de elementos de información que generaron los valores hash. Por ejemplo, si el nodo 102 transmite una lista de valores hash al nodo 104 indicando los elementos de información en el almacenamiento 108, y los valores hash son generados usando un esquema hashing que no tiene valores hash en colisión, entonces si uno de los valores hash recibidos en el nodo 104 corresponde un valor hash generado a partir de un elemento de información en el almacenamiento 110, entonces el nodo 104 puede concluir que el nodo 102 tiene un elemento de información que concuerda con el elemento de información de la lista 110 que se usó para generar el valor hash concordante. Además, si el nodo 104 es capaz de determinar un elemento de información que generaría el valor hash, entonces el nodo 104 puede concluir que dicho elemento de información está almacenado en el nodo 102. En esquemas hashing que tienen casi unicidad, aunque puede no ser posible identificar de forma única un elemento de información a partir de un valor hash concordante, el elemento de información puede ser identificado como uno de solamente unos pocos elementos de información posibles.

El autor de la invención ha observado que se puede usar valores hash concordantes para identificar la presencia de elementos de información concordantes usando un esquema hashing en el que el número de valores hash únicos sea menor que el número de elementos de información único únicos en el sistema. Esto quiere decir que el mismo valor hash es generado a partir de más de un elemento de información único en el esquema hashing. El número de elementos de información únicos indicados por un valor hash en el esquema hashing puede ser elegido dependiendo del contexto en el que se usan los valores hash. Por ejemplo, en un contexto donde los elementos de información son detalles de contactos de un usuario, cada valor hash puede indicar preferiblemente cientos de elementos de información, mientras que en otro contexto donde los elementos de información son identificadores de archivos almacenados en un nodo de compartición, cada valor hash puede indicar preferiblemente menos elementos de información, por ejemplo aproximadamente diez elementos de información. En realizaciones preferidas, el número de valores hash únicos es al menos un orden de magnitud menor que el número de elementos de información únicos en el sistema. En estas realizaciones preferidas, se genera un valor hash único a partir de al menos 10 elementos de información únicos. Como ejemplo, el número de elementos de información únicos puede ser 300 millones, y, dependiendo del contexto, un esquema hashing adecuado puede ser un CRC de 16 bits que tiene un total de 65536 valores hash únicos, de tal manera que cada valor hash indica una media de 4500 elementos de información. Por lo tanto, describiendo el valor hash a otro nodo en la red, el elemento de información no puede ser identificado de forma única (en el ejemplo, el elemento de información puede ser uno de 4500 elementos de información), de modo que se evita que el otro nodo use el valor hash como un identificador de un único elemento de información. El valor hash solamente puede ser usado por el otro nodo para identificar que el elemento de información es uno de un grupo de elementos de información que generan el mismo valor hash.

Ahora se describirán tres realizaciones especialmente útiles. La primera realización se describe con referencia a las figuras 3 y 4, la segunda realización se describirá con referencia a las figuras 5 y 6 y la tercera realización se describirá con referencia a las figuras 7 y 8.

La primera realización especialmente útil se refiere a hallar contactos comunes entre usuarios en una red. La figura 3 representa los nodos de usuario 102 y 104 conectados mediante el enlace 106. El almacenamiento 108 en el nodo 102 contiene los nombres de usuario de contactos del usuario A 112. En el ejemplo representado en la figura 3, el almacenamiento 108 contiene los nombres de usuario “Amy”, “Rosie” y “Martyn”. El almacenamiento 110 en el nodo 104 contiene los nombres de usuario de contactos del usuario B 114. En el ejemplo representado en la figura 3, el almacenamiento 110 contiene los nombres de usuario “Madis”, “Martyn”, “Sonia” y “Michael”. Al usuario A 112 del nodo 102 le gustaría determinar si algunos de sus contactos también son contactos del usuario B 114 del nodo 104, pero el usuario A no quiere describir su lista de contactos al usuario B en el nodo 104. Para hacerlo, se usa el método representado en la figura 4.

En el paso S402, el nodo 102 genera valores hash de los nombres de usuario contacto almacenados en el almacenamiento 108 según un esquema hashing como se ha descrito anteriormente. En el paso S404, los valores hash de contactos del usuario A son transmitidos por la red en enlace 106 al nodo 104. Como se ha descrito anteriormente, un valor hash en el esquema hashing corresponde a más de un nombre de usuario contacto (es

decir, la relación de valores hash a los nombres de usuario contacto en el sistema es 1:muchos). Por lo tanto, el nodo 104 no puede determinar fiablemente los contactos de usuario A 112 usando los valores hash proporcionados a partir del nodo 102.

En el paso S406, el nodo 104 genera valores hash de los nombres de usuario contacto almacenados en el almacenamiento 110 según el mismo esquema hashing que el usado para generar los valores hash de los nombres de usuario contacto del usuario A 112.

En el paso S408, los valores hash recibidos del nodo 102 son comparados con los valores hash generados en el nodo 104 para hallar valores hash concordantes. En el paso S410, los valores hash concordantes son transmitidos por la red al nodo 102 mediante el enlace 106. En el paso S412, el nodo 102 determina los nombres de usuario contacto que se usaron para generar los valores hash concordantes, con el fin de determinar contactos comunes entre el usuario A 112 y el usuario B 114.

