

Redes de contenido: un panorama de sus características y principales aplicaciones

Pablo Rodríguez Bocca

Departamento de Investigación Operativa, Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
prbocca@fing.edu.uy

Héctor Cancela Bosi

Departamento de Investigación Operativa, Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
cancela@fing.edu.uy

Resumen

Las redes de contenido han jugado un papel protagónico en el desarrollo de la sociedad de la información, brindando infraestructura para la oferta de contenido especializado y/o masivo. La propia Internet puede verse bajo este paradigma como una red de contenido, donde el contenido son las direcciones IP; en el sistema de nombres de dominio DNS el contenido son los nombres completos de los hosts, siendo un sistema vital para la mayoría de servicios que se ofrecen en Internet.

Las motivaciones para construir redes de contenido son muy variadas, dado que permiten una adaptabilidad muy grande a las necesidades particulares de las aplicaciones con requerimientos específicos en cuanto a costo, libertad de diseño, confiabilidad, disponibilidad, redundancia, performance, tiempo de acceso, anonimato, seguridad, etc.

En los últimos años se han desarrollado en distintos contextos gran cantidad de redes de contenido, incluyendo redes peer-to-peer, collaborative networks, cooperative Web caching, content distribution network, subscribe-publish networks, content-based sensor networks, backup networks, distributed computing, instant messaging, y multiplayer games.

Este trabajo presenta las principales características de las redes de contenido, tanto desde el punto de su arquitectura como de su comportamiento visto por el usuario, sintetizadas a partir de un relevamiento amplio de la literatura y de las aplicaciones existentes. Asimismo se propone una taxonomía basada en algunas de estas características, ejemplificando cada categoría de esta taxonomía con diversas aplicaciones existentes.

Palabras clave

Redes de contenido, sistemas distribuidos, P2P, trabajo colaborativo.

INTRODUCCIÓN

Una Red de Contenido es una red donde el direccionamiento y ruteo del contenido se basa en la descripción del contenido en lugar de su ubicación [39][41]. Las redes de contenido son redes virtuales que se montan sobre la infraestructura IP de la Internet o de una red corporativa. Para acceder al contenido no es necesario mantener continuamente un vínculo fijo entre el contenido y el host donde

está alojado. Mas aún, para muchas redes de contenido, en un momento dado, parte del contenido puede ser alojado, movido o replicado en algún otro nodo de la red, si se considera conveniente.

Estas características, sumadas a una libertad de diseño producto del continuo desarrollo de protocolos y arquitecturas, representan un atractivo muy importante para nuevas aplicaciones con requerimientos específicos en cuanto a confiabilidad, disponibilidad, redundancia, performance, tiempo de acceso, anonimato, seguridad, etc.

En los últimos años, se han desarrollado en distintos contextos muchas y muy diversas redes de contenido, incluyendo redes peer-to-peer, collaborative networks, cooperative Web caching, content distribution network, subscribe-publish networks, content-based sensor networks, backup networks, distributed computing, instant messaging, y multiplayer games. Las redes de contenido, que hoy suelen implementarse bajo la forma de redes virtuales montadas sobre la infraestructura existente de Internet, son sumamente atractivas dada su flexibilidad a adecuarse a los requerimientos de cada aplicación específica, como se ha mencionado anteriormente. Sin embargo, en general no contemplan al armar sus topologías la red física de Internet (es decir dos nodos conectados en la red virtual pueden estar muy alejados en la red física), provocando mayor retardo a los clientes y un fuerte impacto sobre la lógica de negocios de los proveedores de Internet [50][53][56].

El objetivo de este trabajo es estudiar las redes de contenido, especialmente a nivel de las diferentes características que corresponden a su arquitectura y a su comportamiento, de manera de dar un marco que permita al diseñador y al usuario comprender las principales opciones existentes y sus interrelaciones. Para esto se hizo un relevamiento amplio de la literatura, así como de la información disponible de las distintas aplicaciones de redes de contenido existentes. Si bien hay mucho material sobre diversas propuestas de arquitecturas específicas, no es fácil encontrar estudios de síntesis, ya que es un área de desarrollo muy rápido y reciente. Del material relevado, dos trabajos han servido como puntos de apoyo: por un lado el artículo de H. T. Kung y C. H. Wu [39], que presenta una taxonomía de redes de contenido basadas en la forma de agrupar lógicamente contenidos y de distribuir físicamente los mismos; y por otro el reporte técnico de D. S. Milojicic et.al. [42],

que estudia las redes P2P y discute una cantidad importante de características relevantes desde un punto de vista empresarial y su lógica de negocio.

La próxima sección presenta las principales características de las redes de contenido, por un lado aquellas referentes a la arquitectura, y por otro las que tienen en cuenta el comportamiento de la red. La sección siguiente presenta una taxonomía de las redes de contenido y una discusión de cada categoría de esta taxonomía, presentando ejemplos de aplicaciones existentes, extraídas de la literatura.

CARACTERÍSTICAS DE LAS REDES DE CONTENIDO

Las redes de contenido son redes virtuales, cuyo direccionamiento y ruteo se basa en la descripción del contenido en lugar de su ubicación; tienen por lo tanto la flexibilidad de crear una topología virtual propia, no importando la red física subyacente: por ejemplo, dos nodos conectados en la red virtual pueden estar muy alejados en la red física. Los nodos que pertenecen a dicha red virtual son llamados nodos de red o nodos de contenido. En base a su funcionalidad, pueden clasificarse básicamente en tres tipos: fuentes de contenido, que tienen algún tipo de contenido puesto a disposición para el resto de los nodos participantes en la red; solicitantes de contenido, que desean obtener algún contenido de la red; y ruteadores de contenido, que intercambian mensajes de control con otros nodos de la red (de cualquier tipo). Esta división es conceptual, puesto que un mismo host físico puede corresponder a varios nodos de contenido de diverso tipo.

Las funcionalidades básicas de un nodo fuente de contenido es la entrega de contenido a sus solicitantes y la publicación de su contenido mediante descriptores en mensajes de control. Por otro lado el solicitante de contenido envía mensajes de control solicitando algún contenido en base a un descriptor. Los ruteadores de contenido se encargan de circular de la manera más inteligente posible los mensajes de control, dirigiendo el descubrimiento del contenido dentro de la red de manera de lograr que los solicitantes encuentren fuentes que puedan responder a sus pedidos.

La gran flexibilidad de las redes de contenido (en cuanto a confiabilidad, disponibilidad, redundancia, performance, tiempo de acceso, anonimato, seguridad, etc.) es posible gracias a que:

- Los clientes presentan muchos recursos a compartir (almacenamiento, procesamiento, etc.) instalados en computadores cada vez más potentes; logrando mayor disponibilidad de contenido, entrega más rápida y distribución más eficiente.
- El consumo del ancho de banda ha aumentado enormemente, gracias al incentivo por parte de los proveedores de Internet (ISP) y sus políticas de tarifas planas.

