Universidad Autónoma de Madrid

Escuela Politécnica Superior

Máster en Ingeniería Informática y Máster en Investigación e Innovación en las TIC

TRABAJO DE FIN DE MÁSTER

PLATAFORMA PARA LA LUCHA CONTRA LA PORNOGRAFÍA INFANTIL

Carlos Navarro Criado
Tutor: Álvaro Ortigosa Juárez

14 de septiembre de 2016

PLATAFORMA PARA LA LUCHA CONTRA LA PORNOGRAFÍA INFANTIL

Autor: Carlos Navarro Criado
Tutor: Álvaro Ortigosa Juárez

Escuela Politécnica Superior
Universidad Autónoma de Madrid

14 de septiembre de 2016

Agradecimientos

Especiales agradecimientos al capitán de la Guardia Civil Carlos Igual y a mi tutor Álvaro Ortigosa.

Abstract

Abstract — This paper has studied the problem that represents child pornography on P2P networks, particularly in the BitTorrent network. To this end, this study has been developed with the help of some members of the Technical Unit of the Judicial Police of the Guardia Civil, in order to know the current situation and the problems found in this field. Unlike other networks, BitTorrent does not provide an a priori identification of users.Therefore, the main goal of this project has been to create a platform that can provide support to those researchers dedicated to the fight against child pornography. For this propurse, a study of the BitTorrent network for indicators to recognize users engaged in illegal material sharing has also been conducted. The results of this work are the creation of a platform for fighting on P2P networks and of a monitoring tool to track users that share illicit files on the BitTorrent network. In addition, there have been found indicators to take into account in order to identify users or user groups in the BitTorrent network.

Key words — p2p, bittorrent, child pornography

Resumen

Resumen —

En este trabajo se ha estudiado la problemática que representa la pornografía infantil en las redes P2P, en concreto en la red BitTorrent. Con este fin se ha trabajado con miembros de la Unidad Técnica de la Policía Judicial de la Guardia Civil para conocer la situación actual y los problemas con los que se encuentran en este campo. A diferencia de otras redes, BitTorrent no proporciona a priori una identificación de los usuarios. Por ello, en este proyecto se propone la creación de una plataforma que pueda servir de apoyo a los investigadores dedicados a la lucha contra la pornografía infantil. Asimismo, se ha realizado un estudio de la red BitTorrent en busca de indicadores que permitan identificar a usuarios que se dediquen a compartir material ilícito de este ámbito. Los resultados de este trabajo son la creación de una plataforma para la lucha en las redes P2P y la creación de una herramienta de monitorización de usuarios que compartan archivos ilícitos en la red BitTorrent. Además, se han encontrado la existencia de indicadores a tener en cuenta para identificar a usuarios o grupos de usuarios en las red BitTorrent.

Palabras clave — p2p, bittorrent, pornografía infantil

Índice general

1. Introducción **1**
1.1. Motivación . . . 1
1.1.1. Algunos datos sobre la pornografía infantil en la Red . . . 2
1.1.2. Las redes P2P . . . 2
1.2. Problemática . . . 3
1.3. Objetivos . . . 5
1.4. Estructura del documento . . . 6

2. Planteamiento **7**
2.1. Metodología . . . 7
2.2. Estudios . . . 7
2.3. Herramientas . . . 9
2.4. Resumen . . . 10

3. Análisis **11**
3.1. Alcance del sistema . . . 11
3.2. Casos de uso . . . 12
3.3. Requisitos del sistema . . . 12

4. Diseño **25**
4.1. Arquitectura del sistema . . . 25
4.2. Solución escogida . . . 27
4.2.1. Ventajas y desventajas . . . 27
4.2.2. Alternativas . . . 28
4.3. Diagramas de actividad . . . 28
4.4. Diagramas de estado . . . 33

5. Implementación del sistema **35**
5.1. Sistema . . . 35
5.2. Cliente BitTorrent . . . 40
5.3. Aceptación del sistema . . . 42
5.3.1. Historias . . . 42
5.3.2. Pruebas realizadas . . . 42

6. Estudio de viabilidad de un reconocimeinto automático de usuarios **45**
6.1. Método . . . 45
6.1.1. Procedimiento . . . 45
6.1.2. Participantes . . . 47

6.2. Análisis 47
6.3. Resultados 47
6.3.1. Identificación de características comunes de los registros del usuario monitorizado 48
6.3.2. Ejecución de distintos algoritmos de Machine Learning sobre las bases de datos 48
6.4. Conclusiones 50

7. Conclusiones **53**
7.1. Objetivos alcanzados 53
7.2. Metodología empleada 53
7.3. Limitaciones encontradas 54
7.4. Futuras líneas de trabajo 54
7.5. Conclusiones finales 55

8. Bibliografía **57**

Índice de tablas

3.1. Caso de Uso - Login . . . 13
3.2. Caso de Uso - Creación de cuenta . . . 13
3.3. Caso de Uso - Administración de usuarios . . . 14
3.4. Caso de Uso - Visualización de datos . . . 14
3.5. Caso de Uso - Exportación de resultados . . . 14
3.6. Caso de Uso - Visualización de estadísticas . . . 14
3.7. Caso de Uso - Registro de cliente . . . 15
3.8. Caso de uso - Añadir Archivo . . . 15
3.9. Caso de uso - Eliminar Archivo . . . 15
3.10. Caso de uso - Información del sistema . . . 16
3.11. Tabla resumen requisitos funcionales . . . 16
3.12. Tabla resumen requisitos no funcionales . . . 16
3.13. RF-01 . . . 18
3.14. RF-02 . . . 18
3.15. RF-03 . . . 19
3.16. RF-04 . . . 19
3.17. RF-05 . . . 19
3.18. RF-06 . . . 19
3.19. RF-07 . . . 20
3.20. RF-08 . . . 20
3.21. RF-09 . . . 21
3.22. RF-10 . . . 21
3.23. RF-11 . . . 21
3.24. RF-12 . . . 21
3.25. RF-13 . . . 21
3.26. RF-14 . . . 22
3.27. RF-15 . . . 22
3.28. RF-16 . . . 22
3.29. RF-17 . . . 22
3.30. RF-18 . . . 22
3.31. RF-19 . . . 23
3.32. RF-20 . . . 23
3.33. RNF-01 . . . 23
3.34. RNF-02 . . . 23
3.35. RNF-03 . . . 23
3.36. RNF-04 . . . 24

5.1. Historias comprobadas . . . 42

6.1. Resultados algoritmos sobre Dataset-1 49
6.2. Resultados algoritmos sobre Dataset-1 con eliminación de variables 49
6.3. Resultados algoritmos sobre Dataset-2 49
6.4. Resultados algoritmos sobre Dataset-2 con eliminación de variables 49

Índice de figuras

3.1. Diagrama de caso de uso de la plataforma 13
3.2. Diagrama de caso de uso del cliente 15

4.1. Arquitectura del sistema 26
4.2. Arquitectura del sistema 27
4.3. Diagrama - Registro 29
4.4. Diagrama - Login 30
4.5. Diagrama - Visualización de resultados 31
4.6. Diagrama - Añadir archivo Torrent 32
4.7. Diagrama - Ejecución del cliente P2P 33
4.8. Diagrama - Ejecución Tracker 34
4.9. Diagrama - Ejecución Peer 34

5.1. Diagrama del sistema 36
5.2. Ejemplo de Login 37
5.3. Ejemplo de Recuperación de contraseña 37
5.4. Ejemplo de Visualización de resultados 38
5.5. Ejemplo de Visualización de estadísticas 38
5.6. Ejemplo de Filtrado de resultados 39
5.7. Ejemplo de Exportación de resultados 39
5.8. Diagrama del cliente P2P 41

1

Introducción

1.1. Motivación

El abuso sexual y la explotación sexual de los niños, incluida la pornografía infantil, constituyen graves violaciones de los derechos fundamentales, en particular de los derechos de los niños a la protección y el cuidado que sean necesarios para su bienestar, conforme a lo estipulado por la Convención de las Naciones Unidas sobre los Derechos del niño y la Carta de los Derechos Fundamentales de la Unión Europea (European Commission, 2010).

Las tecnologías de la información han proporcionado nuevos medios de distribución de imágenes de abuso sexual de los niños. Esto parece haber estimulado un aumento en la producción y distribución de las imágenes y, a su vez, ha alimentado dichas conductas a través de en un aumento en el abuso sexual de los niños, ya sea a efectos de producir las imágenes o porque los usuarios de las mismas han buscado cumplir con las fantasías sexuales que han sido alimentadas por el uso de dichas imágenes (Carr, 2004).

La naturaleza global de Internet trae consigo problemas logísticos y jurisdiccionales que añaden otra capa de complejidad. La detección, la prevención, la identificación y el rescate de las víctimas se hacen aún más difíciles. Además, la velocidad a la que la tecnología crece y cambia, resulta en otro problema añadido (Carr, 2004).

En España, tanto la Guardia Civil, como el Cuerpo de Policía Nacional tienen grupos encargados de este tipo de delitos. En el caso de la Guardia Civil, se encuentra en la Unidad Técnica de Policía Judicial el Grupo de Delitos contra Menores y Mujeres.

Existiendo la necesidad que tienen los Fuerzas y Cuerpos de Seguridad del Estado de España (en adelante FFCCSE) de contar con una herramienta capaz de luchar contra el intercambio de pornografía infantil en las redes P2P, se propone como Trabajo de Fin de Máster la creación de una herramienta que ayude en la lucha contra este grave problema dentro del territorio español.

1.1.1. Algunos datos sobre la pornografía infantil en la Red

La red International Association of Internet Hotlines (INHOPE) es una red global encargada de responder a las denuncias sobre contenido ilegal en la web, así como de la lucha contra el material de abuso sexual infantil. Merece la pena atender a algunos datos alarmantes proporcionados por esta organización en su INSAFE-INHOPE Annual Report de 2013, con la finalidad de comprender el fuerte impacto que tiene en la red el contenido sobre abuso sexual infantil. Sólo en 2013 se procesaron 1.210.893 denuncias por dicho contenido en Internet. Mediante la creación de una base de datos unificada, la red de INHOPE empleando el sistema IHRMS (INHOPE Report Management System) en 2011 registró 29.908 denuncias sobre CSAM (Child Sexual Abuse Material), con un incremento del 25 % en 2012 de 37.404 denuncias y alcanzando en 2013 hasta 54.969 informes registrados (un aumento del 47 %) de contenido encontrado en la web.

Este tipo de contenido se encontró en distintos tipos de sitios web, en concreto, el 37 % en páginas web, el 29 % en repositorios de archivos, el 22 % en repositorios de imágenes, el 7 % en redes sociales y el 5 % restante en sitios de banners. Cabe mencionar brevemente acerca del contenido, que en 2013 material denunciado consistía en un 10 % en contenido relativo a niños menores de 3 años, en un 71 % a niños en la pre pubertad (4-12 años) y el 19 % restante correspondía a niños en la pubertad (13-17).

En cuanto a la ubicación de los sitios que alojaban contenido CSAM, el 52 % se encontraban en Europa y el 44 % en Norte América. Así mismo, el 97 % de las denuncias en 2013 fueron reportadas a los cuerpos policiales siendo el 93 % eliminados en una semana.

En España, mediante la asociación Protégeles, se han registrado más de 35.000 denuncias de contenido infantil ilegal. Entre 2001 y 2012 PROTEGELES recibió más de 200.000 denuncias o informaciones, y realizó más de 15.000 notificaciones sobre páginas y archivos de todo el mundo a las unidades policiales de distintos países, y de forma sistemática a la Brigada de Investigación Tecnológica –BIT- de la Policía española, así como a la Guardia Civil.

1.1.2. Las redes P2P

Merece la pena centrar la atención brevemente en las redes P2P y conocer si son o no una vía potencialmente peligrosa para el tráfico de contenido sexual infantil. Las redes P2P permiten a los usuarios buscar y descargar archivos electrónicos directamente de otros ordenadores. Los participantes que descargan software se encuentran conectados a la red junto con los otros usuarios. Éstos suben archivos a la red en carpetas para compartirlos con los demás participantes usando términos de búsqueda para encontrar aquellos archivos que desean descargar. Mientras que la atención mediática se ha centrado en el problema que estas redes suponen para las violaciones de copyright cuando los objetos de descarga e intercambio son archivos musicales o vídeos, también han de tenerse en cuenta las redes P2P como un medio de intercambio de pornografía tanto legal, como ilegal (Wolak, Finkelhor y Mitchell, 2012). De esta forma, y según los estudios de Cisco (2014), el tráfico de intercambio de archivos donde se incluyen todas las redes P2P, corresponde a un 20,9 % y un 18,2 % en 2013 y 2014 respectivamente, siendo la cantidad de datos transmitida en torno a los 6.000 en 2013 y 6.500 petabytes mensuales en 2014, lo cual supone una gran cantidad de archivos siendo compartidos por usuarios de todo el mundo (Cisco, 2014).

En 2003, un estudio llevado a cabo por el US Government Accountability Office estimaba

que entre el 42 al 44 % del contenido compartido por la red P2P KaZaA pertenecía a contenido sexual infantil (GAO, 2003a). Otro estudio como el realizado en 2008 (Steel, 2008) muestra que respecto al tráfico en redes P2P, como por ejemplo en la red Gnutella, el porcentaje de búsquedas relacionadas con la pornografía infantil ascendía al 1 % (de 235,513 consultas) del total de búsquedas realizadas en un periodo de tiempo, y el 1.45 % de los archivos mostrados pertenecían a contenido sexual infantil (de 194,444 archivos mostrados).