En el ejemplo representado en la figura 3, el usuario en el sistema con nombre de usuario “Martyn” es un contacto tanto del usuario A 112 como del usuario B 114. El valor hash generado para el nombre de usuario “Martyn” se hallará como un valor hash concordante en el paso S408, y luego, en el paso S412, se determinará que Martyn es un contacto común del usuario A 112 y del usuario B 114. El esquema hashing se elige para tener significativamente menos valores hash únicos que usuarios en el sistema, de tal manera que un valor hash pueda ser generado a partir de muchos nombres de usuario. Sin embargo, hay significativamente más valores hash únicos en el esquema hashing que nombres de usuario contacto en los almacenamientos 108 y 110. Como un ejemplo, en todo el sistema de comunicaciones puede haber 300 millones de usuarios con nombres de usuario diferentes. Un esquema hashing adecuado es un CRC de 16 bits que tiene un espacio clave total de 65.536, es decir, hay 65.536 valores hash únicos que pueden ser usados para indicar los nombres de usuario. En este ejemplo, cada valor hash único indicaría una media de 4500 nombres de usuario en todo el sistema P2P, pero el usuario A tiene tres contactos y el usuario B tiene cuatro contactos, de modo que es improbable que más de un contacto del usuario A o el usuario B sea indicado por el mismo valor hash. En el ejemplo representado en la figura 3, la probabilidad de una colisión hash aleatoria viene dada por: (número de contactos del usuario A) x (número de contactos del usuario B)/(número de valores hash únicos en el esquema hashing) = 12/65536 ≈ 0,00018. Esto es tan bajo que, en este caso, cualesquiera valores hash concordantes pueden ser indicativos de contactos concordantes.

La razón por la que se puede suponer que los valores hash concordantes son indicativos de contactos concordantes a pesar de que cada valor hash representa 4500 nombres de usuario contacto como media en este caso es porque el hecho de que los usuarios 112 y 114 tengan contactos comunes es más probable que el hecho de que dos contactos diferentes del usuario 112 o del usuario 114 generen aleatoriamente el mismo valor hash. Esto es verdadero porque hay una asociación entre los contactos del usuario A 112 y los contactos del usuario B 114. En particular, el usuario A 112 conoce al usuario B (el usuario B 114 podría ser un contacto del usuario A 112, o el usuario A 112 simplemente podría tener idea del usuario B 114), lo que quiere decir que es mucho más probable que un contacto del usuario B 114 sea un contacto de usuario A que un usuario completamente aleatorio en la red 111. De esta forma, eligiendo un esquema hashing adecuado, puede asegurarse que, al comparar contactos de usuarios que se conocen entre sí, los valores hash concordantes indican predominantemente contactos concordantes más bien que colisión hash de contactos diferentes. Esto permite suponer que los valores hash concordantes son indicativos de contactos concordantes. Un escenario especialmente útil en el que se puede usar el método es permitir que se ejecute una búsqueda de Amigos de Amigos en nodos P2P sin revelar los nombres de contacto entre nodos P2P que participan en la búsqueda, permitiendo al mismo tiempo la identificación de contactos compartidos.

Este método no se limita al uso al determinar contactos comunes entre usuarios que se conocen. De hecho, el método funcionará con cualquier tipo de elemento de información almacenado en listas en nodos diferentes en la red, donde hay una asociación entre las listas que incrementa la probabilidad de que los elementos de información en una lista estén presentes en la otra lista. Donde las listas están asociadas de tal forma que la probabilidad de que uno de los elementos de información en una lista concuerde con uno de los elementos de información en la otra lista es más grande que la probabilidad de que dos elementos de información diferentes sean identificados por el mismo valor hash en el esquema hashing, entonces se puede considerar que los valores hash concordantes son indicativos de elementos de información concordantes. De esta forma, se pueden hallar elementos de información comunes entre los almacenamientos 108 y 110 sin describir los elementos de información almacenados en el nodo 108 o en el nodo 110 a ningún nodo en la red. Un nodo que recibe los valores hash de los elementos de información en el almacenamiento 108 del nodo 102 no puede determinar los elementos de información en el almacenamiento 108 porque cada valor hash identifica muchos elementos de información diferentes en el esquema hashing.

En una realización alternativa, los valores hash de los elementos de información en el almacenamiento 108 pueden ser transmitidos a un tercer nodo (no representado en la figura 3), y los valores hash de los elementos de información en el almacenamiento 110 también pueden ser transmitidos al tercer nodo, donde los valores hash son comparados en el tercer nodo para hallar valores hash concordantes entre los almacenamientos 108 y 110. En esta realización, los valores hash no son enviados al nodo 104, lo que puede ser beneficioso si el usuario A 112 tiene más confianza en el tercer nodo que en el nodo 104 y/o en el usuario B 114.