- Dada la reciente aparición de la mayoría de las redes de contenido actualmente utilizadas, en su diseño ha sido posible aplicar diversas técnicas informáticas modernas, en especial la estandarización y las técnicas de Interoperabilidad (por ejemplo el uso de Web-Services [62]).
- Las redes de contenido son redes virtuales, por tanto se puede adaptar su topología y en particular la cantidad y tipo de conectividad de sus nodos, logrando mayor o menor confiabilidad y mayor o menor latencia.
- La inexistencia de una dirección de alojamiento (IP) fija determinan menos vulnerabilidad a diversos ataques, en particular los del tipo de denegación de servicio (Denial of Service: DoS).
- Esta misma característica permite soportar nuevos tipos de aplicaciones de cliente donde no se tiene una dirección IP fija, como peer-to-peer o computación móvil.
- Se reduce la carga administrativa (operación y mantenimiento) de la red, cuando se requieren características de alta disponibilidad, redundancia y defensa contra ataques (debido a que estas características están incorporadas nativamente en el diseño de la red y por tanto no requieren elementos de red extras).
- Desde el punto de vista empresarial, las redes de contenido al utilizar aplicaciones propias del lado de cliente, permiten una fidelización por parte del mismo. Esto es la situación contraria a la de la Web, donde los wrappers son comúnmente utilizados por portales que ofrecen mas funcionalidades y servicios sin importar el verdadero proveedor.
- El “efecto de red” [46], que consiste en que la utilidad crece muy rápidamente con el número de usuarios que utilizan un servicio, en este tipo de redes es muy fuerte, y potencializado por la globalización de la Internet, permite que algunos mercados con determinadas exigencias de servicios encuentren natural la utilización de las redes de contenido para satisfacer su demanda.

En esta sección se consideran un conjunto bastante amplio de características distintivas de las redes de contenido, basándose principalmente en el análisis de los trabajos de H. T. Kung y C. H. Wu [39], y D. S. Milojevic et.al. [42], complementados con una extensa bibliografía adicional.

Agrupamos las diversas características abordadas en la literatura en dos grandes conjuntos: por un lado las características ligadas a la arquitectura de las redes de contenido, y por otro características ligadas a su comportamiento. El primer conjunto se refiere a los aspectos ligados a la forma de construcción e implementación de las redes, abarcando las propiedades siguientes: Descentralización, Autonomía, Agrupación lógica del Contenido, Colocación física del Contenido, Heterogeneidad, Auto-Organización, Interoperabilidad. El segundo conjunto en cambio se refiere a aspectos del funcionamiento tal como percibido por el usuario y abarca las siguientes características: Anonimato, Perfor-

mance, Escalabilidad, Transparencia y Usabilidad, Seguridad, Robustez y Resistencia a Fallas. En muchos casos, decisiones ligadas a la arquitectura tienen efecto directo en las características ligadas al comportamiento; es importante tenerlas en cuenta, ya que desde el punto de vista del usuario, suelen ser las más importantes.

A continuación se define cada uno de estos aspectos de las redes de contenido, discutiendo brevemente los posibles casos que se presentan comúnmente.

Características relacionadas con la arquitectura de las redes de contenido

Descentralización

En una red de contenido la descentralización se encuentra determinada por la capacidad que tiene la red de funcionar con varios nodos de un mismo tipo (fuentes, solicitantes y ruteadores). Los principales modelos de descentralización son el modelo cliente-servidor, el de distribución jerarquizada, y el completamente distribuido.

En el modelo tradicional de *cliente-servidor*, la información se encuentra localizada en servidores centralizados y distribuida a través de la red a los clientes finales. De esta forma se simplifica enormemente el protocolo de comunicación y descubrimiento del contenido, pero pueden aparecer problemas de escalabilidad a nivel del servidor central, resultando en un costo importante de su mantenimiento. En general, la información de control es la que se encuentra centralizada (es decir, la red funciona con un único ruteador) y el contenido se encuentra distribuido entre una cantidad arbitraria de fuentes. Ejemplos de este modelo son Napster[44][45] (un único nodo ruteador), Seti@home [57] (un único nodo ruteador/solicitante), y Avaki [4] (un único nodo ruteador).

En un sistema *completamente distribuido* cada nodo tiene la misma participación en la red, ofreciendo potencialmente la misma cantidad de contenido (existen una cantidad arbitraria de nodos fuentes, solicitantes y ruteadores). De esta forma se reduce en gran medida el costo de recursos y mantenimiento asociado al escalado de los servidores centrales. Pero este tipo de redes tienen que resolver el problema de la vista global; al presentarse de forma distribuida el contenido es necesario consolidar dicho contenido para ofrecerlo a los nodos solicitantes, esto trae aparejado la utilización de protocolos de comunicación más complejos y mayor utilización de la red subyacente (alto ancho de banda), que también puede representar un problema de escalamiento al crecer el número de participantes. Para resolver el problema de la vista global, básicamente existen dos filosofías posibles (que pueden ser combinadas):

- La más comúnmente usada, pero que presenta mayor consumo de ancho de banda (y por tanto mayor limitante al escalamiento), es el broadcast, donde cada nodo solicitante inunda la red solicitando el contenido deseado.

- Otra alternativa es que los nodos fuentes se encarguen de mantener informados a los nodos ruteadores, para que cuando algún contenido sea solicitado, sepan donde se encuentra. Esta alternativa presenta múltiples formas de implementarse, destacándose la utilización de tablas de hash distribuidos [55] o técnicas de ordenamiento estructurado (auto-ordenamiento) [3][5] [39].

Ejemplos de sistemas completamente distribuidos son Gnutella [26] (y sus variantes Limewire[40], BearShare, Gnucleus[27], etc.), Morpheus[24] y Freenet [11][24]. Algunos de estos ejemplos serán descritos más adelante, mostrando en detalle el funcionamiento típico de los tres estilos de redes completamente distribuidas (broadcast, tabla de hash distribuida y auto-ordenamiento).

Entre los dos modelos antagónicos anteriores (cliente-servidor y completamente distribuido) existe una serie muy variada de modelos con *distribución jerarquizada*, donde la idea es agrupar en niveles según la capacidad de contenidos de cada participante. Esta variante es la de mejor escalamiento, pues distribuye mejor la carga tanto de procesamiento como de consumo de ancho de banda de comunicaciones (debe recordarse que las redes cliente-servidor presentan una limitante en escalamiento debido al importante costo de su servidor central, mientras que las redes completamente distribuidas presentan una limitante debido al consumo exponencial de ancho de banda necesario por todos los nodos para participar en la red). Algunos ejemplos son DNS[43], KaZaA[36], FastTrack, Grokster, Groove[30], y OpenNap[49]; todos presentan una cantidad arbitraria de nodos fuente y solicitantes, y una cantidad proporcionalmente mucho menor de ruteadores.

Agrupación lógica del Contenido

Otra característica, que está muy ligada al modelo de descentralización de la red, es la agrupación lógica del contenido. Por agrupación lógica del contenido entendemos la estrategia para asignar los contenidos individuales dentro de grupos de contenido.

Se pueden identificar tres etapas separadas en el proceso de agrupación:

- Mapeo: Un primer paso, desarrollado en general por los nodos fuente, es clasificar el contenido en valores de un espacio de contenido bien definido (distintas estrategias son utilizadas dependiendo del contenido: el caso más estudiado es aquel en el que el contenido corresponde a archivos, sea de datos o multimedia).
- Agrupamiento: Un segundo paso, en general desarrollado por los nodos ruteadores, es mantener lógicamente agrupados en la red los contenidos que presentan valores similares (o idénticos).
- Consolidación: En la mayoría de las redes de contenido se explota la existencia de múltiples fuentes del mismo contenido, ofreciéndole todas las fuentes posi-

bles al solicitante. Si este es el caso, el solicitante debe consolidar el contenido brindado por todas las fuentes, para tener una vista única del mismo.