Ya en 2011 se estimaba que la mayor cantidad de tráfico (57 %) pertenecía a tráfico P2P y que el 73 % de dicho tráfico corresponde al protocolo BitTorrent (Cowdroy, 2010, citado por Prichard, Watters y Spiranovic, 2011)

De la misma forma, en estudio de un año de duración realizado en 2013 (Wolak, Liberatore y Levine, 2014), se detectaron 244,920 ordenadores en Estados Unidos con 120,418 archivos de pornografía infantil diferentes, de los cuales más del 80 % compartieron menos de 10 archivos durante el año de estudio, y menos del 1 % (n= 915) de los ordenadores realizaron una contribución anual alta de archivos conocidos (100 o más archivos). Si la policía hubiera arrestado a estos últimos contributarios y hubiera eliminado sus archivos offline, el número de archivos CP conocidos distintos disponibles en la red P2P podría haberse reducido hasta en un 30 %. A la vista de los datos, se puede concluir, por lo tanto, que existe un problema grave en lo que respecta al intercambio de datos de contenido sexual infantil en las redes P2P. A diferencia del contenido en páginas webs, las cuales se encuentran y son denunciadas y eliminadas desde la red internacional de INHOPE, con las redes P2P no se puede realizar este proceso de "depuración". Para entender esta diferencia se debe tener en cuenta la descentralización de la red, es decir, comprender que los archivos no se encuentran concentrados únicamente dentro de una fronteras geopolíticas, sino que gracias a Internet, los archivos son compartidos en las redes P2P por todo el mundo. Teniendo en cuenta que los cuerpos policiales normalmente sólo pueden actuar dentro de las jurisdicciones que los competen, luchar contra la transmisión de pornografía infantil pasa a ser una ardua y casi imposible tarea. Si hacer desaparecer un único archivo de contenido sexual infantil es de por sí complicado, cuando este archivo se encuentra además interterritorialmente compartido, se hace necesario la cooperación simultánea de numerosos países para la eliminación de tal archivo, lo cual, en un principio, puede ser inviable. Debido a la idiosincrasia del traspaso de contenido en las redes P2P y a la complejidad que supone la lucha contra la pornografía infantil, se hace imperativo encontrar nuevas formas de lucha y detección de usuarios que cometen este tipo de delito.

1.2. Problemática

Como se ha podido observar, la monitorización e identificación de usuarios que comparten archivos de contenido pornográfico infantil es un fenómeno complejo. Las FFCCSE, y en concreto, la Guardia Civil (GC), tiene el problema de no contar con un cliente capaz de monitorizar la red BitTorrent de archivos PI. En concreto, la GC debe afrontar las siguientes cuestiones:

- La localización de los archivos de BitTorrent de carácter PI para ser monitorizados, y además en ningún momento ser descargados (dado que es un delito). Solo es posible visualizarlos en un ordenador especial propiedad de la GC.
- Aunque la descarga de un único archivo PI es una acción ilegal (Art 189 del Código Penal), es posible que la descarga de un único archivo por un usuario se deba a un error no intencionado por parte del mismo. De esta forma, en un proceso judicial, es necesaria la

recopilación de pruebas fehacientes que identifiquen a un usuario a lo largo del tiempo que no solo haya descargado, sino que haya compartido con otros usuarios. Además, y en la medida de lo posible, hay que tener la certeza de que el usuario ha compartido más de un archivo. Por tanto, el cliente BitTorrent que se implante ha de ser capaz de monitorizar e identificar a los usuarios que compartan archivos PI a lo largo de un periodo de tiempo.

- El sistema que se requiere ha de proteger en todo momento a los usuarios inocentes y en el estudio de este sistema se ha de proteger los derechos de anonimato de todos los usuarios hasta que se demuestre que ha cometido un delito. Aún así, debido a la sentencia del 3 de octubre de 2014 (recurso número 6153/2011), dado que el sistema necesitará almacenar direcciones IP, el sistema debe enmarcarse bajo la protección de la LOPD.
- Existe una limitación del poder de investigación, debido a las restricciones legales, las FCSE tienen limitaciones de qué y cómo pueden investigar estos delitos, quedando amparado por la Ley de Enjuiciamiento Criminal actualizada el 5 de octubre de 2015.
  - Un ejemplo es la obtención de los datos asociados de la persona física o entidad que tiene asociada una dirección IP desde donde se ha compartido un archivo de PI, las FFCCSE necesitan solicitar previamente una autorización judicial para poder solicitar al PSI la identidad del usuario (Artículo 588 ter k de la LEC).
  - Otro ejemplo, relacionado a la forma de investigación es la forma de monitorización, las FFCCSE no pueden ser proveedoras de material para inducir a un usuario a que cometa el delito, a excepción de casos de canales cerrados donde, previa autorización judicial sí que es permitido (reforma LEC del 5 de octubre de 2015), y en los canales abiertos solo pueden "preguntar" a un usuario si dispone de material ilícito y verificar que realmente lo posea.
- Los distintos clientes que tiene la GC proporcionan datos en distintas redes P2P. Teniendo en cuenta que, como se ha expuesto anteriormente, un usuario puede descargarse un solo archivo por error, si se pudiera detectar que ese mismo usuario se ha descargado archivos en distintos clientes, podría considerase un indicativo de intencionalidad.
- El protocolo BitTorrent es abierto a diferencia de otros protocolos P2P que son cerrados, además, está pensado para que terceras partes puedan ampliar el protocolo e implementar su propio cliente. Lo que conlleva este aspecto es que han surgido cientos de programas diferentes para la compartición de archivos a través de la red BitTorrent y cada uno tiene sus propias características, por ello, se puede hacer más difícil monitorizar a todos los usuarios.

Para conocer las características y limitaciones de la red BitTorrent, se ha llevado a cabo un estudio previo de su protocolo. Como resultado se ha identificado una restricción que será el eje principal en este trabajo. El protocolo de BitTorrent está pensado en el anonimato de los usuarios, es decir, el protocolo para compartir archivos está pensado para no aportar información que permita identificar a un usuario a lo largo del tiempo, este aspecto hay que aclarar. Ésta es una de las características principales por las que surge BitTorrent: como respuesta a otros protocolos que existían previamente que sí permiten identificar unívocamente a los usuarios.

Asimismo, otros obstáculos en la identificación de usuarios de la red BitTorrent son:

- El uso de la dirección IP como identificación de los usuarios, dado que disponer únicamente una dirección IP como único elemento identificador puede conllevar problemas, dado que, según se monitorice la red P2P en cuestión, pueden existir direcciones falsas.

- Las direcciones IP tal y como informan las FFCCSE ya no son suficientes para identificar a un usuario cuando se solicita a los PSI la revelación de la identidad del usuario que emplea una dirección IP. Actualmente es necesario proporcionar también el puerto desde el que se llevó a cabo el delito, esto se debe a la sobrepoblación de conexiones a Internet a través de IPv4. Es decir, habitualmente una dirección IP estaba asignada a una única persona que tiene un punto de conexión en su domicilio, o una empresa que en su instalaciones tiene un router. En el caso tradicional, una dirección IP estaba asignada únicamente al individuo o empresa que dispone de una conexión a Internet, a través de este punto de conexión, numerosos individuos se pueden conectar a Internet compartiendo la misma dirección IP. Actualmente, una dirección IP puede estar asignada a distintas personas.
- Además, están apareciendo nuevos mecanismos de comunicación dentro del protocolo de BitTorrent, que permiten un mayor nivel de anonimato, como es el uso de direcciones a través de la la red TOR y la red I2P que complican la identificación de los usuarios puesto que se oculta la dirección IP real del usuario.

1.3. Objetivos

Como fruto de las reuniones realizadas con la UTPJ de la GC, se han establecido una serie de objetivos a plantearse durante la realización de este trabajo. Estos objetivos se agruparán en tres puntos principales.

Como objetivo principal, se establece la creación de un cliente de la red BitTorrent que sea capaz de monitorizar una serie de archivos, estos archivos serán los que sean etiquetados como PI. Para ello, el sistema a crear debe ser capaz de registrar aquellos usuarios que compartan archivos a través de la red BitTorrent, y su posterior implantación en un servidor en las instalaciones de la GC. Por ello será necesario obtener un listado inicial de archivos que sean identificados como PI y sean compartidos en la red de BitTorrent.

Tal y como se ha descrito en la sección de Problemática, tras analizar el protocolo de compartición de archivos BitTorrent, éste está pensado en no identificar de forma unívoca sus usuarios como ocurre con otras redes P2P. Dada esta situación, el segundo objetivo planteado para este trabajo será la realización de un estudio de investigación para la búsqueda de características o patrones que permitan identificar a un usuario de esta red a lo largo del tiempo.

Finalmente, como tercer objetivo de este trabajo, se establecerá en la creación e implantación de una plataforma que permita la visualización de los registros obtenidos del cliente de monitorización BitTorrent, así como de otros clientes de monitorización de PI en otras redes P2P que se encuentran implantados en los servidores de la GC. Esta plataforma debe ser capaz de:

- Permitir incorporar de forma semiautomática la visualización de los datos obtenidos por otros clientes de monitorización de distintas redes P2P que se encuentren actualmente implantados, como aquellos que puedan ser implantados en el futuro.
- La visualización de forma individual y conjunta de los datos obtenidos por los distintos clientes de monitorización P2P.
- La búsqueda y filtrado de resultados en base a unos parámetros preestablecidos de los datos obtenidos por los distintos clientes de monitorización P2P implantados..

1.4. Estructura del documento

Este informe constará de las siguientes secciones y subsecciones:

Una introducción que contendrá la motivación por la que surge la realización de este trabajo, la problemática técnica y legal en torno al tema escogido, los objetivos que se plantean llevar a cabo. Por último la estructura del informe.

En la sección planteamiento, se describirá la metodología escogida para llevar a cabo este trabajo, la planificación temporal mediante diagramas de Gantt, un análisis del estado del arte de los proyectos que podrían ser similares al sistema planteado como tema de este trabajo. Asimismo, se describirá un análisis de las investigaciones enfocadas en la identificación de usuarios en las redes BitTorrent.

En el apartado de análisis se detallará de forma concreta el alcance del proyecto planteado, donde se acotarán aquellos puntos principales a implantar, así como aquellos puntos secundarios deseables. Para la descripción del sistema, se esquematizarán mediante unos casos de usos las funcionalidades que debe disponer el sistema, y, como resultado del análisis de los diagramas de casos de uso y de las entrevistas con la GC, se obtendrán aquellos requisitos funcionales y no funcionales del sistema planteado.

Una vez definidas las características del sistema, se describirá el diseño del sistema en base a la arquitectura del mismo, presentando una serie posibles soluciones y argumentando aquella escogida con las ventajas y desventajas de la misma. En este punto, se describirán de forma gráfica los diagramas de actividad de las funcionalidades del sistema y también los diagramas de estado de aquellos componentes del sistema planteado.

Se describirá como se ha implantado el sistema y el cliente propuesto en el apartado de implantación, que tecnologías se han empleado, y la organización del mismo. Asimismo, se detallarán una serie de historias que el usuario que va a emplear el sistema puede llevar a cabo y comprobar si estas acciones se llevan a cabo con éxito. Por último, se describirán los tests y pruebas realizadas y los resultados de los mismos de los sistemas construidos.

En el apartado investigación, se describirá la investigación llevada a cabo en torno a la cuestión de la identificación de usuarios de la red BitTorrent, exponiendo el procedimiento, metodología, los análisis llevados a cabo y los resultados obtenidos.

Finalmente, en el apartado de conclusiones se podrá encontrar los objetivos alcanzados, las limitaciones encontradas, asimismo, las futuras líneas de trabajo que se presentan una vez realizado el trabajo y por último las conclusiones obtenidas en el transcurso de este trabajo y de la investigación llevada a cabo.

Al final de este trabajo se podrán encontrar las referencias bibliográficas empleadas en este trabajo en formato A.P.A.

2

Planteamiento

2.1. Metodología

El sistema propuesto no está planteado para uso abierto, sino que se encuentra dirigido a un cliente concreto, que, como se ha mencionado, será la Guardia Civil y las FFCCSE. Por lo tanto, y atendiendo a las características concretas de este trabajo, la metodología planteada se ha basado en la reunión regulares con el tutor de este TFM y periódicamente con los expertos de la Sección de Menores y de Explotación Sexual Infantil, dentro de la Unidad Técnica de la Policía Judicial de la Sección del Análisis del Comportamiento Delictivo de la Guardia Civil. En concreto, se llevaron a cabo una serie de reuniones iniciales con los expertos de la Guardia Civil para ver el estado de la situación, problemáticas y proporcionar datos sobre la forma de enfocar este trabajo, como por ejemplo en base a que principios legales se rigen, así como cuál es el proceso de investigación de estos casos.

2.2. Estudios

Piatek, M., Kohno, T., & Krishnamurthy, A. (2008).Challenges and Directions for Monitoring P2P File Sharing Networks — or — Why My Printer Received a DMCA Takedown Notice. In Proceedings of the USENIX Workshop on Hot Topics in Security, San Jose, CA, USA, 2008.

En este trabajo los autores realizan un análisis de los distintos métodos que tienen las agencias de monitorización para analizar en las redes BitTorrent las infracciones de la DMCA, analizan los pros y contras y cómo los usuarios se pueden defender ante ellos. Por ejemplo explican cómo hacer para provocar penalizaciones a direcciones IP que están asignadas por ejemplo a impresoras IP y qué nuevas vías deberían tomar los cuerpos de seguridad y los ISPs. En este estudio se definen dos tipos de monitorización:

- La indirecta, donde se pregunta al servidor tracker qué usuarios están conectados y están

compartiendo los archivos, este listado de direcciones IP puede ser manipulado (falsos positivos).

- La directa donde se pregunta al usuario si dispone un archivo y se comparte con él para verificar que realmente está en posesión de dicho archivo, evita los falsos positivos pero tiene el inconveniente de requerir de muchos recursos computacionales.

Steel, C. (2009). Child pornography in peer-to-peer networks. Child Abuse & Neglect, 33(8), 560-568

El objetivo de este estudio es cuantificar y analizar el "consumo" de pornografía infantil en las redes P2P mediante el análisis de las consultas realizadas en dicha red. Empleando la red Gnutella y monitorizando las consultas, se analizaron el contenido tanto de los resultados, como de las consultas emitidas por los usuarios. Para ello emplearon la red Gnutella, que en 2007 era la red mas popular con un 40 % del tráfico P2P. Los resultados obtenidos mediante su monitorización indicaron que la pornografía infantil representaba un 1 % de las consultas y un 1.45 % el número de archivos mostrados, además estudiaron la jerga empleada y la geolocalización de los usuarios.

Liberatore, M., Erdely, R., Kerle, T., Levine, B. N., & Shields, C. (2010). Forensic investigation of peer-to-peer file sharing networks. digital investigation, 7, S95-S103.

Esta publicación supone el análisis de cómo en las redes P2P Gnutella y BitTorrent se puede recabar información que sea relevante y que sirva como evidencia. Para ello presentan la herramienta RoundUp centrada en monitorizar dichas redes y compartir datos entre distintas investigadores que se instalen la herramienta.

de la Cruz, I. P., Aller, C. F., Garcia, S. S., & Gallardo, J. C. (2010, June). A careful design for a tool to detect child pornography in P2P networks. In Technology and Society (ISTAS), 2010 IEEE International Symposium on (pp. 227-233). IEEE.