Una segunda realización especialmente útil se describe con referencia a las figuras 5 y 6 y se refiere a compartición de archivos en una red. La figura 5 representa un primer nodo de compartición 502 que almacena Archivo 1, Archivo 2 y Archivo 3 en un almacenamiento 508 y un segundo nodo de compartición 504 que almacena Archivo 11, Archivo 12 y Archivo 13 en un almacenamiento 510. Los nodos de compartición 502 y 504 pueden comunicar con un índice central 512 por la red mediante respectivos enlaces 506 y 507. De forma similar al enlace 106 descrito anteriormente, los enlaces 506 y 507 pueden ser enlaces directos o indirectos por la red.

En el paso S602, valores hash de los archivos almacenados en el nodo de compartición 502 son generados en el nodo 502 usando un esquema hashing adecuado como se ha descrito anteriormente para evitar que los valores hash sean usados como identificadores de archivos únicos, y valores hash de los archivos almacenados en el nodo de compartición 504 son generados en el nodo 504 usando el mismo esquema hashing. En el paso S604, los valores hash generados en el nodo 502 son transmitidos por el enlace 506 al índice central 512, y los valores hash generados en el nodo 504 son transmitidos por el enlace 507 al índice central 512. Los valores hash son almacenados en el almacenamiento 514 en el índice central 512. Puede haber muchos nodos de compartición en la red, pero solamente se representan dos (nodos 502 y 504) en la figura 5 para claridad. El almacenamiento 514 guarda los valores hash de tal forma que pueda asociar un valor hash en el almacenamiento 514 con uno de los nodos de compartición (por ejemplo, 502 o 504) del que se recibió el valor hash. Esto se puede hacer por cualquier método adecuado tal como almacenando los valores hash de diferentes nodos de compartición en listas diferentes, o enlazando cada valor hash en el almacenamiento 514 con el nodo de compartición relevante.

Un nodo solicitante 102 tal como el nodo 102 representado en la figura 1 asociado con el usuario A 112 puede comunicar con el índice central 112 por la red mediante un enlace 505. De forma similar al enlace 106 descrito anteriormente, el enlace 505 puede ser un enlace directo o indirecto por la red. El nodo 102 tiene un almacenamiento 108 que puede almacenar uno o más valores hash de archivos que el nodo 102 desearía hallar en la red. Por ejemplo, como se representa en la figura 5, el usuario 112 desea hallar Archivo 2 y el almacenamiento 108 guarda “Hash 2” que corresponde a Archivo 2 almacenado en el nodo de compartición 502. En el paso S606, el nodo solicitante transmite una petición al índice central usando el enlace 505 para localizar el archivo que corresponde al valor hash almacenado en el almacenamiento 108. La petición incluye el valor hash “Hash 2” del almacenamiento 108.

En el paso S608, el valor hash recibido del nodo solicitante 102 es comparado con los valores hash almacenados en el almacenamiento 514 en el índice central 512 para determinar cualesquiera valores hash concordantes. Se determina(n) la(s) posición(es) de los nodos de compartición que almacenan archivos identificados por los valores hash concordantes. Dado que un valor hash corresponde a muchos archivos en el esquema hashing, es posible que un valor hash concordante no indique el archivo correcto que el nodo 102 esté buscando. Sin embargo, la(s) posición(es) de los nodos de compartición determinada(s) en el paso S608 es (son) devuelta(s) al nodo solicitante 102 en el paso S610. Por ejemplo, en el sistema representado en la figura 5, la posición del primer nodo de compartición 502 será devuelta al nodo solicitante porque Hash 2 corresponde a Archivo 2 almacenado en el nodo 502. Sin embargo, también es posible que, por ejemplo, Hash 11 generado a partir de Archivo 11 sea el mismo que Hash 2 generado a partir de Archivo 2, de modo que la posición del segundo nodo de compartición 504 también pueda ser devuelta al nodo solicitante 102.

En el paso S612, el nodo solicitante 102 contacta el (los) nodo(s) de compartición identificado(s) por la red para determinar si el archivo correcto (Archivo 2) está almacenado en el (los) nodo(s) de compartición. Esto se puede hacer transmitiendo un identificador del archivo al nodo de compartición que es más preciso que los valores hash usados previamente. El identificador más preciso puede ser un valor hash diferente calculado usando un esquema hashing diferente. En el esquema hashing diferente, un valor hash puede identificar un archivo único. Alternativamente, el identificador más preciso puede ser el nombre de archivo del archivo. Se puede usar otros identificadores como será evidente a los expertos. En el paso S614, el (los) nodo(s) de compartición usa(n) el identificador más preciso para determinar si el nodo de compartición contiene el archivo que el nodo solicitante está buscando. En el ejemplo representado en la figura 5, el primer nodo de compartición 502 determinará que Archivo 2 es el archivo que el nodo solicitante 102 está buscando, y entonces puede transmitir el archivo al nodo solicitante 102. Sin embargo, el segundo nodo de compartición 504 determinará que Archivo 11 no es el archivo que el nodo solicitante está buscando aunque Hash 11 concordase con Hash 2. Aunque el archivo en el nodo de compartición es el archivo que el nodo solicitante está buscando, el nodo de compartición puede decidir no transmitir el archivo al nodo solicitante.