Existen dos filosofías bien distintas respecto a las variantes de agrupación lógica. La agrupación puede ser semántica o sintáctica según como sea el espacio de contenido al que se mapea en el primer paso:

- **Semántica:** El espacio de contenido es una taxonomía, donde los valores tienen significado. Este tipo de redes permiten una búsqueda por proximidad semántica (es decir puede buscar contenido parecido al solicitado) lo cual puede resultar muy útil para cumplir el objetivo de la red de ofrecer contenido relevante a los nodos participantes.
- **Sintáctica:** En este caso el espacio de contenido no tiene ningún significado, por ejemplo las redes basadas en tablas de hash distribuidas simplemente transforman cualquier archivo en un valor numérico, por tanto se desconoce la similaridad semántica entre valores, impidiéndose la búsqueda por proximidad. Una ventaja de este enfoque es que se obtienen en general sistemas más escalables debido a que no existen criterios o limitaciones de diseño externas para la agrupación del contenido (en estas redes se tienen problemas de escalado en la implementación de búsquedas semánticas de contenido).

Siguiendo la nomenclatura propuesta por H. T. Kung y C. H. Wu en [39], diremos que una red de contenido es *semántica* o *sintáctica* según como se realice su agrupación lógica. Asimismo hay redes particulares donde no se realiza ningún tipo de agrupación lógica, estas se clasificarán como sintácticas.

La agrupación lógica determina las posibilidades de la red de ofrecer una vista global del contenido. En redes con servidores centrales es relativamente sencillo obtener esta vista global. En cambio, en las redes descentralizadas (no cliente-servidor), la misma es el resultado del conjunto de los esfuerzos individuales de los nodos para mantener el conocimiento del contenido existente.

Colocación física del Contenido

Dependiendo de la naturaleza del contenido y de la arquitectura de la red, hay situaciones en que el contenido puede migrarse o replicarse entre fuentes para lograr una mejor performance y/o mayor anonimato. Esto no siempre es posible o conveniente (por ejemplo, las fuentes pueden no permitir alojar un contenido arbitrario puesto que es una forma de violar su propia autonomía). La característica de colocación física del contenido determina el método utilizado para elegir el nodo fuente donde debe ser alojado un contenido. Dos grandes posibilidades se presentan:

- **Colocación Sensible al Contenido:** En este caso se elige el nodo fuente de un contenido específico en función del contenido mismo. Este método facilita enor-

memente la tarea a los nodos ruteadores, ya que a partir de la descripción del contenido buscado, es posible encontrar la ubicación del mismo.

- **Colocación Insensible al Contenido:** cada contenido es colocado sin tomar en cuenta cuál es el mismo. En este método las fuentes deben periódicamente anunciar su contenido o los nodos ruteadores inundar la red en busca del mismo, aumentando el consumo de recursos de transmisión. Este método funciona mejor cuando el modelo de distribución es centralizado (cliente-servidor).

Se dice que una red de contenido es *sensible* o *insensible* al contenido según como se realice la colocación física del contenido. Como mencionamos, hay redes donde no es posible modificar la colocación física de contenido, por tanto estas redes son siempre insensibles al contenido.

Con el fin de diferenciar los conceptos de agrupamiento lógico y colocación física es importante hacer notar que ambos tipos de agrupamientos (semántico y sintáctico) pueden utilizar una colocación física sensible o insensible. Por ejemplo, el sistema de DNS[43] presenta una agrupación semántica (jerarquía arborescente por nombre de dominio) y la colocación del contenido es sensible a dicha jerarquía dado que cada nombre debe ser colocado en el servidor que tiene asignado ese dominio en la jerarquía. Por otro lado las redes basadas en tabla de hash distribuidas donde el contenido es colocado en los nodos según los valores de hash, como Past(Pastry) [17] o OceanStore (Tapestry)[38], son un ejemplo de redes sintácticas sensibles. Respecto a redes insensibles los ejemplos más claros son Gnutella[26] y TRIAD[29]. Gnutella es una red P2P sin agrupación lógica, ni colocación física del contenido, mientras que TRIAD es una propuesta de nuevos protocolos de red, que sustituyen al método actual basado en ruteo IP, búsqueda de hosts por DNS y recuperación de contenidos vía http. Este protocolo permite una red donde el contenido es mapeado al espacio de las URL (agrupación lógica semántica), pero no hay colocación física porque el contenido no debe ubicarse arbitrariamente en ninguna fuente en particular.

Autonomía

La autonomía en una red de contenido corresponde a la capacidad de “independencia” de los nodos participantes. En los estudios clásicos de interoperabilidad [7][19] se definen cuatro tipos de autonomía para un sistema formado por varias componentes:

- **Autonomía de Diseño:** Libertad en la elección de su propio: modelo de datos, lenguaje de consulta, implementación, restricciones, gerenciamiento de datos a ser usado, funcionalidades soportadas, etc.
- **Autonomía de Comunicación:** Capacidad de los componentes de decidir cuándo y cómo responder a requisitos de otros componentes.

- **Autonomía de Ejecución:** Cada componente puede ejecutar sus operaciones sin interferir en la ejecución de operaciones no locales.
- **Autonomía de Participación:** Capacidad de los componentes de decidir cuánto de sus funciones, operaciones y datos compartir.

Muchas de las redes de contenido presentan protocolos de comunicación abiertos y conocidos (otras son propietarios), en ellas existe libertad de diseño excepto en el protocolo.

Dado que una red de contenido es una red virtual que se monta sobre la infraestructura IP existente, la misma utiliza los recursos de los nodos participantes según las características de cada red. Los recursos que poseen los nodos son de tres tipos: procesamiento, almacenamiento y transmisión. Teniendo en cuenta que el contenido que pueden ofrecer los nodos fuentes debe ser satisfeco con alguno de sus recursos, entonces las autonomías de comunicación, ejecución y participación pueden verse simplemente como la autonomía de un nodo fuente a decidir qué y cuanto contenido ofrecer.

La autonomía es un requerimiento muy importan en algunos de los tipos de redes de contenido mencionadas, especialmente para las redes de pares (peer-to-peer) que presentan siempre una completa autonomía de participación.

Heterogeneidad

La heterogeneidad puede presentarse a tres niveles, según la teoría de interoperabilidad entre componentes [7][19]:

- **Heterogeneidad de sistemas:** los nodos presentan diferentes plataformas de hardware y software, diferentes sistemas operativos y/o diferentes protocolos de comunicación.
- **Heterogeneidad sintáctica:** diferentes modelos de datos del contenido en cada nodo.
- **Heterogeneidad semántica:** los mismos contenidos en distintos nodos son representados de manera diferente en cada uno.

En las redes de contenido, es muy común encontrar heterogeneidad de sistemas, donde los nodos presentan distinta capacidad de brindar sus recursos, además de distinto hardware, sistema operativo y versión del software de red de contenido. Esta ultima posibilidad es también frecuente (distintas versiones de software), en cuyo caso puede aparecer también heterogeneidad sintáctica o semántica.

Auto-Organización

En cibernética un sistema auto-organizado es aquel que tiene la habilidad de organizar sus componentes en un framework de trabajo sin la necesidad de control externo [33]. Las redes de contenido presentan un extremo dinamismo en escala, topología, contenido y carga, por tanto, algunas propiedades de auto-ordenamiento son deseables, e incluso necesarias para obtener una buena performance y escalamiento.

Dependiendo del tipo de descentralización utilizado por la red, es posible alcanzar distintos niveles de auto-ordenamiento:

- En los sistemas centralizados el problema es más fácil de abordar, buscando soluciones en el sistema central.
- En los sistemas completamente distribuidos la solución no es sencilla y distintas técnicas son utilizadas. Resaltamos la utilización de hipercubos para la topología y búsqueda de contenido; la replicación y migración de contenido; y los protocolos de conexión/desconexión de nodos, etc. Ejemplos son Past (Pastry) [17], CAN[54], Chord[59], OceanStore (Ta-pestry)[38], etc.
- En los sistemas jerarquizados, la propia jerarquía impone un nivel de organización. Además se utilizan muchas veces algunos de los métodos de las redes centralizadas o completamente distribuidas para lograr mayor ordenamiento. Un ejemplo son las redes SuperPeer como KazaA [36].