Los autores aportan una visión legal y social del consumo de pornografía infantil en España y una propuesta en base a los anteriores aspectos de una herramienta capaz de detectar y registrar usuarios que comparten contenido sensible en distintas redes P2P. Toda esta información está oculta al operador que emplea el sistema hasta que no tenga una orden judicial que permita desvelar dicha información. El reto tecnológico proviene de crear un sistema que sea auditable por terceras personas, que los datos estén protegidos de manipulaciones externas y el anonimato reversible de las pruebas almacenadas. Consecuentemente se propone la creación de un sistema de manejo de autorizaciones para respetar los canales burocráticos respetando así en todo momento la ley y protegiendo rigurosamente la privacidad del usuario hasta que jurídicamente se tenga permiso. De esta forma, los autores tienen muy en cuenta el derecho a la privacidad de los usuarios para que el sistema solo sea empleado con el único fin de luchar contra la pornografía infantil.

Liberatore, M., Levine, B. N., & Shields, C. (2010, November). Strengthening forensic investigations of child pornography on p2p networks. In Proceedings of the 6th International COnference (p. 19). ACM

Liberatore, Levine y Shields proponen una nueva vía al clásico monitoreo y registro de las IPs de las descargas que realizan los usuarios en redes P2P. Esta propuesta consiste en realizar un "etiquetado" en el ordenador de la descarga que identifique, dado un registro llevado a cabo por las fuerzas de seguridad, qué dispositivo es la fuente de las descargas dentro de una red. De esta forma. aportan un valor forense importante, dado que, aunque no aparezca el contenido sensible

dentro del dispositivo, mediante la técnica que proponen se dejan rastros en el ordenador que pueden ser empleados por los cuerpos de seguridad como prueba de delito. Ejemplo de este tipo de rastros son:

- Caché del cliente p2p: rastros temporales de las descargas en el disco duro del ordenador que muestren la existencia del archivo.
- DNS lookup: visualizar los registros DNS del ordenador que indiquen que el ordenador investigado efectivamente se ha comunicado con el monitor.
- Logs del cliente: la aplicación del cliente P2P puede almacenar registros, en ellos se puede provocar al cliente que registre un evento con la dirección IP del monitor, lo que da una evidencia de que se ha compartido el archivo monitorizado.

2.3. Herramientas

P2P PATROL to Combat Child Pornography (2004)

La Distributed Computing Industry Association (DCIA) lanzó una serie de programas denominados P2P PATROL (PEER-TO-PEER PARENTS AND TEENS REACT ONLINE)

- El primer programa que fue lanzado como componente enviaba a los usuarios una notificación en tiempo real cuando realizaban consultas de archivos relacionados con la pornografía infantil dentro de programas P2P que están asociados a la DCIA.
- Otros módulos incluyen la notificación a los proveedores de servicio de internet (ISPs) y a las agencias y cuerpos de seguridad de los estados.
- Por último incluía una herramienta de enfoque educativo para los usuarios en general, donde les ayuda a detectar, denunciar y eliminar el contenido que se encuentren.

PROGRAMA VICUS (2010)

Trabajo de fin de carrera donde la autora creó una herramienta de monitorización de archivos con contenido sexual infantil. Esta herramienta está siendo empleada actualmente por la Guardia Civil en la búsqueda de usuarios que comparten dicho contenido en la red P2P Emule. Como los usuarios emplean constantemente un identificador único en el protocolo Emule, dada una base de datos de archivos conocidos con contenido PI, se monitorizan las descargas de los archivos conocidos para procesar aquellos en los que se demuestra que los usurarios se han descargado un buen número de dichos archivos a lo largo del tiempo. Se pretende que esta herramienta sea integrada en el sistema propuesto en este TFM, aunque esto ocurrirá en una segunda fase del proyecto.

ICAC TASK FORCE

El Internet Crimes Against Children Task Force Program es un red nacional de Estados Unidos compuesta por agencias de justicia y fuerzas y cuerpos de seguridad. Se dedican a la investigación y persecución de las personas envueltas en el abuso y explotación de menores relacionados con Internet. Disponen de una herramienta de monitorización restringida a las agencias estadunidenses y permiten el acceso a otros agencias de los distintos países.

La herramienta de monitorización de la que disponen es capaz de monitorizar la compartición de archivos en diferentes redes y redes P2P y permite visualizar los datos a través de una interfaz web. Esta herramienta ha sido inspiración para la creación de la plataforma planteada, pues debido a que dispone de un registro básico de los archivos monitorizados aunque sin aportar, en el caso de la red BitTorrent, información que permita identificar a usuarios. Esta herramienta deja de ser útil para las FFCCSE de España, dado que como se ha descrito en la sección de problemática, el problema de la red BitTorrent radica en la identificación de usuarios a lo largo del tiempo.

2.4. Resumen

Actualmente existen numerosos programas en las distintas redes de P2P tanto para luchar contra la piratería, como contra la pornografía infantil. Algunos de estos programas son conocidos (RoundUp) y se conoce abiertamente cómo trabajan, mientras que existen otros de los cuales no se tiene información sobre el funcionamiento pero se conoce como actúan. Se ha podido observar que hay dos grandes grupos de monitores; los directos y los indirectos. Por un lado, los indirectos fueron muy utilizados hace años y aún existen pero son fácilmente manipulables (producen falsos positivos) siendo las evidencias recogidas poco fiables. Por otro lado, se encuentran los monitores activos que interactúan con los usuarios "compartiendo" archivos, por lo que a diferencia de los indirectos, las evidencias son más fuertes aunque no tiene por que ser totalmente concluyentes.

Algo que se ha observado es que ninguno analiza la problemática de que los usuarios cambien de direcciones IPs, es decir, un seguimiento a largo plazo de los usuarios. Esto se debe a que, mientras que en España es necesario estar seguros que a lo largo del tiempo un usuario ha estado compartiendo contenido ilícito para poder conocer su identidad, en otros países al primer aviso las FFCCSE pueden conocer la identidad del usuario y actuar.

Puesto que en redes como BitTorrent no existen identificadores unívocos, la parte de investigación se plantea buscar alguna forma de detectar a individuos a lo largo del tiempo aunque la dirección IP empleada varíe.

3 Análisis

En esta sección se plasmará el conjunto de características que definirán al sistema. Asimismo, se especificarán las distintas funcionalidades necesarias para que el sistema pueda lograr los objetivos planteados.

Para ello, se acotará el alcance del sistema, definiendo los requisitos necesarios y los distintos casos de uso que lo compondrán.

3.1. Alcance del sistema

Atendiendo a las limitaciones de tiempo que el TFM plantea - no excederse de 600 horas en el desarrollo del trabajo - y de limitaciones de recursos tanto humanos, legales, como materiales, el proyecto se centrará en:

Creación de una plataforma de gestión de las distintas redes P2P:

- Login: para acceder a las funcionalidades del sistema se deberá de iniciar sesión mediante un usuario y contraseña.
- Administración de usuarios: un usuario con rol de administrador, podrá dar de alta usuarios, editarlos o eliminarlos.
- Gestión de archivos: dado que para monitorizar las descargas de los archivos sensibles en las redes P2P se deben identificar previamente que archivos son los que hay que vigilar, se proponen dos vías:
  - Añadir archivos manualmente: registrar los archivos manualmente especificando en base a la red P2P objetivo, los parámetros necesarios.
  - Búsqueda automática: se proveerá de un módulo de búsqueda de archivos, tanto por las redes, como por internet en busca de archivos sensibles mediante una serie de parámetros.

    - Eliminar archivos: Un usuario administrador, podrá eliminar aquellos archivos de monitorización que desee.

- Visualización de datos: todos los registros recabados por los clientes P2P se podrán visualizar mediante filtros interactivos se podrán acotar y visualizar de forma más dinámica. Asimismo, se podrá visualizar con filtros la gestión de usuarios, archivos, entre otros.

    - Exportación de resultados: los datos mostrados y acotados por filtros podrán ser almacenados en un archivo en función del formato especificado para su posterior uso.
    - Guardar resultados: los datos mostrados y acotados por filtros podrán ser almacenados dentro del sistema para ser empleados en otras secciones.

- Identificación de usuarios: en esta sección se plantea la inclusión de reglas de identificación de usuarios, dichas reglas serán las obtenidas durante la fase de investigación.

- Visualización de estadísticas: para la visualización del estado global del sistema como de cada uno de los clientes registrados en el sistema de forma resumida y gráfica.

- Estado del sistema: el usuario con permisos de administrador podrá visualizar el uso de los recursos del sistema y los clientes en el ordenador donde se ejecute. Además podrá controlar la ejecución de los distintos clientes (parar, iniciar, etc)

- Registro de cliente: para poder agregar nuevos clientes de monitorización de redes P2P, se propone incorporar un módulo que permita registrarlos mediante la especificación de un formato con las características de dicho cliente.

Creación de un cliente en la red P2P Torrent que sea capaz de:

- Listado de archivos: siguiendo el protocolo de Torrent, para poder monitorizar archivos en esta red, el cliente ha de ser capaz de leer un listado de Torrents que se deban de monitorizar, se tiene que encontrar un archivo .torrent con la información del tracker donde se pueden encontrar y compartir dichos archivo.

- Monitorización de archivos: simulando la compartición de archivos, sea capaz de registrar aquellos usuarios que están en disposición de los archivos que se conocen que son etiquetados como sensibles.

- Comunicación asíncrona: el cliente de monitorización ha de ser capaz de permitir una comunicación con la plataforma anteriormente descrita. En esta comunicación podrá recibir órdenes como la solicitud de información del estado del cliente, la recepción de información de nuevos archivos a monitorizar, o la eliminación de ya existentes.

3.2. Casos de uso

3.3. Requisitos del sistema

A continuación se detallarán todos los requisitos anteriormente definidos, contando cada requisito con los siguientes campos:

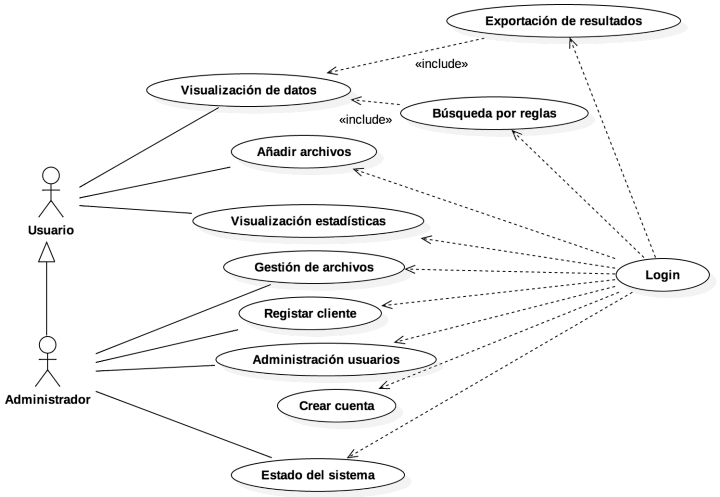

Figura 3.1: Diagrama de caso de uso de la plataforma

Tabla 3.1: Caso de Uso - Login

Nombre	**Login**
Actores	Usuario registrado
Objetivo	El usuario se identificará para acceder a las funcionalidades del sistema.
Pre-condiciones	El usuario se ha registrado previamente en el sistema.
Post-condiciones	El usuario podrá acceder a todas las partes restringidas.
Escenario básico	1.El usuario rellenará los datos del formulario. - Nombre de usuario - Contraseña 2. Enviará los datos

Tabla 3.2: Caso de Uso - Creación de cuenta

Nombre	**Creación de cuenta**
Actores	Usuario administrador
Objetivo	El administrador creará una nueva cuenta en el sistema para que un usuario pueda acceder a parte o a todas las funcionalidades de la aplicación en base al rol impuesto al usuario a registrar.
Pre-condiciones	El administrador tendrá que haber iniciado sesión en el sistema y el usuario no debe existir previamente.
Post-condiciones	La cuenta quedará almacenada en el sistema y el usuario no registrado pasará a convertirse en usuario registrado.
Escenario básico	1.Rellena el formulario de registro - Nombre de usuario - Contraseña - Repetir contraseña - Rol 2.Envía los datos

Tabla 3.3: Caso de Uso - Administración de usuarios

Nombre	**Administración de usuarios**
Actores	Usuario administrador
Objetivo	El administrador podrá editar los datos de los usuarios registrados
Pre-condiciones	El administrador tendrá que haber iniciado sesión en el sistema y el usuario a editar debe existir previamente.
Post-condiciones	Se editarán los datos de un usuario del sistema o será eliminado.
Escenario básico	1. Selecciona el usuario en cuestión. 2. Edita o se solicita la eliminación de un usuario. 3. Envía los datos.

Tabla 3.4: Caso de Uso - Visualización de datos

Nombre	**Visalización de datos**
Actores	Usuario registrado
Objetivo	El usuario visualizará los datos con unos filtros
Pre-condiciones	El usuario se encontrará registrado en el sistema
Post-condiciones	El usuario visualizará los datos solicitados
Escenario básico	1. Selecciona los filtros 2. Enviar la información

Tabla 3.5: Caso de Uso - Exportación de resultados

Nombre	**Exportación de resultados**
Actores	Usuario registrado
Objetivo	El sistema generará un archivo que será enviado al usuario.
Pre-condiciones	El usuario ha seleccionado y filtrado la visualización de datos.
Post-condiciones	El usuario dispondrá de un archivo con los datos solicitados.
Escenario básico	1. Selecciona el formato de los resultados 2. Envía la solicitud

Tabla 3.6: Caso de Uso - Visualización de estadísticas

Nombre	**Visualización de estadísticas**
Actores	Usuario registrado
Objetivo	Se visualizarán las estadísticas individuales o globales del sistema.
Pre-condiciones	El usuario ha de haber iniciado sesión previamente.
Post-condiciones	El usuario visualizará las estadísticas.
Escenario básico	1. Envía la solicitud de visualización de datos de las estadísticas.

Tabla 3.7: Caso de Uso - Registro de cliente

Nombre	**Registro de cliente**
Actores	Usuario administrador
Objetivo	Subir un archivo al sistema para que identifique y tenga en cuenta a un,nuevo cliente P2P.
Pre-condiciones	El usuario ha de haber iniciado sesión previamente y deberá tener un archivo con unos datos y formato especial.
Post-condiciones	El sistema registrará y creará los archivos necesarios para ser incluido en el sistema.
Escenario básico	1. El usuario crea un archivo y lo rellenará. 2. Adjunta el archivo en la solicitud.