En algunas realizaciones, en el paso S612, el nodo solicitante tendrá que enviar una autenticación al nodo de compartición. El nodo de compartición comprobará entonces la autenticación y solamente si el nodo solicitante es autenticado, el nodo de compartición transmitirá el archivo al nodo solicitante en el paso S614. En un sistema P2P como el descrito anteriormente, la autenticación podría ser el certificado digital del nodo solicitante 102.

Cuando el método se usa en un sistema para compartición de archivos tal como el descrito anteriormente en relación a las figuras 5 y 6, el índice central puede almacenar valores hash de archivos almacenados en nodos de compartición, y dado que los valores hash no identifican archivos únicos, los archivos reales almacenados en los

nodos de compartición no pueden ser determinados a partir de una inspección del índice central. Además, el nodo solicitante 102 puede enviar una petición al índice central 512, y la información real que está buscando no se describe. Sin embargo, todavía es posible usar el índice central para buscar archivos almacenados en nodos de compartición en la red usando el método descrito anteriormente.

El método de la segunda realización especialmente útil se describe anteriormente con referencia a las figuras 5 y 6 en relación a elementos de información que son identificadores de archivos almacenados en los nodos. Sin embargo, el método puede ser usado para otros elementos de información donde se desea identificar nodos de compartición que almacenan elementos de información particulares.

Una tercera realización especialmente útil se describe con referencia a las figuras 7 y 8 y se refiere a hallar usuarios con intereses similares en una red. La figura 7 representa un primer nodo de usuario 102 que está asociado con el usuario A 112. El almacenamiento 108 en el nodo 102 contiene identificadores de intereses de usuario A 112. Igualmente, el nodo 702 está asociado con el usuario C 710 e incluye un almacenamiento 708 que contiene identificadores de intereses de usuario C 710. Un nodo central 712 en la red puede comunicar con el nodo 102 mediante el enlace 705 y puede comunicar con el nodo 702 mediante el enlace 706. De forma similar al enlace 106 descrito anteriormente, los enlaces 705 y 706 pueden ser enlaces directos o indirectos por la red.

Como ejemplos, se puede usar URLs y palabras clave de búsqueda como identificadores de los intereses de un usuario. Las imágenes que un usuario ve en una base de datos de imágenes también pueden usarse como identificadores de los intereses del usuario.

En el paso S802, el nodo 102 genera un conjunto de valores hash de los identificadores de los intereses en el almacenamiento 108 usando un esquema hashing adecuado como el descrito anteriormente. Igualmente, el nodo 702 genera un conjunto de valores hash a partir de los identificadores de los intereses en el almacenamiento 708 usando el mismo esquema hashing. En el paso S804, el conjunto de valores hash generados en el nodo 102 es transmitido al nodo central 712 usando el enlace 705, y el conjunto de valores hash generado en el nodo 702 es transmitido al nodo central 712 usando el enlace 706. Los conjuntos de valores hash son almacenados en el nodo central 712. En el ejemplo representado en la figura 7, el conjunto de valores hash recibido del nodo 102 se almacena en un almacenamiento 714 y el conjunto de valores hash recibido del nodo 702 se almacena en un almacenamiento 716. En otras realizaciones, los conjuntos de valores hash recibidos de diferentes nodos son almacenados en el mismo almacenamiento en el nodo central 712, teniendo el nodo central 712 algún mecanismo para identificar el nodo del que se recibieron los conjuntos de valores hash.

En el paso S806, el conjunto de valores hash recibido del nodo 102 es comparado con el conjunto de valores hash recibido del nodo 702 en el nodo central 712 para hallar valores hash concordantes entre los conjuntos. En el paso S808, se cuenta el número de valores hash concordantes entre los conjuntos. El número de valores hash concordantes da una indicación sobre si los usuarios 112 y 710 tienen intereses similares. Dado que en el esquema hashing cada valor hash identifica muchos intereses diferentes, algunos valores hash concordantes pueden no estar relacionados con intereses concordantes. De hecho, si los intereses de los usuarios 112 y 710 se seleccionasen de forma completamente aleatoria a partir de todos los intereses que pueden ser identificados en el sistema, habría un número esperado de valores hash concordantes entre los dos conjuntos que dependería del tamaño de los dos conjuntos de valores hash y del número de valores hash únicos en el esquema hashing usado para generar los valores hash.

En el paso S810, se determina si el número de valores hash concordantes es mayor que el número esperado de valores hash concordantes entre los dos conjuntos de intereses si los intereses se seleccionaron aleatoriamente. Si hay más valores hash concordantes de los esperados, entonces eso indica que los usuarios 102 y 710 tienen intereses similares. El nivel de semejanza entre los intereses de los dos usuarios puede ser cuantificado por la cantidad que el número de valores hash concordantes excede del número esperado de valores hash concordantes. De hecho, es posible atribuir un valor de intensidad a la semejanza de los intereses entre los usuarios. Esta información puede ser usada para muchos fines. Por ejemplo, si el nodo central 712 se usa para alojar un grupo de red social, la intensidad de la semejanza entre dos usuarios que son parte del grupo de red social podría ser usada para uno de los usuarios para identificar el otro usuario como alguien que tiene intereses similares a los suyos propios.