Es necesario resaltar que para los sistemas distribuidos no existe una vista global centralizada del sistema, y por tanto el auto-ordenamiento solo puede ser alcanzado mediante criterios locales (de auto-ordenamiento). Un estudio detallado sobre estos conceptos se encuentra en [3].

Interoperabilidad

El concepto de Interoperabilidad clásico [7][19] refiere al acceso uniforme a múltiples fuentes de datos heterogéneas y autónomas. La intención de la interoperación es intercambiar datos y funcionalidades cooperando con un fin común. Los problemas básicos que se enfrentan cuando se realiza un sistema integrado es resolver la heterogeneidad de las fuentes, respetar la autonomía de las mismas, y diseñar de forma que el sistema sea gerenciable y mantenible frente a cambios en la cantidad de las fuentes (escalabilidad) o cambios en las estructuras de las fuentes (evolución). La arquitectura de un sistema interoperable debe entonces, adaptarse a la naturaleza de las fuentes y los objetivos específicos de integración en cada caso.

Las redes de contenido pueden verse como un sistema interoperable en si mismos, montado sobre la infraestructura IP, en general de gran escala, donde los sistemas a integrar son cada nodo y los datos intercambiados son contenidos de distinto tipo (recursos como procesamiento, almacenamiento y transmisión).

Al momento de ser escrito este documento existe una gran explosión de redes de contenido, donde los usuarios que quieren pertenecer a determinada red deben instalar un software específico que los habilite a participar como nodo de la misma. Cuando esta situación madure, comenzará a ser muy útil la creación de software como sistemas integrados de distintas redes de contenido. Los problemas que tendrán que resolver estos futuros softwares de integración son justamente los planteados por la interoperabilidad.

Características relacionadas con el comportamiento de las redes de contenido

Anonimato

Muy relacionado con el concepto de autonomía, aparece el anonimato. Un atractivo de estos sistemas es brindarle a los nodos participantes cierto nivel de anonimato, evadiendo problemas legales (debido a posibles evasiones al derecho de autor del contenido ofrecido), una potencial censura (contenido que se considere no adecuado) y otras ramificaciones (por ejemplo utilizar recursos del trabajo para juegos multi-jugador).

En Freehaven [15] se identifican las siguientes clases de anonimato:

- Anonimato de Autor: El autor o creador del contenido no puede ser identificado.
- Anonimato de Editor: La persona que publica el documento en la red no puede ser identificado.
- Anonimato de Lectura: Las personas que leen o consumen información no pueden ser identificados.
- Anonimato de Servidor: Los servidores de un contenido no pueden ser identificados a partir de la información.
- Anonimato de Información: Los servidores no saben realmente cual contenido están sirviendo.
- Anonimato de Consulta: Los servidores no pueden informar que contenido es usado para responder una consulta dada.

Aplicando estas clases de anonimato al modelo presentando de redes de contenido se obtiene:

- Anonimato de Fuente: No se conoce las fuentes de un contenido dado.
- Anonimato de Editor: No se conoce quien introdujo el contenido a la red. A veces puede ser aceptable conocer las fuentes, pero no la primera fuente.
- Anonimato de Solicitante: No se conoce realmente quien solicita los contenidos.
- Anonimato de Entrega: No se conoce a quién se le entrega determinado contenido (en muchos sistemas se pueden hacer las consultas anónimas, pero si se quiere obtener el contenido es necesario identificarse).
- Anonimato de Contenido: Las fuentes no saben exactamente qué contenidos poseen (solo es posible en redes con baja autonomía de fuente de contenido).

Distintas técnicas son utilizadas por las redes de contenido para mejorar alguno de los aspectos del anonimato: puertos TCP/IP dinámicos, comunicación encriptada, técnicas de Multicast [52], spoofing de solicitante, alojamiento no voluntario [17][55][59], covert paths, etc.

Performance

La performance esta dada por la efectividad y eficiencia con que son utilizados los recursos de los nodos de la red. Dependiendo de la arquitectura de la red de contenido se tiene que la efectividad puede ser la probabilidad de encon-

trar las fuentes de determinado contenido, la precisión del ruteo, etc.; por otro lado la eficiencia está dada por el consumo de ancho de banda, la latencia en el acceso o consulta del contenido, la cantidad de replicas necesarias, la carga del procesador, etc.

Escalabilidad

La escalabilidad de la red de contenido representa la capacidad que tiene la red de funcionar, manteniendo una performance aceptable, frente a un crecimiento en la cantidad de nodos (fuentes, solicitantes y ruteadores).

La importancia de la escalabilidad de una red de contenido depende de los objetivos de utilización de la red. Por ejemplo si la red de contenido es una red p2p para trabajo cooperativo dentro de una institución no tiene sentido pensar en un escalamiento mas allá del tamaño de la empresa (en general a lo sumo unos miles de nodos). Por el contrario si la red de contenido es una red para intercambiar archivos en Internet, es importante que presente un excelente escalado a cientos de millones de nodos, puesto que de esta forma se maximiza el efecto de red y la probabilidad de éxito de la red.

La escalabilidad se encuentra determinada por las características de arquitectura de la red, principalmente por el modelo de descentralización, pero también de otras características como el tipo de colocación física de contenido y la capacidad de auto-ordenamiento. En un sistema centralizado (con un único nodo ruteador) se puede lograr una buena performance y escalabilidad si los nodos fuentes se encuentran distribuidos, este tipo de redes presentan una limitante de escalamiento en la cantidad de mensajes de control que debe manipular el servidor central (Napster[44][45] logró escalar hasta 6 millones de usuarios, cuando por litigios legales de derecho de autor tuvo que cambiar su estrategia y filtrar el contenido ilegal. Seti@home[57] tiene, al momento de escrito este documento, 5 millones de usuarios).

En una red completamente distribuida basada en broadcast la limitante de escalabilidad se presenta en la cantidad de mensajes de control que tienen que manejar los nodos, inclusive aquellos con pocos recursos de la red (como los que se conectan de forma asincrónica). Esta gran consumo de ancho de banda hace costoso pertenecer a la red, además de obtener un alto grado de indeterminismo en las búsquedas. Una solución de compromiso entre los dos sistemas de distribución antes descritos se puede encontrar en las redes jerarquizadas o en las completamente distribuidas pero basadas en una estructura de ordenamiento. Estas redes distribuyen mejor la carga tanto de procesamiento como de consumo de ancho de banda de comunicaciones entre los participantes según la propia capacidad de cada uno, pudiendo escalar en teoría a miles de millones de usuarios.

Algunas variantes de auto-ordenamiento y colocación física del contenido permiten un mayor escalado utilizando básicamente el concepto de acercamiento del contenido a los

potenciales nodos solicitantes, de esta forma se pueden lograr importantes mejoras en la cantidad de comunicación de la red y por tanto en su escalado.

Transparencia y Usabilidad

La transparencia en un sistema distribuido es asociada con la habilidad y simplicidad de conectar el sistema distribuido en un sistema local. Un mayor grado de transparencia determina mayor usabilidad, y por tanto mayor probabilidad de éxito de la red.