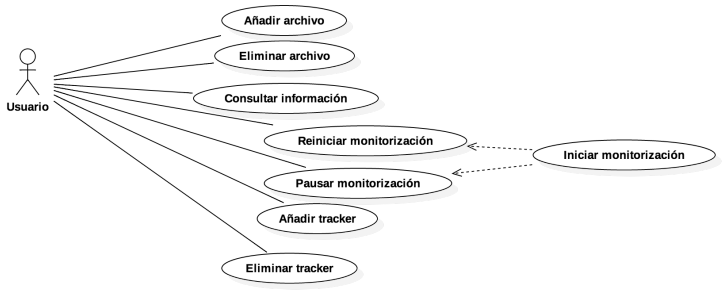

Figura 3.2: Diagrama de caso de uso del cliente

Tabla 3.8: Caso de uso - Añadir Archivo

Nombre	**Añadir Archivo**
Actores	Usuario
Objetivo	El sistema tenga en cuenta este archivo para monitorizar.
Pre-condiciones	El archivo no debe estar previamente registrado.
Post-condiciones	El archivo será registrado.
Escenario básico	1. El usuario indica la ruta del archivo. 2. Envía la solicitud.

Tabla 3.9: Caso de uso - Eliminar Archivo

Nombre	**Eliminar Archivo**
Actores	Usuario
Objetivo	El sistema deje de tener en cuenta este archivo para monitorizar.
Pre-condiciones	El archivo debe estar previamente registrado.
Post-condiciones	El archivo será eliminado.
Escenario básico	1. El usuario indica identificador del archivo. 2. Envía la solicitud.

Tabla 3.10: Caso de uso - Información del sistema

Nombre	**Información del sistema**
Actores	Usuario
Objetivo	El sistema provee al usuario de información del sistema.
Pre-condiciones	-
Post-condiciones	Información del sistema es provista al usuario.
Escenario básico	1. Envía la solicitud.

Tabla 3.11: Tabla resumen requisitos funcionales

Requisitos funcionales	
Código	**Nombre**
RF-01	Sistema: Registro de usuarios
RF-02	Sistema: Inicio de sesión
RF-03	Sistema: Edición de usuarios
RF-04	Sistema: Cierre de sesión
RF-05	Sistema: Cambiar contraseña olvidada
RF-06	Sistema: Recordar nombre de usuario
RF-07	Sistema: Contacto
RF-08	Sistema: Registro de cliente P2P
RF-09	Sistema: Acceso a almacén de datos
RF-10	Sistema: Consultas de datos
RF-11	Sistema: Filtrado de datos
RF-12	Sistema: Creación de archivos de datos
RF-13	Sistema: Comunicación clientes
RF-14	Sistema: Descarga de archivos Torrent
RF-15	Sistema: Procesamiento de archivos Torrent
RF-16	Cliente: Comunicación con Tracker vía TCP
RF-17	Cliente: Comunicación con Tracker vía UDP
RF-18	Cliente: Comunicación con Peer vía TCP
RF-19	Cliente: Comunicación Base de datos
RF-20	Cliente: Canal de comunicación con programas
RF-21	Cliente: Obtención de datos de geolocalización
RF-22	Cliente: Obtención datos de ISP

Tabla 3.12: Tabla resumen requisitos no funcionales

Requisitos no funcionales	
Código	**Nombre**
RNF-01	Multiplataforma
RNF-02	Impacto reducido en el sistema
RNF-03	Idioma castellano
RNF-04	Diseño ampliable
RNF-05	Cliente: diseño asíncrono

Identificador-Nombre: Identificador y nombre extraídos de la tabla anterior para referenciar al requisito sobre el cual se está describiendo.

Prioridad: nivel de preferencia de incluir el requisito en el sistema respecto a los demás, para ello se establece el siguiente orden ascendente de prioridad: Bajo/Medio/Alto.

Estado: estado actual en el que se encuentra dicho requisito, puede encontrarse en los siguientes estados:

- Completado: el requisito se ha incluido satisfactoriamente en el sistema.
- Iniciado: se ha comenzado ha incluir en el sistema.
- No comenzado: no se ha comenzado ha incluir en el sistema.
- Revisión: se ha incluido en el sistema pero no se ha comprobado si se cumple satisfactoriamente.

Necesidad: en una escala de Bajo/Medio/Alto, se definirá la necesidad de incluir dicho requisito en el sistema.

Autor: como todos los requisitos los ha obtenido el mismo autor se obviará dicho campo.

Requisitos relacionados: aparecerán los identificadores de otros requisitos relacionados con dicho requisito ya sea porque necesitan a dicho requisito para existir o viceversa.

Descripción: texto más detallado sobre el propio requisito.

Tabla 3.13: RF-01

RF-01 Sistema: Registro de usuarios		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-02 - RF-10		
Descripción	Para acceder a todas las funcionalidades del sistema, es necesario que el administrador cree una cuenta en el sistema en el sistema para el usuario. Para ello, el administrador ha de rellenar un formulario con los siguientes campos: -Nombre de usuario: combinación alfanumérica,permitiendo que identifique al usuario dentro del sistema de forma unívoca, por tanto no deben existir duplicados y no se permitirá cambiarlo posteriormente -Correo electrónico: correo del usuario necesario para confirmar la creación de la cuenta en el sistema y para otras funcionalidades -Contraseña: combinación alfanumérica de longitud mínima de 6 caracteres y máxima de 13 -Idioma: podrá seleccionar un idioma del sistema de entre una lista predefinida, por defecto el sistema aceptará el inglés y el castellano.	

Tabla 3.14: RF-02

RF-02 Sistema: Inicio de sesión		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-01, RF-03 - RF-10		
Descripción	Para acceder a todas las funcionalidades del sistema, es necesario que el usuario rellene un formulario para acceder a las funcionalidades del sistema. Para ello ha de rellenar un formulario con los siguientes campos: -Nombre de usuario o correo electrónico: combinación alfanumérica - Contraseña: combinación alfanumérica de longitud mínima de 6 caracteres y máxima de 13.	

Tabla 3.15: RF-03

RF-03 Sistema: Edición de usuarios		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-01, RF-02		
Descripción	Para acceder a todas las funcionalidades del sistema, es necesario que el usuario rellene un formulario para acceder a las funcionalidades del sistema. Para ello ha de rellenar un formulario con los siguientes campos: -Nombre de usuario o correo electrónico: combinación alfanumérica - Contraseña: combinación alfanumérica de longitud mínima de 6 caracteres y máxima de 13.	

Tabla 3.16: RF-04

RF-04 Sistema: Cierre de Sesión		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completado**
Requisitos relacionados: RF-01, RF-02		
Descripción	El usuario podrá cerrar sesión, lo que le devolverá a la página inicial y se dejará de poder acceder a las secciones restringidas del sistema.	

Tabla 3.17: RF-05

RF-05 Sistema: Cambiar contraseña olvidada		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completado**
Requisitos relacionados: RF-01		
Descripción	En el caso de que un usuario olvide la contraseña de acceso al sistema, puede rellenar un formulario con el nombre de usuario o el correo electrónico y recibirá en el correo electrónico un mensaje con la nueva contraseña.	

Tabla 3.18: RF-06

RF-06 Sistema: Recordar nombre de usuario		
Prioridad: Media	**Necesidad: Media**	**Estado: Completado**
Requisitos relacionados: RF-01		
Descripción	En el caso de que un usuario olvide el nombre de usuario de acceso al sistema puede rellenar un formulario con el nombre de usuario o el correo electrónico y recibirá en el correo electrónico un mensaje con el nombre de usuario asociado al correo electrónico introducido.	

Tabla 3.19: RF-07

RF-07 Sistema: Contacto		
Prioridad: Baja	**Necesidad: Baja**	**Estado: Completado**
Requisitos relacionados: RF-01, RF-02		
Descripción	En caso de problemas, un usuario puede dejar un aviso al administrador para que el problema sea solventado, para ello, el usuario deberá rellenar un formulario con la siguiente información: -Selección de tipo de cuestión de un listado de cuestiones predefinidas -Correo electrónico: al que el administrador puede contestar en caso necesario Motivo de la cuestión: texto con el contenido de la cuestión.	

Tabla 3.20: RF-08

RF-08 Sistema: Registro de cliente P2P		
Prioridad: Media	**Necesidad: Alta**	**Estado: Incompleto**
Requisitos relacionados: RF-01, RF-02		
Descripción	Para que el sistema pueda acceder a los datos de monitorización de distintos clientes P2P, es necesario registrar y configurar una serie de formularios. El administrador será el único usuario capaz de acceder a esta sección. Una vez rellenados estos datos, se crearán las vistas y accesos al sistema para visualizar los datos. Los formularios a rellenar son los siguientes: -Un formulario obligatorio con los datos de acceso a la Base de Datos donde se almacenan los datos del cliente P2P, que constarán de: - Tipo de Base de Datos - Nombre de usuario de la BD - Contraseña del usuario - Dirección,de la BD - Un formulario obligatorio en el que se deberá seleccionar: - El nombre de la tabla principal donde se registran los datos de monitorización del cliente. Una vez seleccionado el nombre de la tabla principal aparecerán distintos campos. - Selección,del campo asociado a la dirección IP. - Selección del campo hora. - Selección,del campo identificador del archivo - Un formulario opcional al que se le puede añadir acciones de comunicación con el cliente, por cada acción se deberá rellenar los siguientes campos: - Nombre de la acción - Listado de acciones predefinidas - Tipo de acción - Cadena de caracteres que indique el comando o dirección de la acción.	

Tabla 3.21: RF-09

RF-09 Sistema: Acceso a almacén de datos		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-08		
Descripción	El sistema podrá acceder a los datos almacenados por cada cliente P2P registrado en el sistema.	

Tabla 3.22: RF-10

RF-10 Sistema: Consulta de datos		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-08		
Descripción	El usuario podrá acceder a los datos registrados por los distintos clientes P2P registrados en el sistema y almacenados en las bases de datos.	

Tabla 3.23: RF-11

RF-11 Sistema: Filtrado de datos		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-09		
Descripción	El usuario podrá filtrar los datos visualizados mediante la selección de filtros.	

Tabla 3.24: RF-12

RF-12 Sistema: Creación de archivos de datos		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-09, RF-10		
Descripción	El usuario podrá exportar los datos visualizados y filtrados a un archivo que se almacene en el disco local del usuario.	

Tabla 3.25: RF-13

RF-14 Sistema: Comunicación con los clientes P2P		
Prioridad: Media	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-08		
Descripción	El sistema ha de ser capaz de poder ejecutar las acciones definidas durante el proceso de registro del cliente.	

Tabla 3.26: RF-14

RF-14 Sistema: Descarga de archivos Torrent		
Prioridad: Media	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-08		
Descripción	El sistema podrá descargar un archivo Torrent en base al id (btih) proporcionado por el usuario de repositorios de archivos Torrent.	

Tabla 3.27: RF-15

RF-15 Sistema: Procesamiento de archivos de archivos Torrent		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-14		
Descripción	El sistema podrá procesar un archivo Torrent para leer los datos almacenados en su interior para que el cliente P2P pueda monitorizarlo.	

Tabla 3.28: RF-16

RF-17 Cliente: Comunicación con Tracker vía TCP		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-15		
Descripción	El cliente P2P de Bittorrent ha de ser capaz de comunicarse con los Trackers mediante el protocolo TCP para solicitar el listado de usuarios peers que tiene que disponen de un archivo en base al identificador del archivo.	

Tabla 3.29: RF-17

RF-17 Cliente: Comunicación con Tracker vía vía UDP		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-15		
Descripción	El cliente P2P de Bittorrent ha de ser capaz de comunicarse con los Trackers mediante el protocolo UDP para solicitar el listado de usuarios peers que tiene que disponen de un archivo en base al identificador del archivo.	

Tabla 3.30: RF-18

RF-18 Cliente: Comunicación con Peer vía Protocolo TCP		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados: RF-15		
Descripción	El cliente Bittorrent ha de comunicarse mediante el protocolo TCP y solicitar datos de bloques de archivos.	

Tabla 3.31: RF-19

RF-19 Cliente: Comunicación Base de Datos		
Prioridad: Alta	**Necesidad: Alta**	**Estado: Completo**
Requisitos relacionados:		
Descripción	El cliente Bittorrent ha de comunicarse con una Base de datos para leer y escribir datos.	

Tabla 3.32: RF-20

RF-20 Cliente: Canal de comunicación exterior		
Prioridad: Media	**Necesidad: Media**	**Estado: Completo**
Requisitos relacionados:		
Descripción	El cliente ha de poder disponer de un canal que permita a un ente externo acceder a ciertas funcionalidades de cliente.	

Tabla 3.33: RNF-01

RNF-01 Multiplataforma		
Prioridad: Media	**Necesidad: Media**	**Estado: Completo**
Requisitos relacionados:		
Descripción	El sistema y el cliente han de ser diseñados de forma que sean compatibles para distintos sistemas operativos por ejemplo Windows y GNU/Linux.	

Tabla 3.34: RNF-02

RNF-02 Impacto reducido en el sistema		
Prioridad: Media	**Necesidad: Media**	**Estado: Completo**
Requisitos relacionados:		
Descripción	Debido a las limitaciones técnicas de los sistemas donde podría ejecutarse el cliente y el sistema, estos han de ser diseñados para que consuman la menor cantidad de recursos posibles.	

Tabla 3.35: RNF-03

RNF-03 Idioma castellano		
Prioridad: Media	**Necesidad: Media**	**Estado: Completo**
Requisitos relacionados:		
Descripción	La interfaz del sistema ha de encontrarse en castellano, pues los usuarios manejan este idioma.	

Tabla 3.36: RNF-04

RNF-04 Diseño ampliable		
Prioridad: Media	**Necesidad: Media**	**Estado: Completo**
Requisitos relacionados:		
Descripción	Tanto el sistema como el cliente deben estar preparados para poder disponer más funcionalidades sin que requiera un gran esfuerzo.	

4

Diseño

4.1. Arquitectura del sistema

A continuación se presentará el diseño de la arquitectura del sistema a desarrollar, presentándolo en el nivel más alto posible de abstracción, explicando los distintos componentes que lo conforman.

Como primera aproximación, la arquitectura a emplear en este proyecto será la clásica estructura Cliente-Servidor, es decir, contamos con un cliente que realizará peticiones solicitando información a un programa servidor.

Tal y como se describen, tanto en los requisitos, como en los objetivos del proyecto, se concluye que el sistema a implementar corresponde a un modelo de sistema centralizado, en el que solamente habrá un dispositivo aunque no ha de descartarse la posibilidad de que se amplíe en un futuro. Las arquitecturas clásicas existentes que subyacen de un sistema centralizado son la arquitectura C/S – Cliente-Servidor. La arquitectura C/S es empleada cuando distintos componentes se conectan a través de la red a un sistema central. Esto ocurre cuando unas funcionalidades están ubicadas en la parte servidor y es el cliente el que se comunica con el servidor para solicitarle una respuesta ante una petición. Por otro lado, se encuentra la arquitectura P2P – Peer-to-Peer -, que a diferencia de la arquitectura C/S, cada dispositivo puede funcionar tanto como cliente como de servidor ante el resto de ordenadores con los que se encuentra interconectado, pero debido a la limitación de dispositivos se descarta esta arquitectura pues es más típico en sistemas distribuidos.