Dado que el esquema hashing usa valores hash que no identifican intereses únicos, el método permite usar información referente a intereses de personas para localizar personas con intereses comunes sin revelar exactamente cuáles son los intereses. Los valores hash almacenados en el nodo central 712 no pueden ser usados para identificar de forma única los intereses de un usuario. Los sitios web que una persona visita pueden ser muy indicativos de los intereses que tiene una persona. Esto quiere decir que las URLs usados por un usuario pueden usarse como los identificadores de los intereses del usuario. Igualmente, se puede usar palabras clave de búsqueda como los identificadores de los intereses. El número de valores hash en común entre dos listas puede ser considerado indicativo de intereses comunes.

El método también puede aplicarse a archivos, por ejemplo generando valores hash de contenido de archivo de

imagen de un sitio de compartición de imágenes y comparando listas de valores hash entre dos usuarios puede determinarse si les gusta ver imágenes similares. Por ejemplo, si un usuario medio ha visto 2000 imágenes de una base de datos de 2 millones de imágenes, entonces usar un valor hash de 16 bits (con 65536 valores hash únicos) identificaría suficientemente bien imágenes en dos listas de imágenes vistas por usuarios mientras que cada valor hash concordaría con alguna de 30 imágenes diferentes. Cada valor hash concordante incrementa la probabilidad de que otros valores hash concordantes hagan referencia de hecho a imágenes concordantes de las 30 posibilidades del sistema. Una vez que se ha hallado un número suficiente de valores hash concordantes, entonces hay una alta probabilidad de que los dos usuarios hayan visto las mismas imágenes. En otros términos, si hay gran número de valores hash concordantes, entonces los dos usuarios son identificados como de intereses similares, y una vez que son identificados como de intereses similares, entonces hay una asociación entre los conjuntos de intereses de un usuario y el conjunto de intereses del otro usuario. Esto puede considerarse similar a la primera realización especialmente útil descrita anteriormente en la que están asociadas las listas de contactos de usuarios que se conocen mutuamente. De la misma forma, donde dos usuarios tienen intereses similares, entonces los intereses están asociados de tal manera que sea más probable que otro valor hash concordante sea debido a un interés concordante más bien que debido a una colisión hash aleatoria de intereses diferentes. Por lo tanto, una vez que los usuarios 102 y 710 han sido identificados como de intereses similares, puede suponerse que más valores hash concordantes identifican intereses concordantes.

En el ejemplo representado en la figura 7, ambos usuarios 102 y 710 tienen intereses 2 y 3 y los valores hash concordantes serán determinados en el nodo central 712 en el paso S806. Tener dos valores hash concordantes será más de lo esperado para el caso donde cada usuario tiene solamente 3 intereses como se representa en la figura 7 y así se determinará que los usuarios 102 y 710 tienen intereses similares.

El método de la tercera realización especialmente útil se describe anteriormente con referencia a las figuras 7 y 8 en relación a elementos de información que son identificadores de intereses de usuarios. Sin embargo, el método puede ser usado para otros elementos de información donde se desee identificar nodos que tengan elementos de información similares.

Las realizaciones descritas anteriormente usan un esquema hashing con menos valores hash únicos que elementos de información únicos en el sistema de tal manera que un valor hash no pueda ser usado para identificar un elemento de información único. Esto asegura la privacidad y la seguridad en el sistema dado que un nodo puede transmitir a otro nodo los valores hash generados a partir de elementos de información almacenados en el nodo y el otro nodo no puede determinar los elementos de información usando los valores hash. Por lo tanto, la privacidad y la seguridad se mejoran mientras que el método mantiene la capacidad de identificar la presencia de elementos de información concordantes en nodos diferentes en el sistema.

Aunque esta invención se ha mostrado y descrito en particular con referencia a realizaciones preferidas, los expertos en la técnica entenderán que se puede hacer varios cambios en la forma y el detalle sin apartarse del alcance de la invención definido por las reivindicaciones anexas.

REIVINDICACIONES

1. Un método de identificar la presencia de elementos de información concordantes en una red (100), incluyendo la red un primer nodo (102) y un segundo nodo (104), incluyendo el método:

usar un esquema hashing para generar (S402) un conjunto de primeros valores hash a partir de un conjunto respectivo de primeros elementos de información almacenados en el primer nodo;

transmitir (S404) el conjunto de primeros valores hash por la red al segundo nodo;

comparar el conjunto de primeros valores hash en el segundo nodo con un conjunto de segundos valores hash generados, usando el esquema hashing, a partir de un conjunto respectivo de segundos elementos de información almacenados en la red, para determinar por ello (S408) al menos un valor hash concordante entre el conjunto de primeros valores hash y el conjunto de segundos valores hash;

usar el al menos valor hash único concordante determinado para identificar (S412) la presencia de al menos un elemento de información concordante entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información,

donde el esquema hashing se elige de modo que un valor hash único en el esquema hashing indique un número suficiente de elementos de información para evitar que el valor hash único sea usado como un identificador de un único elemento de información, de tal manera que la transmisión del conjunto de primeros valores hash al segundo nodo no describa el conjunto de primeros elementos de información al segundo nodo.

2. El método de la reivindicación 1, donde el conjunto de segundos valores hash incluye una pluralidad de segundos valores hash y el conjunto de segundos elementos de información incluye una pluralidad respectiva de segundos elementos de información.