La transparencia mas importante para una red de contenido es la de localización, donde se resuelve una relación entre el contenido y su dirección IP. Muy relacionado con la localización se presenta la transparencia de acceso (que determina la facilidad de un nodo para conectarse a la red) y la de movilidad (que tan fácil es para un nodo cambiar de dirección IP). En [8] y [13] se presentan diversos aspectos en los que es deseable la transparencia: Localización, Acceso, Movilidad, Concurrencia, Replicación, Falla, Escalamiento, Administración y de Dispositivo.

Seguridad

Algunos requerimientos de seguridad son pertinentes a todos los sistemas distribuidos, como ser: encriptación, intercambio de claves y firmas de contenido y pares, etc.

Algunas técnicas y aspectos a tener en cuenta, de especial relevancia en este contexto son:

- **Firewalls:** A muchas personas les gustaría pertenecer a una red, pero su conexión es desde una intranet a través de un firewall. Es muy difícil que este tipo de aplicaciones funciones a través de un firewall, puesto que presentan conexiones bidireccionales. Distintas alternativas son utilizadas resaltándose los nodos Relays, Gateways y Proxys antes mencionados.
- **Sandboxing:** Las aplicaciones de redes de contenido requieren ejecutarse en los nodos participantes. Para evitar potenciales ataques de virus o hackers es una práctica común utilizar un ambiente restringido o maquina virtual, donde se ejecute de forma segura la aplicación.
- **Trusting:** Al compartir contenido entre nodos, por lo general anónimos, es difícil evitar la participación de nodos maliciosos que mientan sobre el tipo de contenido ofrecido [12]. El método más utilizado para redes de contenido masivas es utilizar una especie de puntaje penalizando a los malos nodos. Otros métodos por ejemplo pueden ser un chequeo cíclico del contenido cuando se trata de archivos o la utilización de claves.

Robustez y Resistencia a Fallas

Las fallas en una red de contenido pueden resumirse como la incapacidad del sistema a poder ofrecer a un solicitante algún contenido que ha pertenecido a la red. Una red es mas robusta si resiste mejor a las fallas que se le presentan.

Las fallas pueden deberse entonces a:

- **Una solicitud mal realizada:** los protocolos de comunicación muchas veces son propietarios y cambiantes con nuevas versiones del software que se instala en el nodo solicitante. Es posible que por una falla de software o de versión, las solicitudes se encuentren mal formuladas.
- **Una contenido inalcanzable:** Debido a fallas en la red IP subyacente, problemas de particionamiento de la red, o simplemente por problemas de saturación debido a escala, los nodos ruteadores no pueden resolver las solicitudes de contenido, es decir, desconocen y no pueden alcanzar un contenido deseado. Frente a una falla de la red subyacente difícilmente se pueda encontrar solución a nivel de red de contenido; distinto es para el resto de las falla donde múltiples técnicas son abordadas.
- **Un contenido no disponible:** Un problema bien diferente es cuando el contenido puede ser alcanzado por los nodos ruteadores, pero en el momento de la solicitud el contenido no se encuentra disponible, debido a una desconexión de la fuente o porque dejó de compartir ese contenido. Para solucionar este problema dependiendo de los objetivos de la red se utilizan distintas técnicas: la más común es la redundancia de contenido, que significa que al solicitante se le ofrecen todas las fuentes del contenido solicitado (disminuyendo la probabilidad de falla por disponibilidad); también la replicación o migración de contenido es posible en redes donde se permite la colocación de contenido.

Uno de los objetivos de la descentralización es eliminar el punto central de falla. A diferencia de los sistemas centralizados, donde la responsabilidad de las fallas se le adjudican el sistema central, en un sistema descentralizado la responsabilidad radica de distinta forma en todos los nodos participantes, determinando el éxito de la red aquellas arquitecturas que ofrecen mayor robustez.

TAXONOMÍA DE REDES DE CONTENIDO

La temática de redes de contenido tiene en la actualidad un gran desarrollo, pero se encuentra en gran medida en una etapa de propuesta de diversas arquitecturas y soluciones técnicas, y recién aparecen los primeros trabajos que apuntan a abstraer las propiedades comunes a las diversas propuestas, y a clasificar las redes de acuerdo a estas propiedades. Para el caso de las redes peer-to-peer, algunos trabajos que desarrollan taxonomías son [34], [42], [51] y [60].

Para las redes de contenido en general, H. T. Kung y C. H. Wu [39] proponen una taxonomía, basada en las propiedades de Agrupación lógica del Contenido (Content Aggregation), y de Colocación física del Contenido (Content Placement): sensible o insensible al contenido.

En esta sección presentamos una extensión de esta taxonomía, considerando además de las dos propiedades mencionadas, la del modelo de descentralización de la red. La descentralización es quizás una de las características más importantes de las redes de contenido, puesto que el verdadero valor de la red no se encuentra en la capacidad de contenido de un único componente central, sino en la posibilidad que tiene la red de utilizar los recursos de los bordes, lo que potencializado por un buen “efecto de red” [46], permite una reducción de *costos de propiedad* significativa y mejor servicio a los participantes.

Por lo tanto, la taxonomía propuesta tendrá las siguientes tres dimensiones:

- Agrupación lógica del Contenido: *semántica* o *sintáctica*.
- Colocación física del Contenido: *sensible* o *insensible* al contenido.
- Descentralización: *cliente-servidor*, *distribución jerarquizada*, *completamente distribuida*.

Tabla 1. Tipos de redes de contenido según las dimensiones de estudio.

| Colocación física del contenido | Agrupación lógica del contenido | |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| | Sintáctica | Semántica |
| Insensible al contenido | Tipo A Redes sintácticas insensibles al contenido A.1 Cliente Servidor A.2 Distribución Jerarquizada A.3 Distribución Completa | Tipo C Redes semánticas insensibles al contenido C.1 Cliente Servidor C.2 Distribución Jerarquizada C.3 Distribución Completa |
| | Tipo B Redes sintácticas sensibles al contenido. B.1 Cliente Servidor B.2 Distribución Jerarquizada B.3 Distribución Completa | Tipo D Redes semánticas sensibles al contenido D.1 Cliente Servidor D.2 Distribución Jerarquizada D.3 Distribución Completa |

Las tres dimensiones son ortogonales y fundamentales para determinar el diseño y la arquitectura de la red. Principalmente determinan quién y cómo se toman las decisiones de ruteo, localización y descubrimiento del contenido, lo que afecta en gran medida todas las otras propiedades de una red de contenido. Hay doce tipos de redes de contenido, basados en estas tres dimensiones. La Tabla 1 resume las combinaciones que se pueden presentar. Presentamos a continuación ejemplos de diversas redes de contenido existentes que pertenecen a las distintas categorías de la taxonomía.

Tipo A: Redes Sintácticas Insensibles al contenido

A.1 Cliente-Servidor

Los Web Proxy-Cache [58] son un ejemplo claro. En este caso el contenido son las páginas Web, los nodos fuentes son los sitios Web originales (y los propios proxys puesto que cachean contenido), los nodos solicitantes son los clientes del proxy, y el proxy es el único nodo ruteador. Ninguna agrupación lógica es necesaria, y no es posible la colocación de contenido. Este sistema requiere fuerte disponibilidad de recursos (principalmente transmisión y almacenamiento) por parte del proxy y presenta un punto único de falla.

Un ejemplo muy similar son los Mirrors de contenido, que intentan ofrecer otras alternativas al sitio original para la descarga de grandes archivos, por tanto el contenido son archivos, se tiene un único nodo ruteador/fuente que es el mirror, y los nodos solicitantes son los usuarios del mirror.