En este proyecto existe una separación lógica en el sistema que, debido a las restricciones descritas en los requisitos no funcionales, puede llevar a ubicarse en distintos ordenadores – separación física -. Solo uno de los ordenadores - o grupo de varios funcionando como uno solo – será el que provea de servicios al resto, convirtiéndose así en el "servidor", mientras que el resto serán los que realicen las peticiones a éste, denominándose de esta forma "clientes". Consecuentemente, el modelo de arquitectura que se ha adoptado es el de Cliente-Servidor para el sistema.

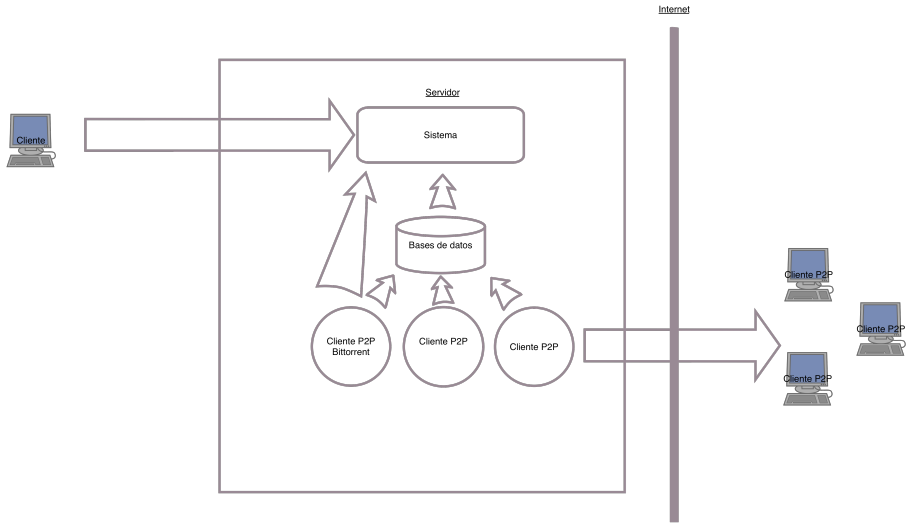

Figura 4.1: Arquitectura del sistema

En cuanto al cliente BitTorrent, el modelo lógico de arquitectura empleado por el protocolo es el P2P, pues todos los clientes del sistema funcionan tanto como clientes como servidores. En cambio el modelo que se empleará será únicamente el del C/S puesto que en ningún momento el cliente BitTorrent a implementar compartirá archivos al resto de usuarios y solo los solicitará. Asimismo, el protocolo BitTorrent requiere contactar como cliente a los servidores Trackers para conocer la existencia de los usuarios de la red. Por tanto, el sistema escogido será únicamente el de C/S, donde los servidores serán los múltiples usuarios.

A continuación se describirán cada uno de los elementos que conformarán esta arquitectura en el sistema propuesto:

- Cliente: software encargado de interactuar directamente con el usuario y de comunicarse con el servidor para solicitar peticiones al sistema.
- Servidor: es el software que hay que implementar dentro un ordenador accesible vía red por los clientes instalados en los distintos terminales de los usuarios. Éste se encargará de comunicarse con los clientes, proveer de servicios tales como el registro dentro del sistema, gestión de la cuenta del usuario, etc. Dentro de él se encuentran los siguientes elementos que deben ser diferenciados del propio servidor:
  - Sistema: encargado de comunicarse con los clientes, tanto el cliente empleado por los usuarios como con los clientes P2P para algunas funcionalidades.
  - Base de datos: elemento donde se almacenará por cada usuario una cuenta de acceso al sistema y toda la información recopilada de los clientes P2P.
- Cliente BitTorrent: software encargado de monitorizar archivos identificados por los usuarios y que almacena en la Base de Datos dichos registros.

Además, tanto el cliente como el sistema emplearán una arquitectura de tipo MVC (modelo-vista-controlador). Esto es, una arquitectura típicamente orientada a interfaces gráficas tales como páginas web. Este modelo tiene el objetivo de separar los componentes en distintas capas para la reutilización de código y su desacoplamiento, facilitando el desarrollo, ampliación y el mantenimiento de la aplicación.

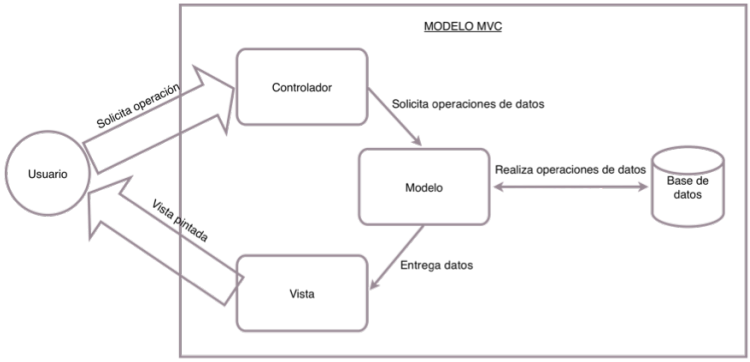

Figura 4.2: Arquitectura del sistema

A continuación se describirán los elementos concretos que conforman la arquitectura de este sistema:

- Modelo: el modelo se encarga de la parte de los datos (lógica de negocio), es decir, la manipulación directa de los datos que el controlador solicitará. En el proyecto, el modelo dentro del cliente estará conformado por los recopiladores de datos de aplicaciones externas como también por los componentes que gestionarán datos privados de la aplicación y también entrarán dentro de esta capa los comunicadores con el servidor. En el servidor en la parte de la aplicación Web, formarán parte del Modelo los repositorios que interactúan directamente con la BD. Dentro del módulo clasificador también existirán componentes que correspondan con esta capa.

- Vista: la vista es la parte de interfaz de usuario con la que va a interactuar directamente el usuario. Se encarga de mostrar de forma estética los datos que recibe de la capa modelo. Tanto el cliente como el servidor contarán con elementos que forman parte de esta capa, como lo son las plantillas y los layouts.

- Controlador: el controlador se encargará de recibir y procesar las peticiones que el usuario envía a través de un cliente. Los usuarios interactúan mediante las interfaces generadas por la capa Vista, mientras que el cliente móvil enviará automáticamente peticiones sin necesidad de una interfaz gráfica. A continuación, el controlador responderá a estas peticiones de dos formas distintas: por un lado, solicitará a la capa Modelo los datos necesarios que serán visualizados mediante la capa Vista, y por otro lado, contestará a estas peticiones sin necesidad de pasar los datos por la capa Vista.

4.2. Solución escogida

4.2.1. Ventajas y desventajas

La solución planteada surge de comparar los pros y contras de las distintas formas de enfocar cómo solucionar o crear un sistema lo más sencillo y eficaz posible para atenerse a los requisitos establecidos.

Aspectos a tener en cuenta

Un punto que se consideró bastante importante desde el inicio del planteamiento del sistema fue que la implementación y uso del sistema y el cliente P2P en el servidor donde pueda ser instalado ejerciera el menor impacto posible al consumo de los recursos - RNF-11 Optimización de recursos -. Esto es, el uso del procesador del propio servidor, el uso de la memoria RAM, el espacio de almacenamiento no volátil y el uso de las comunicaciones con la red.

Sistema

En un principio se consideró que el sistema pudiera residir en un sistema distribuido de servidores, pero debido a las limitaciones impuestas de recursos disponibles se descartó que a corto/medio plazo se pudiera disponer de mejores capacidades de procesamiento. Por tanto, todo el sistema ha de residir en un único servidor que contendrá tanto las Bases de Datos, la aplicación del sistema y los clientes P2P de monitorización. Se empleará el modelo MVC debido a la capacidad de ampliación, abstracción y desacoplamiento de este modelo, para ello, se implantará mediante la tecnología Web dado que no requiere implantar desde cero todo el sistema y solo se ha de concentrarse en la implantación de las funcionalidades específicas del sistema propuesto.

Cliente P2P

El cliente BitTorrent va a ser el que mayor impacto de consumo de recursos que va a ejercer al servidor donde resida, puesto que va a tener que comunicarse con todos los usuarios que de la red BitTorrent que posean los archivos i, teniendo en cuenta que además va a compartir espacio con otros clientes P2P activos, por ello, se ha planteado el uso de un sistema asíncrono

Conexión a internet

Como los dispositivos móviles modernos disponen de dos formas de conexión a internet, se hace imperativo analizar las ventajas e inconvenientes de estos dos tipos de conexión.

4.2.2. Alternativas

Sobre posibles alternativas, para la implantación del cliente BitTorrent existen numerosas librerías creadas mediante distintas tecnologías. El problema de estas librerías reside en lo cerradas que son y que no permiten personalizar ciertas acciones necesarias para el cliente propuesto. Por ello, se desarrollará desde cero un cliente adaptado a las necesidades del proyecto.

Con respecto al sistema planteado, existen numerosas tecnologías, librerías y frameworks para implantar el sistema, por ello, se optará por un sistema Web que emplee algún framework para no tener que construir el sistema desde cero.