3. El método de la reivindicación 1, donde el conjunto de primeros valores hash incluye una pluralidad de primeros valores hash y el conjunto de primeros elementos de información incluye una pluralidad respectiva de primeros elementos de información.

4. El método de la reivindicación 1, donde el conjunto de segundos elementos de información se almacena en el segundo nodo.

5. El método de la reivindicación 1, donde hay una asociación entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información de tal manera que la probabilidad de que uno de los primeros elementos de información concuerde con uno de los segundos elementos de información es más grande que la probabilidad de que dos elementos de información diferentes sean identificados por el mismo valor hash en el esquema hashing, de tal manera que se suponga que los valores hash concordantes son indicativos de elementos de información concordantes, donde los primeros elementos de información identifican contactos de un primer usuario del primer nodo, y los segundos elementos de información identifican contactos de un segundo usuario en la red.

6. El método de la reivindicación 1, incluyendo además:

contar (S808) el número de valores hash concordantes entre el conjunto de primeros valores hash y el conjunto de segundos valores hash; y

determinar (S810) si el número contado de valores hash concordantes es más grande que el número esperado de valores hash concordantes en base al número de primeros valores hash, el número de segundos valores hash y el esquema hashing; y

si el número contado es más grande que el número esperado, identificar que el conjunto de primeros elementos de información es similar al conjunto de segundos elementos de información.

7. El método de la reivindicación 6, incluyendo además usar la diferencia entre el número contado y el número esperado para atribuir un valor de intensidad a la semejanza entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información.

8. El método de la reivindicación 6 o la reivindicación 7, donde los primeros elementos de información identifican intereses de un primer usuario del primer nodo, y los segundos elementos de información identifican intereses de un segundo usuario en la red, y donde la identificación de que el conjunto de primeros elementos de información es similar al conjunto de segundos elementos de información identifica que los usuarios primero y segundo tienen intereses similares.

9. El método de la reivindicación 1, donde el conjunto de segundos elementos de información se almacena en un tercer nodo en la red, incluyendo además el método determinar la posición del tercer nodo, donde el paso de usar el al menos único valor hash concordante determinado para identificar la presencia de al menos un elemento de información concordante incluye:

transmitir la posición del tercer nodo por la red al primer nodo;

determinar el elemento de los primeros elementos de información que corresponde a uno del al menos valor hash único concordante;

transmitir un identificador del elemento de los primeros elementos de información del primer nodo al tercer nodo;

determinar selectivamente un elemento de los segundos elementos de información que corresponde a uno del al menos único valor hash concordante; y

determinar selectivamente en el tercer nodo si el identificador del elemento de los primeros elementos de información también identifica el elemento de los segundos elementos de información.

10. Una red (100) incluyendo:

un primer nodo (102) incluyendo medios para usar un esquema hashing para generar un conjunto de primeros valores hash a partir de un conjunto respectivo de primeros elementos de información almacenados en el primer nodo; y

un segundo nodo (104) incluyendo:

medios para recibir el conjunto de primeros valores hash por la red;

medios para comparar el conjunto de primeros valores hash con un conjunto de segundos valores hash generados, usando el esquema hashing, a partir de un conjunto respectivo de segundos elementos de información almacenados en la red, para determinar por ello el al menos único valor hash concordante entre el conjunto de primeros valores hash y el conjunto de segundos valores hash,

donde el al menos único valor hash concordante determinado se usa para identificar la presencia de al menos un elemento de información concordante entre el conjunto de primeros elementos de información y el conjunto de segundos elementos de información, y

donde el esquema hashing se elige de modo que un valor hash único en el esquema hashing indique un número suficiente de elementos de información para evitar que el valor hash único sea usado como un identificador de un elemento de información único, de tal manera que la transmisión del conjunto de primeros valores hash al segundo nodo no describa el conjunto de primeros elementos de información al segundo nodo.

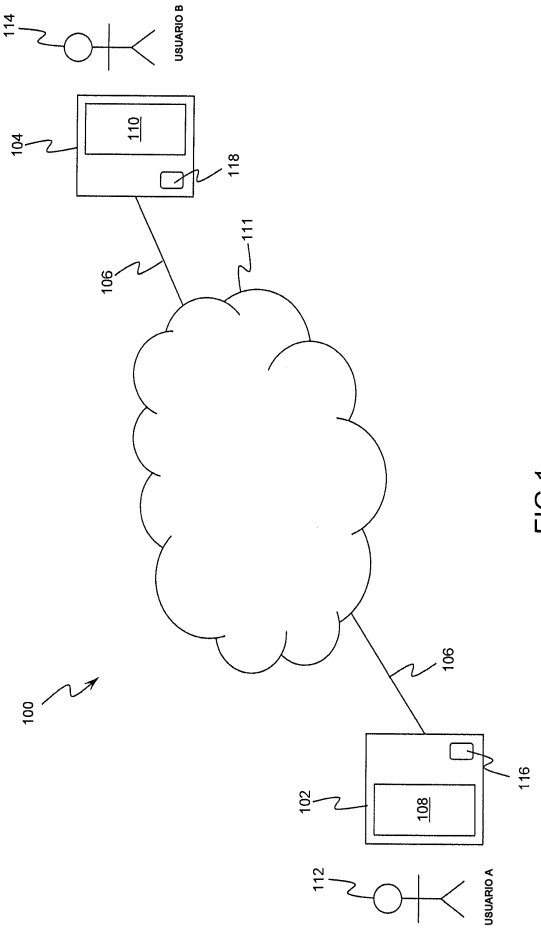
114
USUARIO B
104
110
118
106
111
100
106
116
102
108
112
USUARIO A