En general, dentro de esta categoría entran por definición todos aquellos sistemas centralizados donde no existe agrupación lógica (por definición es sintáctica) o la misma es realizada exclusivamente por el componente central, ni colocación física (por definición es insensible). Ejemplos comunes son las ingenierías de buscadores (Search Engines), como Google y Yahoo; los servidores de servicios Web (Web Services [62]) SOAP, WSDL, UDDI; la primera generación de sistemas P2P para compartir archivos (P2P file sharing), como Napster; etc.

Napster [44][45] es la primera aplicación P2P popular que utiliza este modelo. Los clientes de Napster se identifican al ingreso a la red en el servidor central, informándole los archivos que comparte. Cuando un solicitante busca un contenido, lo hace en el servidor central, el cual conoce completamente los archivos compartidos reduciendo la búsqueda a una búsqueda local eficiente. La entrega del archivo se hace directamente entre pares. Cuando una fuente cambia el contenido ofrecido o se desconecta de la red, esto se le informa al servidor central.

Estos sistemas tienen un punto único de falla y vulnerabilidad, además de altos costos de mantenimiento y utilización de recursos en el servidor central.

A.2 Distribución Jerarquizada

Como alternativa a un Web Proxy-Cache central se utilizan los Cooperative Web Caching [2][53], básicamente tienen el mismo objetivo que los Web Proxy convencionales pero el sistema se encuentra formado por un conjunto de Proxys, dividiendo el costo de propiedad y mantenimiento entre varios equipos. Esta distribución además brinda cierto anonimato de fuente y editor, puesto que si los Proxys son autónomos potencialmente no se conoce quién está habilitando determinado contenido. Es necesaria la agrupación lógica de contenido con el fin de que los solicitantes reciban contenido de varios proxys simultáneamente. Esta agrupa-

ción es sintáctica, porque se basa en la URL del sitio, no en un significado.

Las Content Distribution Networks [1][6][10][14][21][37][47][48] son un ejemplo muy similar al anterior. Podrían pensarse como un conjunto de Mirrors que funcionan de forma cooperativa para distribuir archivos de forma menos costosa y más rápida. Es necesaria la agrupación lógica de contenido con el fin de que los solicitantes reciban un archivo de varios mirrors simultáneamente. Esta agrupación es sintáctica, porque se basa en el nombre (o CRC) del archivo, no en un significado.

Dentro de la tercera generación de redes P2P para compartir archivos (P2P file sharing), existe una variante conocida como redes Super-Peer. Las redes Super-Peer presentan un nivel de jerarquía: los nodos con pocos recursos de transmisión son solo fuente/solicitantes (llamados clientes) y están conectados a un único Super-Peer, los nodos Super-Peer son fuente/ruteador/solicitantes y forman entre sí una red completamente distribuida que funciona, en general, mediante el broadcast de solicitudes de contenido a todos los nodos Super-Peer. Los nodos Super-Peer realizan una agrupación sintáctica del contenido (basada en el nombre o CRC del archivo) ofreciéndole a los clientes una vista consolidada del sistema, en ningún caso existe colocación física del contenido. El ejemplo más notorio es KaZaA [36].

La mayoría de los sistemas de suscripción y publicación (Subscribe-Publish Networks), como Siena [9] y Publius [61], funcionan de forma jerarquizada. El contenido de estas redes son notificaciones y suscripciones. Los nodos fuentes son generadores de eventos, informándole a un servidor cercano de tal evento. Los servidores son nodos ruteadores que se encargan de comunicarse entre sí los eventos y anunciárselo a los subscriptores (nodos solicitantes que previamente especificaron su interés en este tipo eventos). En estas redes, los eventos no pueden ser agrupados lógicamente, ni colocados físicamente en los nodos fuentes. Las redes de sensores (Content-based Sensor Networks), como SCADDS[20][31] y SPIN [32] se manejan exactamente con la misma dinámica.

Dentro de esta categoría se encuentran también el sistema de correo electrónico y la mayoría de los sistemas de mensajería instantáneas, como Jabber [35]. El funcionamiento de estos sistemas es idéntico al presentado para las Subscribe-Publish Networks.

A.3 Distribución Completa

Dentro de la segunda generación de redes P2P para compartir archivos (P2P file sharing), existe una variante que funciona mediante el broadcast de solicitudes de contenido a todos los nodos fuentes, no presentando agrupación lógica, ni colocación física del contenido.

En este tipo de redes sucede que ningún contenido es publicado por parte de las fuentes, sino que los nodos solicitantes realizan búsquedas específicas de contenido a todos sus

pares. Los pares que reciben las consultas en realidad actúan como nodos fuentes contestando al solicitante original si poseen el contenido buscado y como nodos ruteadores volviendo a propagar la consulta a sus pares. Para evitar bucles y esperas muy largas, las búsquedas se propagan con un mecanismo de tiempo de vida limitado (TTL). El ejemplo más notorio es Gnutella[26].

Tipo B: Redes Sintácticas Sensibles al contenido

B.1 Cliente-Servidor

La mayoría de las propuestas para computación distribuida (Distributed Computing) [4][22][25][57] presentan una arquitectura cliente-servidor y tienen aplicabilidad a procesos altamente paralelizables de mucho consumo de CPU. Esto se debe a la complejidad que surge del ordenamiento y sincronización de las tareas a realizar.

En este caso el contenido son las tareas a procesar, los nodos fuentes son llamados clientes y ofrecen el tiempo ocioso de su procesador con un fin común (búsqueda extraterrestre, cura del cáncer, etc.). Existe un único nodo ruteador/solicitante central encargado de administrar las tareas que realiza cada cliente (puede pensarse que este nodo central también es fuente porque inicialmente incluye todas las tareas, es decir el contenido). En general las tareas son asignadas a los clientes por parte del servidor considerando la cantidad de computo que requiere la tarea y la capacidad de computo ofrecida por la fuente, por tanto existe una colocación física sensible al contenido. La agrupación lógica es realizada por el servidor central, sin tomar en cuenta el tipo de tarea que realiza cada cliente (por esto la agrupación es sintáctica y la implementación se limita a tareas altamente repetitivas de mucho procesamiento).

Distintas variantes se han presentado, por ejemplo solicitar la misma tarea a varios clientes simultáneamente como forma de redundancia. Dentro de las opciones más conocidas se encuentran Seti@Home[57], Avaki[4], Genome@Home [25], Distributed.net [16], etc.

B.2 Distribución Jerarquizada

No hemos encontrado ejemplos de redes en esta categoría. Se trataría de redes que agrupan lógicamente el contenido de manera sintáctica, y que modifican su lugar de almacenamiento sobre esta base, trabajando sobre la base de una estructura jerárquica.

B.3 Distribución Completa

Dentro de la segunda generación de redes P2P para compartir archivos (P2P file sharing), existe una variante que funciona en base a índice de hash distribuidos [55]. Se puede expresar que la idea básica es asignar mediante hash una identificación única a cada nodo y a cada archivo (contenido). Un archivo se aloja en el nodo activo con identificación más similar a la identificación del archivo. Los nodos son conectados con aquellos nodos que presentan una iden-

tificación con el mismo prefijo, formando una topología que se conoce como hipercubo.

Los solicitantes deben conocer exactamente la identificación del archivo, de esta forma la búsqueda es sencilla: es necesario rutear una consulta hasta el nodo fuente con identificación más cercana, y allí debe estar el archivo. Puede verse que existe agrupación lógica del contenido (se le asigna una identificación a los archivos), pero la misma es sintáctica porque no se basa en significado; la colocación física debe ser también sensible al identificador del archivo.