4.3. Diagramas de actividad

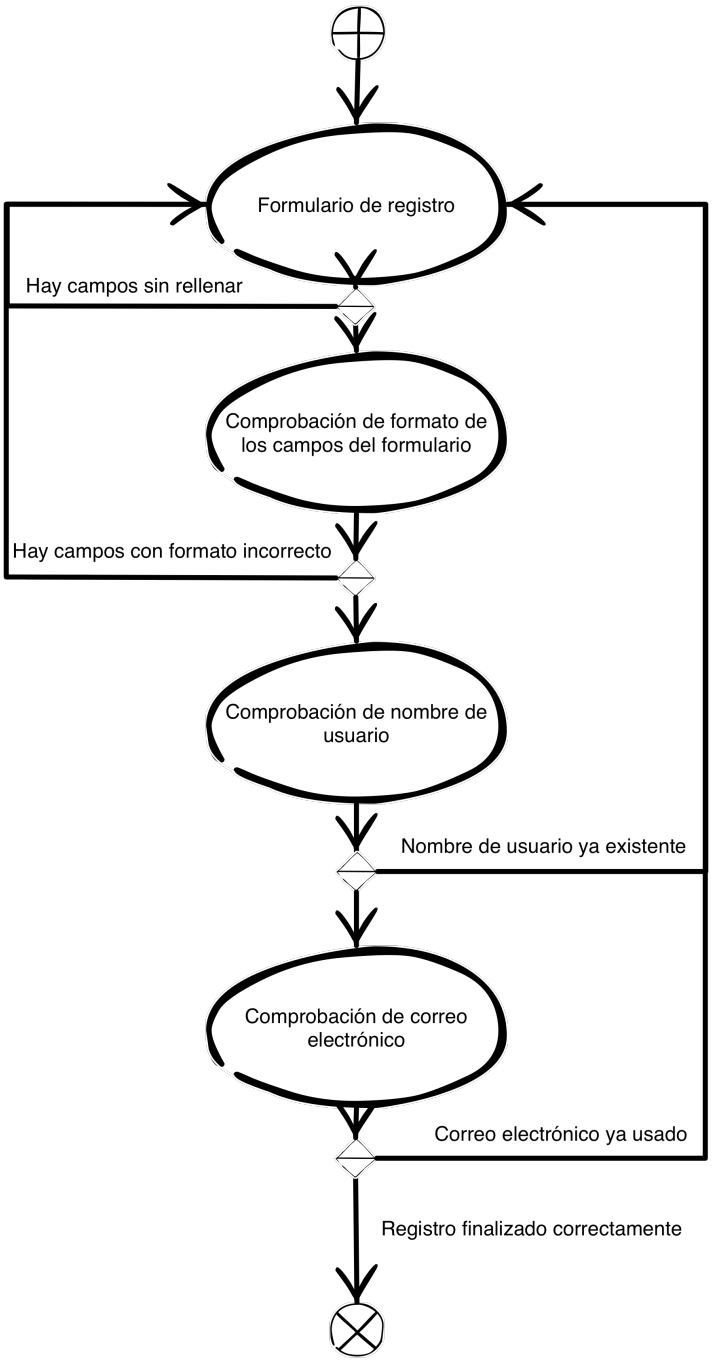

Figura 4.3: Diagrama - Registro

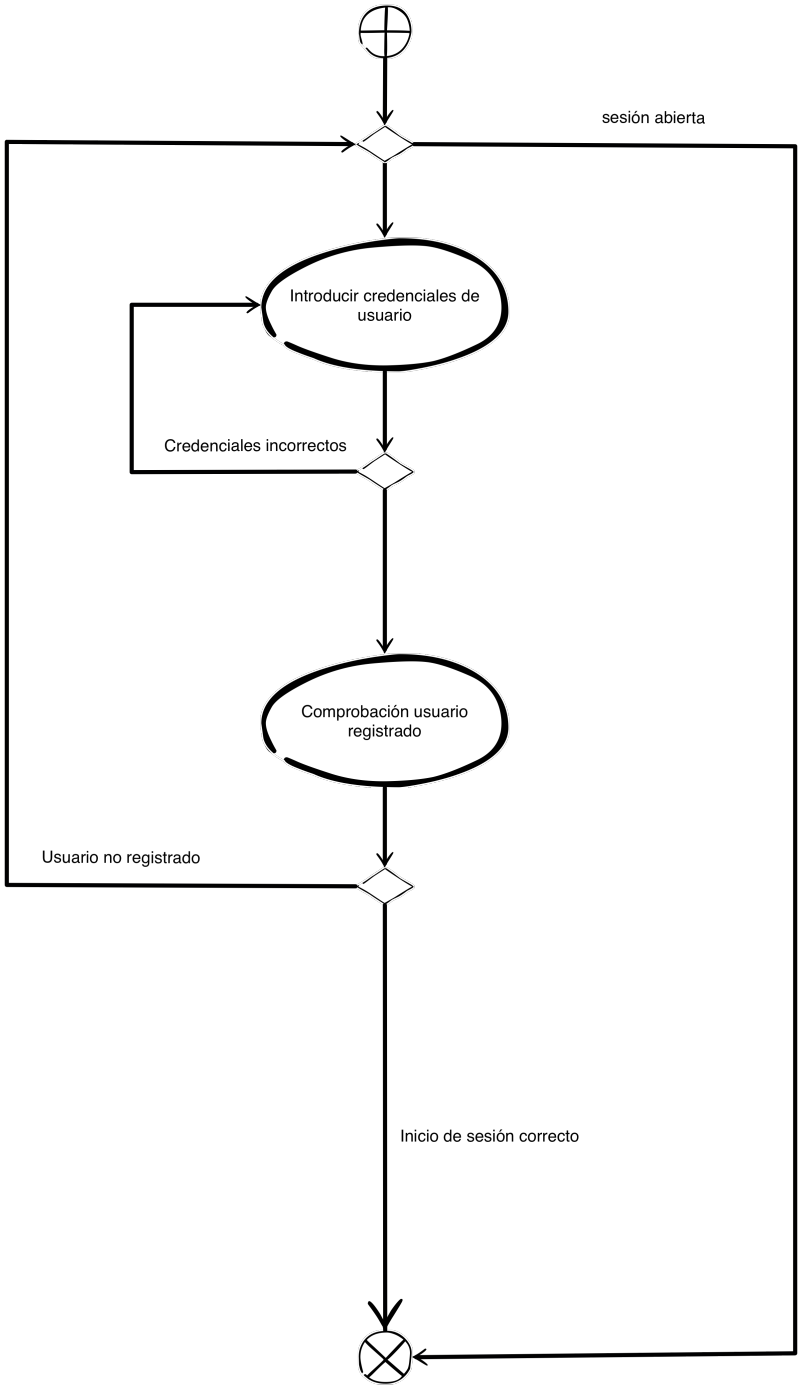

Figura 4.4: Diagrama - Login

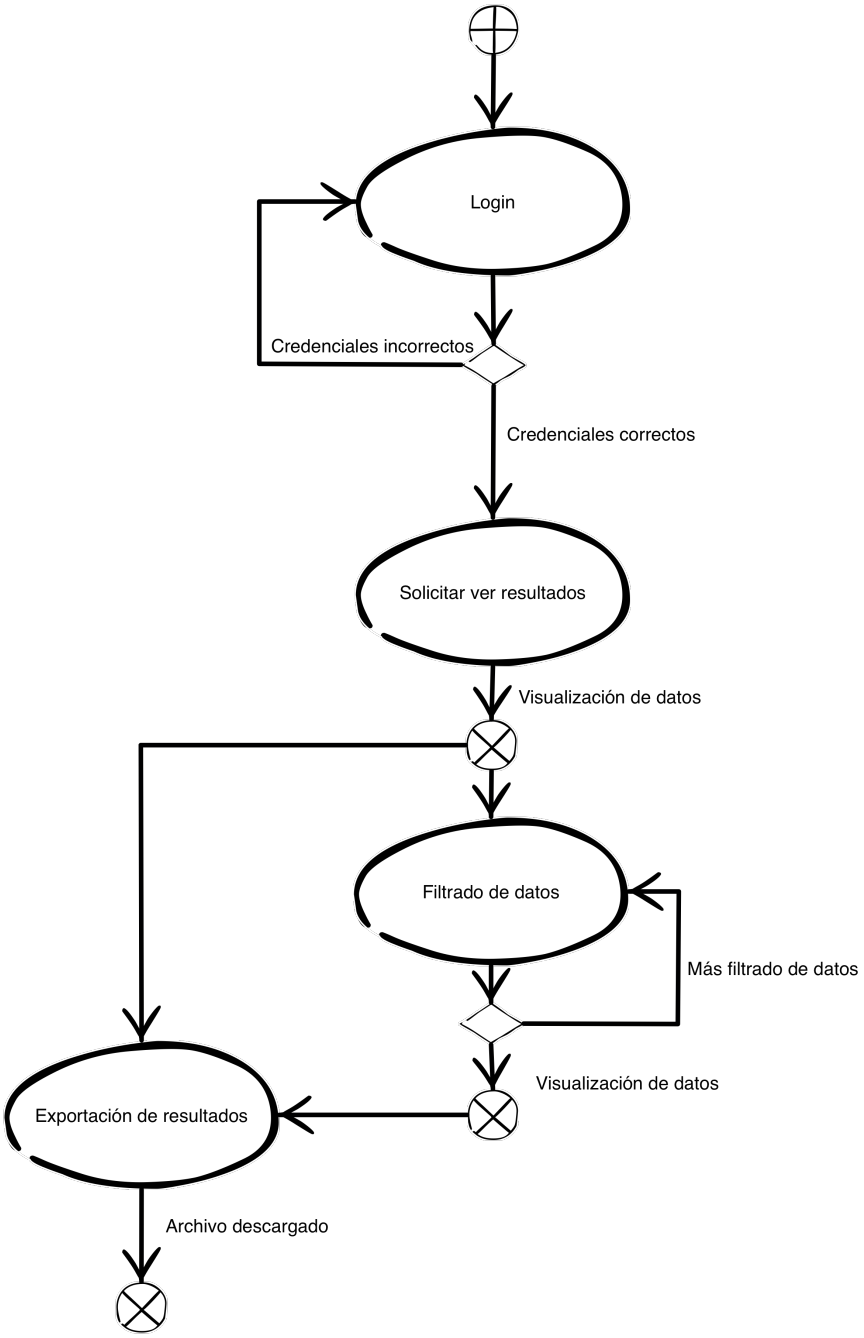

Figura 4.5: Diagrama - Visualización de resultados

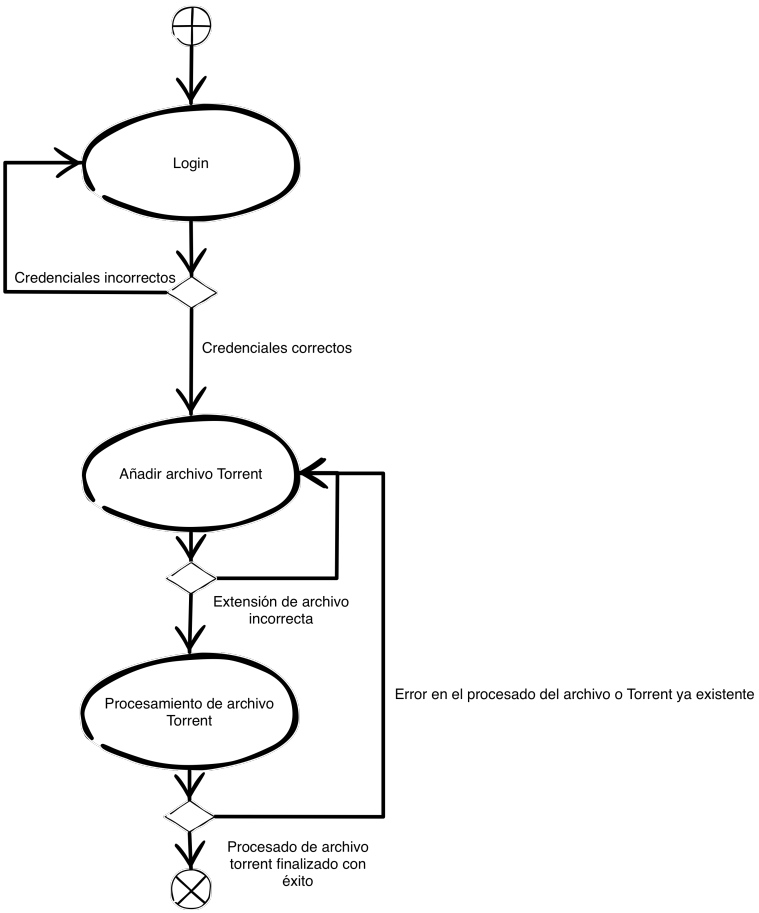

Figura 4.6: Diagrama - Añadir archivo Torrent

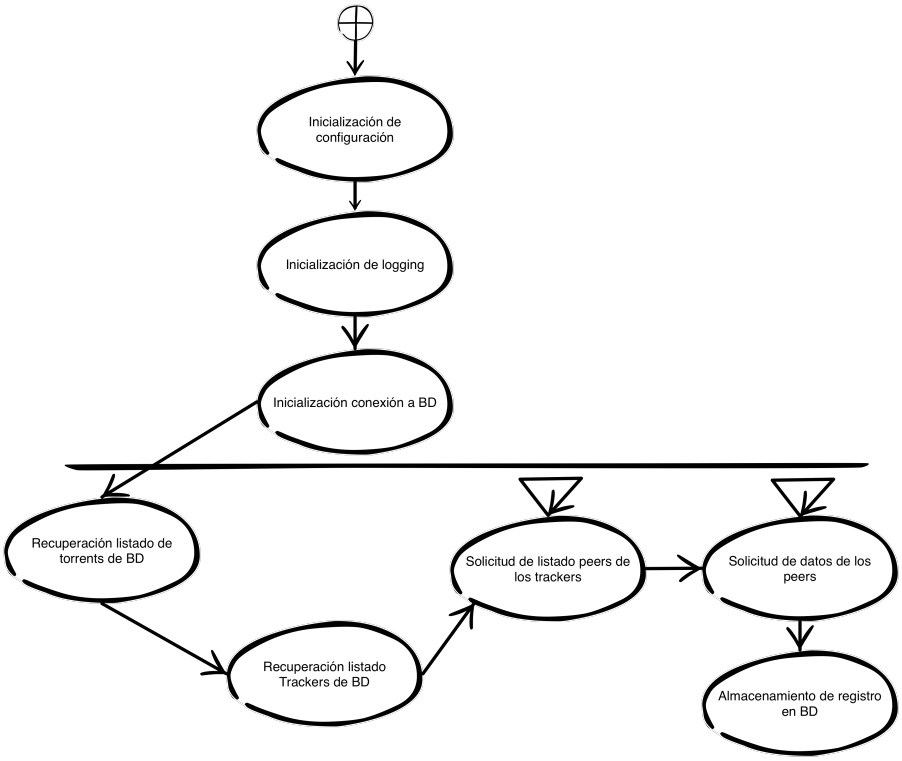

Figura 4.7: Diagrama - Ejecución del cliente P2P

4.4. Diagramas de estado

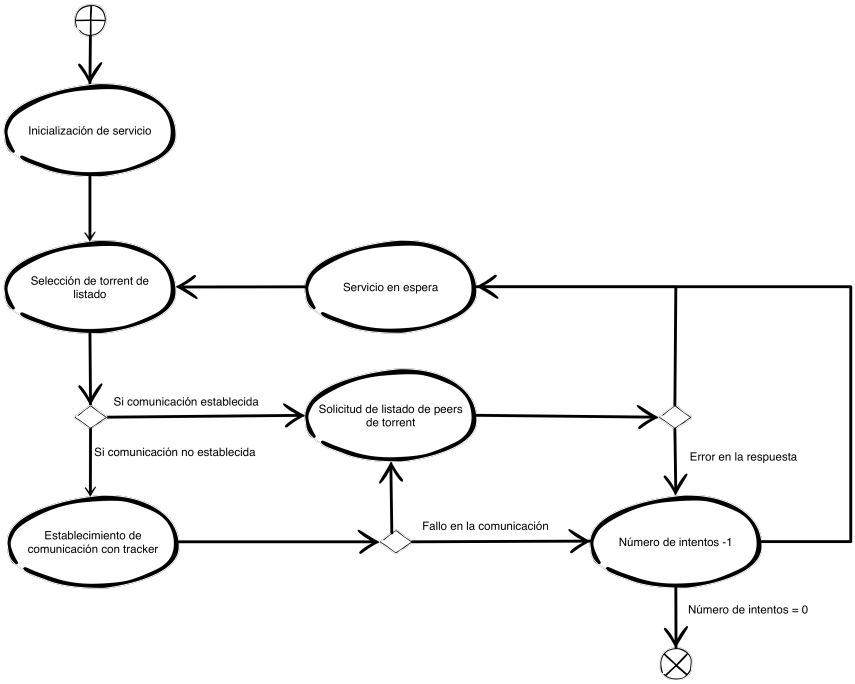

Figura 4.8: Diagrama - Ejecución Tracker

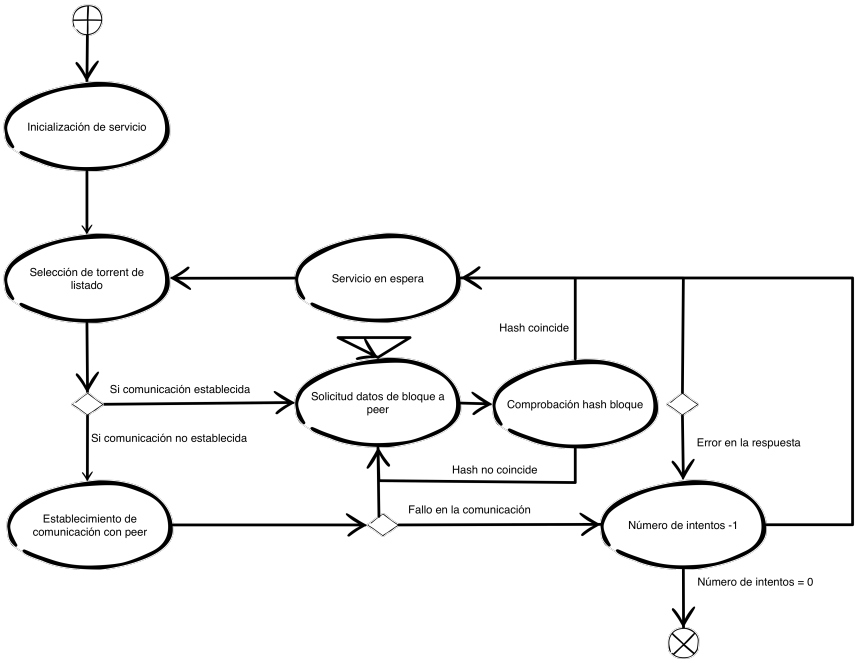

Figura 4.9: Diagrama - Ejecución Peer

5

Implementación del sistema

En esta sección se describirá cómo se ha implementado el sistema planteado, así como el cliente de monitorización BitTorrent.

Como forma de organización para el desarrollo del sistema se ha empleado un sistema de control de versiones Git para poder codificar con un historial de versiones. Esta opción permite tener un registro de todos los cambios y poder revertir fases en caso necesario, asimismo, permite el trabajo colaborativo aunque de momento no haya sido necesario y también facilita la distribución del sistema.