FIG 1

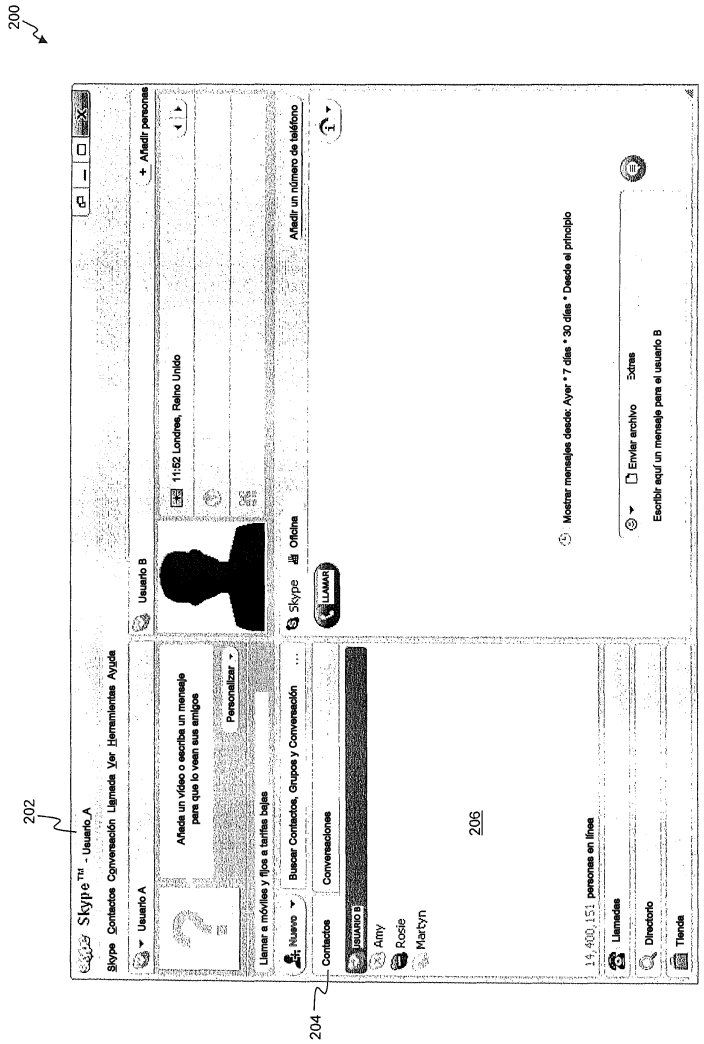
200
202
Skype™ - Usuario A
Skype Contactos Conversación Llamada Ver Herramientas Ayuda
Usuario A
Usuario B
+ Añadir personas
Añada un vídeo o escriba un mensaje para que lo vean sus amigos
Personalizar
11:52 Londres, Reino Unido
Llamar a móviles y fijos a tarifas bajas
Nuevo
Buscar Contactos, Grupos y Conversación
Skype
Offline
LLAMAR
Añadir un número de teléfono
Contactos
Conversaciones
USUARIO B
Amy
Rosie
Martyn
206
Mostrar mensajes desde: Ayer * 7 días * 30 días * Desde el principio
14.400.151 personas en línea
Llamadas
Directorio
Tienda
Enviar archivo
Extras
Escribir aquí un mensaje para el usuario B
204

FIG 2

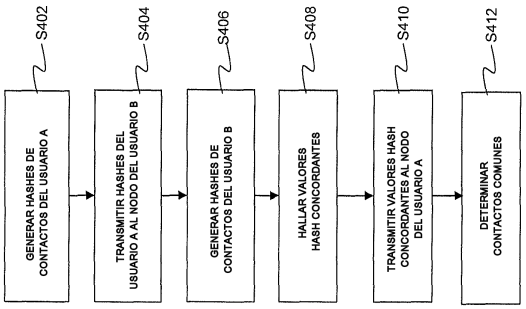
S402
GENERAR HASHES DE CONTACTOS DEL USUARIO A
S404
TRANSMITIR HASHES DEL USUARIO A AL NODO DEL USUARIO B
S406
GENERAR HASHES DE CONTACTOS DEL USUARIO B
S408
HALLAR VALORES HASH CONCORDANTES
S410
TRANSMITIR VALORES HASH CONCORDANTES AL NODO DEL USUARIO A
S412
DETERMINAR CONTACTOS COMUNES