Múltiples variantes se han presentado a la propuesta original, dentro de las cuales se incluye múltiples dimensiones de hipercubos, respaldos, etc. Las redes más difundidas son Past (Pastry)[17], OceanStore [38] (Tapestry), Chord [59], CAN[54], etc. Es de hacer notar que muchas de estas redes tienen el objetivo de almacenamiento persistente (en lugar del más conocido de compartir archivos entre pares), y en la nomenclatura se conocen como redes de respaldo (Backup Networks).

Tipo C: Redes Semánticas Insensibles al contenido

C.1 Cliente-Servidor

No hemos encontrado ejemplos de redes en esta categoría. Se trataría de redes que agrupan lógicamente el contenido de manera semántica, pero que no modifican su lugar de almacenamiento, y trabajan sobre una estructura cliente-servidor.

C.2 Distribución Jerarquizada

El sistema TRIAD[29] es una propuesta para un nuevo protocolo de red, basado en contenidos, que sustituya (subsumiéndolo) al método actual basado en ruteo IP, búsqueda de hosts por DNS y recuperación de contenidos vía HTTP. Este protocolo permite una red donde el contenido es mapeado al espacio (con significado) de las URL. Una agrupación lógica se realiza según el dominio de las URL (semántica), pero no hay colocación física porque el contenido no debe ubicarse arbitrariamente en ninguna fuente en particular. Los solicitantes realizan las búsquedas de contenido utilizando el protocolo INRP (Internet Name Resolution Protocol), mientras que el contenido se publica en la red utilizando NBRP (Name Based Routing Protocol).

C.3 Distribución Completa

Dentro de la segunda generación de redes P2P para compartir archivos (P2P file sharing), existe una variante que funciona en base al auto-ordenamiento de la topología [3].

La idea básica es brindar escalabilidad a las redes completamente distribuidas, utilizando un ordenamiento en la topología según la afinidad de tipo de archivos que tengan los nodos participantes. En una situación ideal los solicitantes estarán conectados a los nodos fuentes que con mayor probabilidad tengan el contenido buscado, de esta forma si se utiliza una estrategia de búsqueda basada en broadcast se

disminuirían enormemente los tiempos de respuesta y los costos de comunicación.

En general se prevén distintos planos de ordenamiento con el fin de aumentar la precisión en afinidad. Dado que no se tiene una vista global del sistema, el auto-ordenamiento tiene que ser basado en una optimización local.

Tipo D: Redes Semánticas Sensibles al contenido

D.1 Cliente-Servidor

Como se aclaró anteriormente, las propuestas estudiadas en este documento para computación distribuida (Distributed Computing) se basan en repetir un único tipo de tarea muchas veces, por tanto la agrupación lógica es sintáctica ya que todas las tareas tienen el mismo “significado”.

Se considera que fácilmente pueden existir variantes de redes de pares para computación distribuida donde existan varios tipos de tareas (no solo una única tarea que es necesario repetir muchísimas veces), en estos casos el servidor central debe hacer una agrupación en base al tipo de tarea y por tanto existiría una agrupación semántica. Debe quedar claro que este tipo de computación requiere por parte del servidor central una asignación, ordenamiento y sincronización de tareas mucho más complejo que el caso de un único tipo de tarea, siendo por esto que este tipo de redes tiene mucho menos éxito y difusión.

D.2 Distribución Jerarquizada

Quizás el mejor ejemplo de este tipo de redes es el sistema de nombre DNS (Internet Domain Name System) [43]. El servicio de DNS traduce nombres de hosts o nombres de dominios en direcciones IP, por tanto el contenido son registros de nombres y dominio. Los nodos solicitantes son todas las PC conectadas a Internet o una intranet que utilizan algún servicio tradicional de comunicación, como ser el Web, e-mail, etc. Los nodos fuente/ruteadores construyen una jerarquía de base de datos distribuida, de forma arborescente según el nombre del dominio.

Las consultas por dominios de los solicitantes canaliza el servidor de nombres más próximo (proximidad según nombre de dominio), si este conoce el dominio solicitado le responde al cliente, sino inicia una búsqueda en los servidores de nombre, comenzando por la raíz.

Aquí existe una agrupación semántica del contenido, puesto que los dominios son agrupados según la propia jerarquía de los nombres de dominio. Cada nombre debe ser colocado en el servidor (fuente) que tiene asignado ese dominio en la estructura arborescente, por tanto existe colocación física de contenido sensible a la semántica (dominio).

D.3 Distribución Completa

No hemos encontrado ejemplos de redes en esta categoría. Se trataría de redes que agrupan lógicamente el contenido de manera semántica, y que modifican su lugar de almacenamiento, trabajando sobre una estructura completamente

distribuida. Una propuesta aún en desarrollo, y que podría entrar en esta categoría, el proyecto Edutella.

Edutella [18] es un proyecto para construir una infraestructura de metadatos para redes P2P, basado en JXTA. En Edutella, la información es agrupada por su contenido, manteniendo de manera distribuida entre los nodos participantes un conjunto de metadatos (especificados según el estándar RDF del W3C) que permiten la búsqueda y recuperación semántica de los contenidos en sí mismos. Edutella prevee la replicación de los metadatos en varios nodos, para proveer persistencia de datos, y mejorar la disponibilidad y la carga. Los contenidos en sí mismos podrían también ser replicados, si bien en una primer etapa no se ha implementado este servicio.

CONCLUSIONES

El área de las redes de contenido se encuentra aún en una etapa inicial de desarrollo, existiendo relativamente pocos trabajos que formalicen marcos conceptuales genéricos para el estudio de sus propiedades. Hay gran diversidad de propuestas concretas, reflejada en la existencia de redes de características muy diversas, pertenecientes a las diversas clases de la taxonomía presentada; la lista de características relevantes presentadas en este trabajo puede ser entonces de ayuda a la hora de buscar elementos para comparar las mismas y eventualmente diseñar nuevas arquitecturas.

AGRADECIMIENTOS

Este trabajo recibió el apoyo del Programa de Jóvenes Investigadores de la CSIC, UDELAR, Uruguay, así como del INRIA (Francia), en el marco del programa de Equipos Asociados (equipo PAIR).