5.1. Sistema

A continuación se detallará cómo se ha implementado el sistema principal

Como se ha mencionado anteriormente, para el desarrollo de la aplicación web del servidor, se ha empleado el modelo MVC. Para el desarrollo del servidor se ha empleado el lenguaje de programación Python y el framework Flask para agilizar su desarrollo, puesto que este framework provee de numerosas librerías de apoyo además de estar orientado hacia la creación de aplicaciones web mediante el modelo MVC. Asimismo, para la visualización de los datos, creación de formularios y gestión de los usuarios se ha empleado la librería de Flask-admin que facilita la creación de las capas de Vista y Controlador.

Controlador: Son los elementos encargados de recibir las peticiones delegando el procesamiento más complejo a los servicios y mostrando los datos obtenidos en las vistas. En esta capa se encuentran los siguientes elementos:

- Controlador de usuarios: encargado de recibir todas las peticiones relacionadas con las cuentas de usuario. A continuación se detallarán las distintas operaciones implementadas:
    - Registro: el proceso de registro recibirá los datos necesarios para crear una cuenta en el sistema, verificando que compruebe que el formato de los datos es correcto,

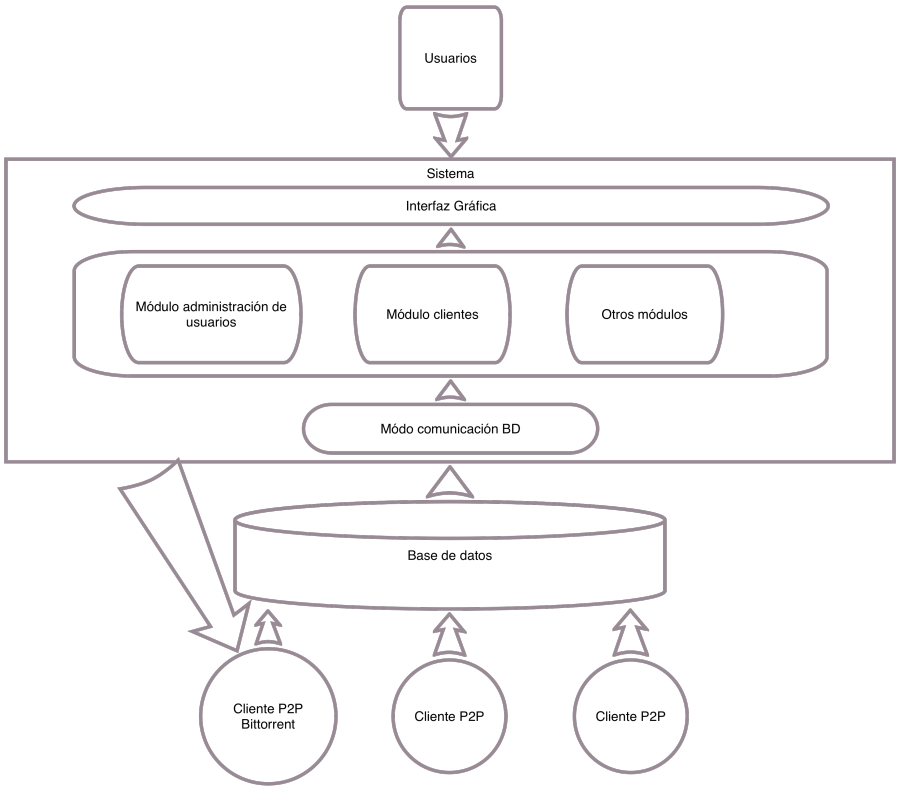

Figura 5.1: Diagrama del sistema

que no exista duplicidad en el nombre de usuario o el correo electrónico introducido comparándolo con el resto de cuentas en el sistema. Según las verificaciones realizadas contestará al cliente con un mensaje de éxito y enviará un correo electrónico para que active la cuenta o, en caso de error, contestará con un mensaje con el motivo del error.

- Inicio de sesión: el controlador recibirá los datos necesarios para el inicio de sesión, comprobará que sean correctos y contestará en caso de éxito con un mensaje de éxito y en caso contrario con un mensaje de error.
- Solicitud de cambio de contraseña: el usuario puede solicitar cambiar su contraseña, para lo que enviará el nombre de usuario asociado a la cuenta. De esta forma, el controlador verificará si existe y enviará un correo electrónico con el formulario si es correcto, o enviará un mensaje de error en caso contrario.
- Cambio de contraseña: controlador que mostrará si procede un formulario para que el usuario cambie la contraseña, verificando que se cumpla el formato de los datos introducidos por el usuario.
- Recordar nombre de usuario: si un usuario se ha olvidado del nombre de usuario asociado a la cuenta, podrá solicitar que se le recuerde dicho nombre enviando el correo electrónico asociado a la cuenta para que pueda recibir un correo electrónico con el nombre de usuario.
- Edición de usuarios: si el usuario que ha iniciado sesión tiene rol de administrador, podrá enviar los datos de los usuarios que desee modificar.

- Controlador del cliente P2P: encargado de proporcionar los datos generados por los distintos clientes P2P registrados en el sistema. Las distintas acciones que permite son:

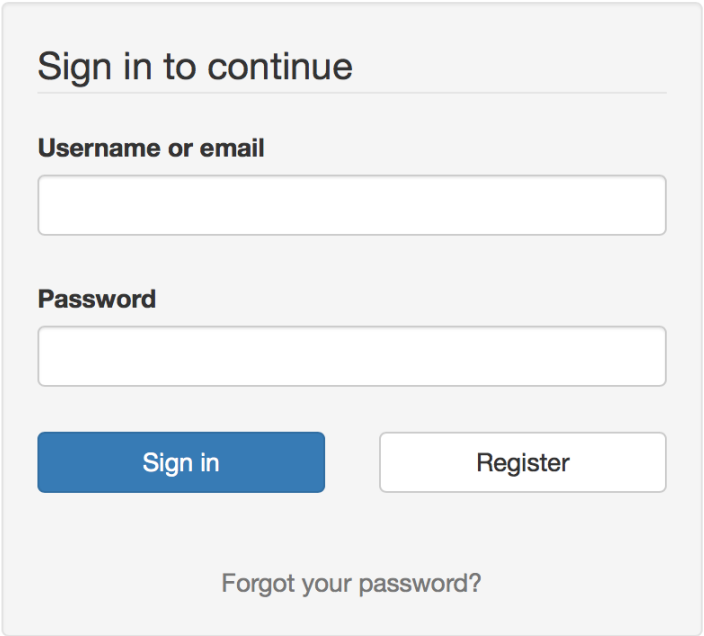

Figura 5.2: Ejemplo de Login

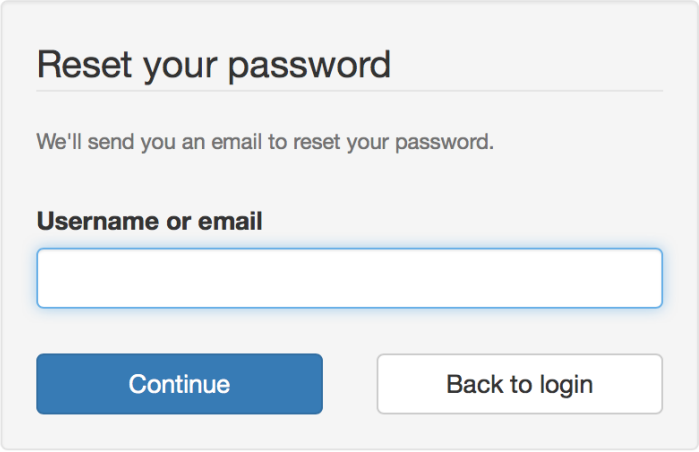

Figura 5.3: Ejemplo de Recuperación de contraseña

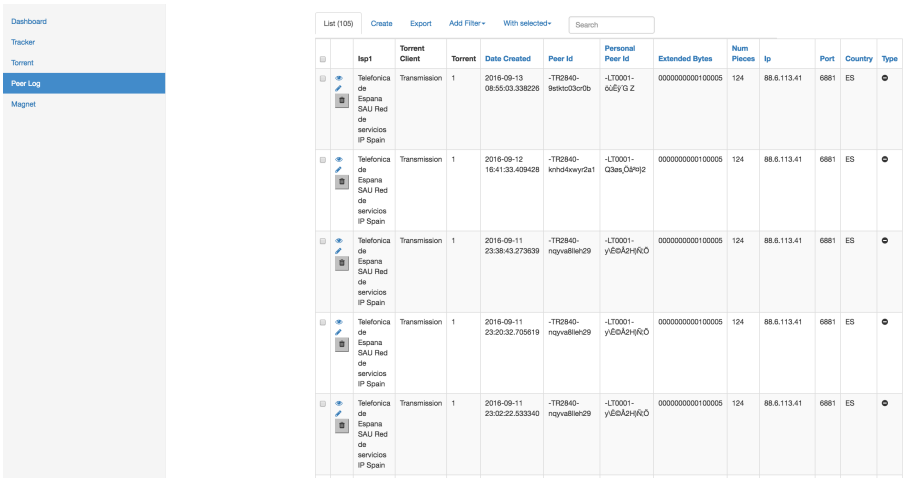

Figura 5.4: Ejemplo de Visualización de resultados

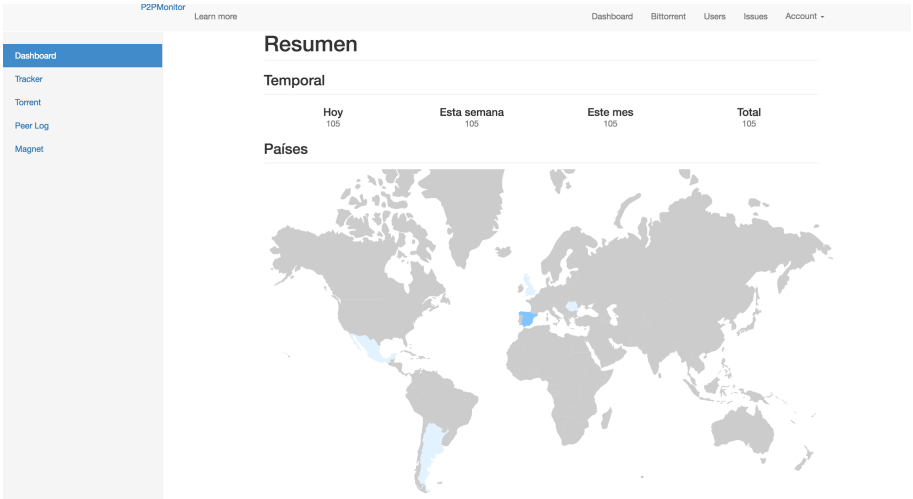

Figura 5.5: Ejemplo de Visualización de estadísticas

- Proporcionar datos: obtener datos de las Bases de Datos y entregarlos a la vista en base a la petición realizada por el usuario, ya sean los datos de varios clientes o de uno solo.
- Generación de archivos: cuando se recibe una petición del usuario para la exportación de datos, el controlador genera un archivo CSV y lo envía al usuario.
- Gestión de datos: solicita la creación, edición o eliminación de los datos generados por el cliente P2P.

Modelo: esta capa es la encargada de interactuar directamente con la BD para realizar tareas tales como consultas, inserciones, actualizaciones o borrado de datos. Se ha implementado un modelo por cada tabla de la Base de Datos de usuarios y del cliente P2P.

Vista: esta capa es la encargada de los elementos visuales que se mostrarán al usuario, tales como plantillas, layouts, entre otros. Para facilitar la creación de las vistas, se han empleado el framework Bootstrap que agiliza la creación de vistas.

- Emails: plantillas de los distintos emails que se puedan enviar al usuario como el de

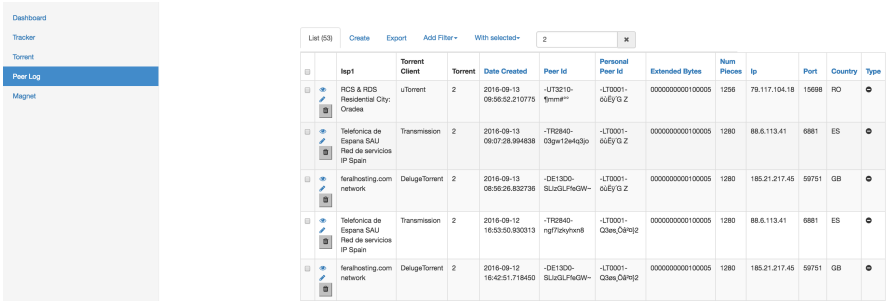

Figura 5.6: Ejemplo de Filtrado de resultados

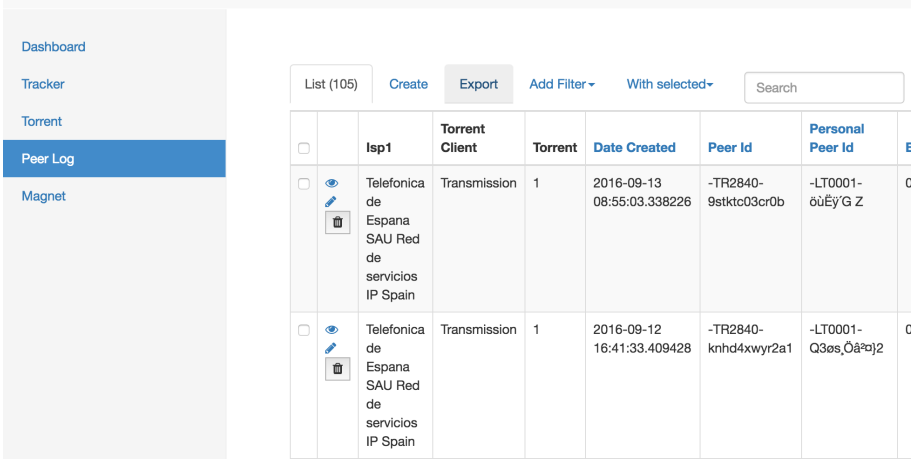

Figura 5.7: Ejemplo de Exportación de resultados

activación de cuenta o el de recordatorio de nombre de usuario.

- Formularios: como es el del cambio de contraseña.