FIG 4

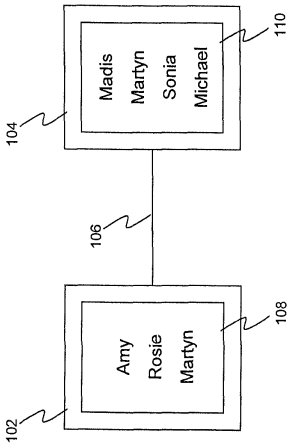
Madis
Martyn
Sonia
Michael
110
104
106
Amy
Rosie
Martyn
108
102

FIG 3

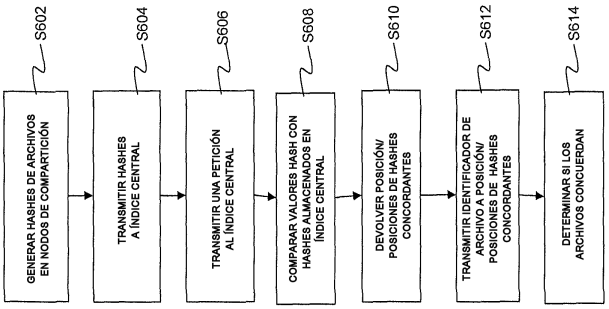
S602
GENERAR HASHES DE ARCHIVOS EN NODOS DE COMPARTICIÓN
S604
TRANSMITIR HASHES A ÍNDICE CENTRAL
S606
TRANSMITIR UNA PETICIÓN AL ÍNDICE CENTRAL
S608
COMPARAR VALORES HASH CON HASHES ALMACENADOS EN ÍNDICE CENTRAL
S610
DEVOLVER POSICIÓN/ POSICIONES DE HASHES CONCORDANTES
S612
TRANSMITIR IDENTIFICADOR DE ARCHIVO A POSICIÓN/ POSICIONES DE HASHES CONCORDANTES
S614
DETERMINAR SI LOS ARCHIVOS COINCIDEN

FIG 6

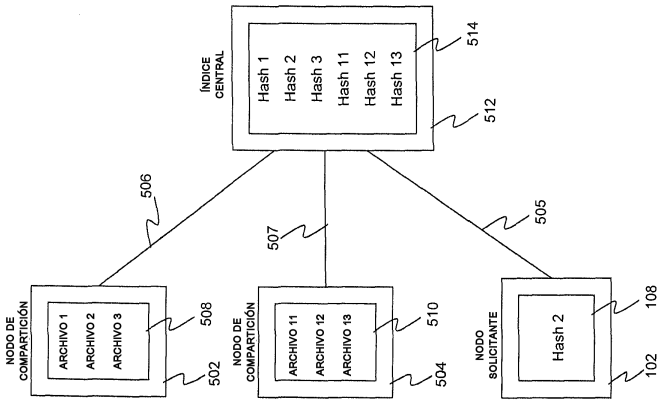
ÍNDICE CENTRAL
Hash 1
Hash 2
Hash 3
Hash 11
Hash 12
Hash 13
514
512
506
507
505
NODO DE COMPARTICIÓN
ARCHIVO 1
ARCHIVO 2
ARCHIVO 3
508
502
NODO DE COMPARTICIÓN
ARCHIVO 11
ARCHIVO 12
ARCHIVO 13
510
504
NODO SOLICITANTE
Hash 2
108
102

FIG 5

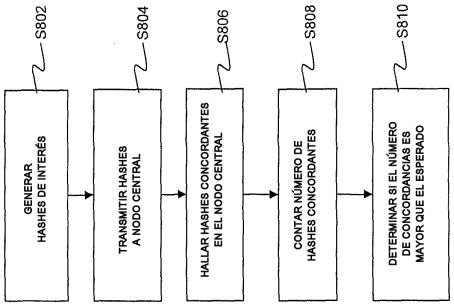
S802
GENERAR HASHES DE INTERÉS
S804
TRANSMITIR HASHES A NODO CENTRAL
S806
HALLAR HASHES CONCORDANTES EN EL NODO CENTRAL
S808
CONTAR NÚMERO DE HASHES CONCORDANTES
S810
DETERMINAR SI EL NÚMERO DE CONCORDANCIAS ES MAYOR QUE EL ESPERADO

FIG 8

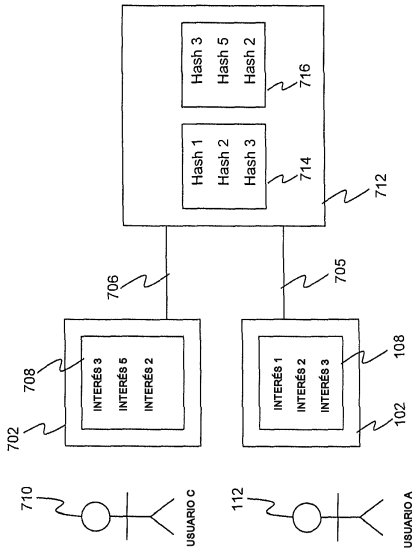
Hash 3
Hash 5
Hash 2
716
Hash 1
Hash 2
Hash 3
714
712
706
705
INTERÉS 3
INTERÉS 5
INTERÉS 2
708
702
INTERÉS 1
INTERÉS 2
INTERÉS 3
108
102
710
USUARIO C
112
USUARIO A

FIG 7