REFERENCIAS

- [1] Akamai Homepage, <http://www.akamai.com/>.
- [2] Almeida, J., Broder, A., Cao, P., and Fan, L., "Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol," in Proc. of ACM SIGCOMM'98, September 1998.
- [3] Anceaume E., Gradinariu M., y Roy M. "Self-organizing Systems Case Study: peer-to-peer networks". Reporte interno PI-1535, IRISA, Rennes, Francia. 2003.
- [4] Avaki Corporation. 2001. Avaki 2.0 Concepts and Architecture. White paper. http://www.avaki.com/papers/avaki_concepts_architecture.pdf.
- [5] Babaoglu O., Meling H., and Montresor A.. Anthill: A framework for the development of agent-based peer-to-peer systems. Technical Report UBLCS-2001-09, University of Bologna, Italy, 2001.
- [6] BitTorrent Home Page <http://sf.net/projects/bittorrent/>.
- [7] Bouguettaya, B. Benatallah y A. Elmagarmid. "Interconnecting heterogeneous information systems" Kluwer Academic, 1998.(ISBN 0-7923-8216-1).
- [8] Carpenter B. Internet Transparency. Internet Society Request for Comments 2775. www.faqs.org/rfcs/rfc2775.html, 2000.
- [9] Carzaniga, A., Rosenblum, D., and Wolf, A., "Design and Evaluation of a Wide-Area Event Notification Service," ACM Transactions on Computer Systems, 19(3):332-383, August 2001.
- [10] Cisco, "Cisco Content Networking Architecture," <http://www.cisco.com/go/cdn>.
- [11] Clarke, I., Sandberg, O., Wiley, B., and Hong, T. W., "Freenet: A Distributed Anonymous Information Storage and Retrieval System," in Proc. of ICSI Workshop on Design Issues in Anonymity and Unobservability, 2000.
- [12] Cooper B., Crespo A., Garcia-Molina H. "Protecting the pipe from malicious peers". Technical report, Stanford University, 2002. Available at <http://dbpubs.stanford.edu/pub/2002-3>.
- [13] Coulouris G., Dollimore J. 2001. Distributed Systems. Concepts and Design. Addison Wesley.
- [14] Digital Island Homepage, <http://www.digitalisland.net>.
- [15] Dingledine R., Freedman M.J., and Molnar D. "The free haven project: Distributed anonymous storage service". In Proceedings of the Workshop on Design Issues in Anonymity and Unobservability (LNCS 2009), July 2001.
- [16] Distributed.net Home Page. <http://www.distributed.net>.
- [17] Druschel, P., and Rowstron, A., "PAST: A Large-Scale, Persistent Peer-to-Peer Storage Utility," in Proc. of HotOS VIII, May 2001.
- [18] Edutella Home Page. <http://edutella.jxta.org>.
- [19] A. Elmagarmid, M. Rusinkiewicz y A. Sheth. "Management of Heterogeneous and Autonomous Database Systems" Morgan Kaufmann, 1999.
- [20] Estrin, D., Govindan, R., Heidemann, J., and Kumar, S., "Next Century Challenges: Scalable Coordination in Sensor Networks," in Proc. of ACM MOBICOM'99, August 1999.
- [21] Exodus, <http://www.exodus.com>.
- [22] Folding@Home 2001. foldingathome.stanford.edu.
- [23] Foster, I. 2000. Internet Computing and the emerging Grid. Nature, Dec. 7.
- [24] "Freenet Protocol 1.0 Specification," <http://freenetproject.org/index.php?page=protocol/>
- [25] Genome@Home. 2001. genomeathome.stanford.edu.
- [26] Gnutella Homepage, <http://gnutella.wego.com/>.
- [27] Gnucleus's Home Page <http://www.gnucleus.com/>.
- [28] Goelle P. Keyton-Brown K., and Mironov I. "Incentives for sharing in peer-to-peer networks". In Proc. Of

- ACM Conference on Electronic Commerce, October 2001.
- [29] Gritter, M., and Cheriton, D. R., "An Architecture for Content Routing Support in the Internet," in Proc. of the 3rd USENIX Symposium on Internet Technologies and Systems, March 2001.
- [30] Groove Home Page. <http://www.groove.net>.
- [31] Heidemann, J., Silva, F., Intanagonwiwat, C., Govindan, R., Estrin, D., and Ganesan, D., "Building Efficient Wireless Sensor Networks with Low-Level Naming," in Proc. of the 18th ACM Symposium on Operating Systems Principles, October 2001.
- [32] Heinzelman, W., Kulik, J., and Balakrishnan, H., "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," in Proc. of ACM MOBICOM'99, August 1999.
- [33] Heylighen, F. 1997. Principa Cybernetica Web. pespmc1.vub.ac.be/SELFORG.html.
- [34] Intel, Peer-to-peer work: The Intel(R) Philanthropic Peer-to-Peer Program <http://www.intel.com/cure/>.
- [35] Jabber Home Page: <http://www.jabber.com/>.
- [36] KaZaA. <http://www.kazaa.com>.
- [37] Kontiki Home page. <http://www.kontiki.com>.
<http://help.kontiki.com/enduser/group.jsp?node=1332>.
- [38] Kubiawicz, J., Bindel, D., Chen, Y., Czerwinski, S., Eaton, P., Geels, D., Gummadi, R., Rhea, S., Waterspoon, H., Weimer, W., Wells, C., and Zhao, B., "OceanStore: An Architecture for Global-Scale Persistent Storage", in Proc. of ASPLOS'00, November 2000.
- [39] Kung, H. T., and Wu, C. H. (2002). "Content Networks: Taxonomy and New Approaches". A chapter in "The Internet as a Large-Scale Complex System", Kihong Park and Walter Willinger (Editors), Oxford University Press. 2002.
- [40] Limewire Home Page. <http://www.limewire.com/>.
- [41] H. T. Kung, et al. MotusNet: A Content Network. Report. Harvard University. 2001.
<http://citeseer.nj.nec.com/443175.html>.
- [42] D. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, Z. Xu. "Peer-to-Peer Computing". Technical report HPL-2002-57, HP Labs. 2002. <http://www.hpl.hp.com/techreports/2002/HPL-2002-57.html>.
- [43] Mockapetris, P.V. Internet Domain Name System Standard: Domain names - concepts and facilities. Rfc-Editor Home Page, <ftp://ftp.rfc-editor.org/in-notes/rfc1034.txt>, 1987.
- [44] Napster Homepage, <http://www.napster.com/>.
- [45] "Napster Messages," <http://opennap.sf.net/napster.txt>.
- [46] E. Nicholas and C. Himmelberg. "Critical Mass and Network Evolution in Telecommunications". Selected Papers from the 1994 Telecommunications Policy Research Conference, Gerard Brock (ed.), 1995.
- [47] Onion Tornado-Cache Home Page. <http://onionnetworks.com/>.
- [48] Open Content Network. <http://www.open-content.net>.
- [49] OpenNap Home Page. <http://opennap.sf.net/>.
- [50] Pandurangan, G., Raghavan, P., and Upfal, E. Building Low-Diameter P2P Networks. In Proceedings of the 42nd Annual IEEE Symposium on the Foundations of Computer Science (FOCS) (2001).
- [51] The Peer-to-Peer Working Group. "Taxonomy of Peer-to-Peer Architectures P2P Working Group Version 09.5", Draft for TAC review.
<http://www.p2pwg.org/tech/taxonomy/Docs/P2P-Taxonomy-v095.doc>.
- [52] Peer-to-Peer Working Group at Internet2, Home Page, <http://p2p.internet2.edu/>.
- [53] L. Qiu, V. N. Padmanabham, and G. M. Voelker. On the placement of web server replicas. In Proc. 20th IEEE INFOCOM, 2001.
- [54] Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S., "A Scalable Content-Addressable Network," in Proc. of ACM SIGCOMM'01, August 2001.
- [55] Ratnasamy, S., Shenker, S., and Stoica, I., "Routing Algorithms for DHTs: Some Open Questions," in Proc. of the First International Workshop on Peer-to-Peer Systems, March, 2002.
- [56] Ripeanu M., Foster I., and Iamnitchi A., Mapping the Gnutella network: Properties of largescale peer-to-peer systems and implications for system design," IEEE Internet Computing Journal 6(1), 2002.
- [57] Seti@home Home Page. <http://setiathome.ssl.berkeley.edu>.
- [58] Squid Web Proxy Cache Project, <http://www.squid-cache.org/>.
- [59] Stoica, I., Morris, R., Karger, D., Kaashoek, F., and Balakrishnan, H., "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in Proc. of ACM SIGCOMM'01, August 2001.
- [60] The Universal Plug and Play Forum (Microsoft) <http://www.upnp.org/>.
- [61] Waldman, M., Rubin, A., and Cranor, L., "Publius: A Robust, Tamper-Evident, Censorship-Resistant Web Publishing System," in Proc. of the 9th USENIX Security Symposium, August 2000.
- [62] Web Services Activity at W3C <http://www.w3.org/2002/ws/>.