Tareas (Tasks): debido a que el modelo de servidor empleado es de tipo síncrono y que existen funciones que tardan un tiempo en llevarse a cabo (se describirán a continuación) y además, para no concluir en una mala experiencia de usuario, se ha configurado un sistema de tareas asíncrono mediante la librería Celery. Estas tareas son enviadas a un sistema independiente de la plataforma web para que las ejecute y una vez finalizado envían una señal al sistema Web para que la gestione. La diferencia entre este sistema y otro modelo, como por ejemplo el de hilos de ejecución tradicional, es la asincronidad. Esto permite que el sistema pueda llevar a cabo otro tipo de operaciones sin tener que esperar a la finalización de estas tareas. Las tareas que se han implementado son:

- Descarga de archivos Torrent: cuando el usuario solicite la búsqueda y descarga de un archivo Torrent que tiene conocimiento que contiene material ilícito, el sistema lanzará una tarea asíncrona que se encargará de buscar y descargar este archivo Torrent por los repositorios de Internet. Si lo encuentra, lo descargará, procesará y finalmente enviará al sistema un mensaje de éxito o, en caso contrario, de error y el usuario que solicitó la tarea lo visualizará en la vista del sistema.
- Procesamiento de los bloques de archivos Torrent: Los archivos Torrent contienen una larga lista de bloques hash que el cliente P2P tiene que tener en la Base de Datos para confirmar que las piezas solicitadas a los usuarios de P2P son correctas. Debido a esto, el procesamiento de este listado es lento, por ello, se lanza una tarea asíncrona que procesa estos bloques.

5.2. Cliente BitTorrent

En el caso del cliente BitTorrent, éste se ha creado prácticamente desde cero, inspirándose en algunas librerías similares como ha sido la librería bt-twisted (https://github.com/staceysern/bt-twisted). No obstante se ha rediseñado la mayoría de funcionalidades y se ha adaptado las necesidades del proyecto. Para la creación del cliente, se ha empleado el lenguaje de programación Python y se ha optado por el modelo asíncrono descrito en la sección de Ventajas y Desventajas a través del framework Twisted que está diseñado especialmente para la creación de clientes y servidores y la implantación en múltiples protocolos. El cliente P2P ha de coexistir en un servidor con otros clientes P2P y el sistema propuesto. Por ello, el consumo de recursos ha de infligir el menor impacto posible, lo cual se puede optimizar a través del framework de Twisted.

El framework Twisted es un framework de red para programación dirigida por eventos (event-drived) y las ventajas que proporciona son la gestión de las tareas bloqueantes a través de un pool de hilos, en vez de tener que gestionarlos manualmente y tener que crear un hilo por cada proceso como puede ser la comunicación con cada Tracker y cada Peers, pues este tipo de tareas puede conllevar al consumo de muchos recursos del ordenador tal y como nos informó la GC con uno de los clientes que tenían implementados es un problema. Con Twisted existe un hilo principal que lleva a cabo todas las tareas; cuando se van a llevar a cabo comunicaciones por la red se emplea un hilo especial que es el que espera a recibir la respuesta o evento y avisa al hilo principal cuando se haya recibido la respuesta. Con esta solución se evitan problemas de gestión de recursos porque principalmente se ejecuta un único hilo, y el pool de hilos prácticamente se

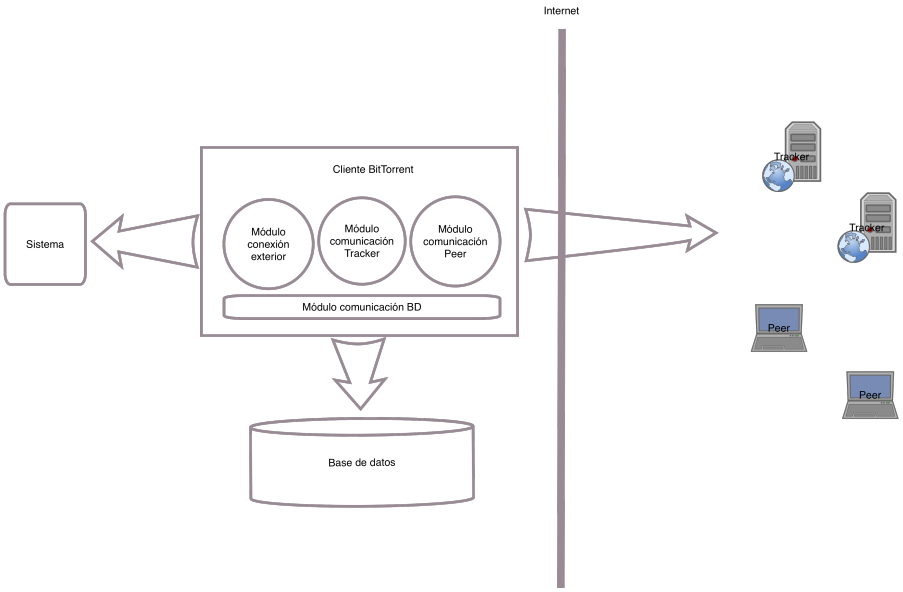

Figura 5.8: Diagrama del cliente P2P

encuentra a la espera sin consumir recursos. Además, ese framework conlleva la estabilidad de una librería especializada que lleva tiempo siendo mantenida por una amplia comunidad.

El cliente está compuesto por los siguientes módulos:

- En primer lugar se encuentra el proceso principal. Entre sus funcionalidades se encarga de inicializar y gestionar todos los "procesos" de comunicación con los Peers, los Trackers, y de crear un servidor para comunicaciones externas como es el caso de la plataforma a implementar.
- El módulo de comunicación de los Trackers: está compuesto a su vez por un módulo de comunicación TCP y otro UDP, dado que los Tracker en BitTorrent suelen soportan uno de los dos.
- El módulo de comunicación de los Peers: solo se ha implementado el protocolo básico de BitTorrrent por TCP que es el estándar. Existen otras implementaciones como es el conocido Micro Transport Protocol (uTP) a través de UDP pero no han sido implementados por el momento.
- El módulo de comunicación de la Base de Datos: se encarga de recibir las peticiones de lectura y escritura de los distintos módulos y llevarlas a cabo. Para ello, se ha adaptado el módulo de Flask-Alchemy para que lo realice de forma asíncrona a través de Twisted.
- Módulo de comunicación externo: para crear un canal de comunicación con el sistema, poder recibir órdenes como es la provisión del estado del cliente, informar de que se ha eliminado un archivo torrent a monitorizar, etc, se ha implementado a través de Twisted un puerto de escucha que permite recibir peticiones GET/POST a través de HTTP.
- Módulo de geolocalización IP: este módulo recibe direcciones IP para obtener el país al que pertenece. Para ello, se ha empleado la librería python-geoip y se ha descargado localmente una copia de la base de datos de geolocalización MaxMind. A través de ella se realizan

Tabla 5.1: Historias comprobadas

Nombre	**Funcionamiento**
CU-01 - Login	Correcto
CU-02 – Creación de cuenta	Correcto
CU-03 – Edición de usuarios	Correcto
CU-04 – Visualización de datos	Correcto
CU-05 – Exportación de resultados	Correcto
CU-06 – Visualización de estadísticas	Correcto
CU-07 – Registrar cliente	Incompleto
CU-08 – Añadir archivo	Correcto
CU-09 – Eliminar Archivo	Correcto
CU-11 – Información del sistema	Correcto

peticiones por cada dirección IP que no haya sido registrada previamente y se obtiene el país donde se encuentra dicha IP.

- Módulo de obtención de información del ISP: este módulo recibe direcciones IP, por cada una de ellas comprueba lo siguiente, si la dirección IP no se encuentra registrada en la Base de Datos del cliente P2P, se realiza una petición mediante la librería ipwhois a una Base de Datos remota y obtiene los datos del PSI. En caso de que este PSI no se encuentre en la BD, ésta lo almacena.

5.3. Aceptación del sistema

Para garantizar el correcto funcionamiento del sistema, se han planteado una serie de pruebas de distintos tipos.

5.3.1. Historias

De los casos de uso obtenidos, se realizará un chequeo de las historias comprobando que se cumplen las postcondiciones.

5.3.2. Pruebas realizadas

Antes de comprobar que las historias cumplieran sus objetivos, se han realizado numerosos test durante la codificación del proyecto para comprobar que cada funcionalidad interna estuviera correctamente implementada.

Por un lado, se han codificado pruebas unitarias para probar que individualmente en cada módulo las distintas funciones desarrolladas cumplieran con los propósitos esperados. Una vez implementados las pruebas unitarias se ha procedido a desarrollar las pruebas de integración para comprobar que los distintos módulos codificados encajaran correctamente.

Para comprobar el correcto funcionamiento del cliente P2P, se ha empleado la suite de pytest para probar el conjunto de clases codificadas. En total, en torno al 50 % del código se encuentra bajo tests unitarios, pero para el resto del código no se han podido codificar tests debido a lo

especializado de las librerías empleadas como ha sido el caso de las comunicaciones. Aun así, se han realizado numerosas pruebas reales para comprobar que no hubiese problemas y solventarlos en caso contrario.

En el caso del sistema, no se han codificado tests, pues la parte de las Vistas y los Controladores han sido puestas a prueba de forma manual y se han empleado librerías que se encontraban testeadas, mientras que la capa Modelo ha sido testeada previamente con el cliente P2P.

6

Estudio de viabilidad de un reconocimeinto automático de usuarios

En esta sección se describirá la investigación propuesta en la sección de Introducción con el objetivo de encontrar indicios que permitan identificar a un usuario o grupo de usuarios dentro de la red BitTorrent a lo largo de un periodo de tiempo.

Como se ha mencionado anteriormente, la identificación de usuarios de forma prospectiva supone el requisito sine qua non para poder señalar a un usuario como sospechoso de compartir archivos ilícitos de pornografía infantil en las redes P2P. Por ello se pretende determinar parámetros para clasificar automáticamente a usuarios que compartan archivos en la red BitTorrent.

6.1. Método

6.1.1. Procedimiento

A lo largo de un año, y tras haber estudiado el protocolo BitTorrent en profundidad, se ha construido un cliente de monitorización para la red BitTorrent y una plataforma Web para la gestión y visualización de los datos que genera la herramienta de monitorización.

Una vez finalizada la construcción de todo el sistema y para poder llevar a cabo la investigación de posibles indicadores que permitan identificar a un usuario o grupo de usuarios, se ha seleccionado una serie de archivos Torrent de forma aleatoria para ser monitorizados. Posteriormente se crearon dos archivos Torrent.

En primer lugar se ha procedido a monitorizar durante 2 semanas dos archivos Torrent, de entre los cuales se encontrara un cliente BitTorrent. La monitorización consiste en solicitar a usuarios que provean de un bloque aleatorio perteneciente a los archivos Torrent y verificar que el bloque solicitado realmente pertenece al archivo (comprobando el hash del bloque con el listado

de hashes del archivo Torrent). Durante el proceso de monitorización se activó un cliente control para así dispone de registros de un usuario conocido. Los parámetros que se registraron fueron:

- Identificador: valores numéricos, autoincremental de la base de datos.
- Fecha de creación: valores en formato fecha que contiene la fecha y hora en la que se ha monitorizado al usuario.
- Fecha de modificación: valores en formato fecha con la fecha de última modificación del registro en la base de datos.
- Peer_id: cadena de caracteres del identificador del cliente monitorizado.
- Torrent_id: valor numérico del identificador del Torrent dentro de la Base de datos.
- Torrent_client_id: valor numérico del identificador del cliente monitorizado en la base de datos.
- Personal_peer_id: cadena de caracteres del identificador del cliente P2P implementado.
- Extended_bytes: valor numérico enviado por el cliente monitorizado que corresponde a la extensión de protocolos que soporta el cliente monitorizado.
- Num_pieces: valor numérico de piezas del archivo monitorizado.
- IP: cadena de caracteres que representa la dirección IP del cliente monitorizado.
- Puerto: valor numérico del puerto empleado por el cliente monitorizado.
- ISP: identificador numérico del ISP obtenido mediante una consulta de la dirección IP empleada por el cliente monitorizado.
- Country: identificador en formato ISO-3166-1 (2 caracteres) que registra el país que está asociado a la dirección IP del cliente monitorizado.

De la monitorización llevada cabo durante 2 semanas se ha obtenido una muestra compuesta por 3614 registros. Por problemas de formato de algunos campos se ha tenido que reducir a 2543 registros. Esta base de datos cuenta con las variables anteriormente mencionados. No obstante, de los registros obtenidos no aparece en ningún momento el usuario control.

Debido a la ausencia de datos del usuario control, en la muestra obtenida de 3614 registros, se crearon dos archivos Torrents que solo tuviera el usuario control. Además, se monitorizó un tercer archivo aleatorio que tuviera muy pocos usuarios activos, abarcando el proceso de monitorización 4 días. El resultado obtenido es una muestra de 100 registros a la que no ha sido necesario llevar a cabo una limpieza de valores perdidos.

Posteriormente, la variable peer_id de las bases de datos obtenidas se ha transformado a su vez en 3 variables distintas. Éstas son:

- Identificador del cliente: cadena de caracteres compuesta por 1 y 3 caracteres.
- Versión del cliente: valor numérico que indica la versión del cliente empleado.
- Valor aleatorio: cadena de caracteres generada aleatoriamente.

Asimismo, se tiene registrado del usuario monitorizado cuál es la dirección IP empleada en cada momento.

6.1.2. Participantes

Se dispone de dos bases de datos que constan de:

- 100 registros extraídos de la monitorización de 3 archivos Torrent, dos creados específicamente para este proyecto y un tercero ya existente para incluir usuarios reales. De los 100 registros, 72 corresponden a un único usuario monitorizado a través de varios días que disponía de los 3 archivos Torrent.
- 3614 registros extraídos de la monitorización de 2 archivos Torrent ya existentes durante varios días.

6.2. Análisis

Las variables seleccionadas para su posterior análisis son las siguientes:

- Dirección IP
- Puerto
- Extended_bytes
- Identificador del cliente
- Versión del cliente
- Valor aleatorio
- ISP
- País

Todas las variables se considerarán como nominales, puesto que sus valores no se encuentran sometidos a un criterio de orden. Las variables numéricas "puerto" o "extended_bytes" no corresponden a una escala o medición, y por ello también se consideran nominales.

Los análisis llevados a cabo son:

- Extracción de características comunes de los registros del usuario monitorizado.
- Ejecución de distintos algoritmos de Machine Learning sobre la base de datos.

Todos los análisis se han llevado a cabo a través de la herramienta RapidMiner y Weka.

6.3. Resultados

Los resultados obtenidos son los siguientes:

6.3.1. Identificación de características comunes de los registros del usuario monitorizado

Tras estudiar la base de datos de 100 registros, se observa lo siguiente:

- Aparece el 100 % de las veces que se registra al cliente "control" el mismo:
  - Identificador del cliente
  - Versión del cliente
  - ISP
  - Puerto
  - La extensión de bytes

Por otro lado, se ha observado que la variable Valor Aleatorio tiene diferentes valores. De esta forma, por cada archivo Torrent monitorizado la variable tiene siempre un valor diferente, en cambio, si se monitoriza varias veces el mismo archivo, el cliente proporciona siempre el mismo valor. No obstante, estos valores se modifican cuando se cambia de dirección IP o se resetea el cliente P2P del usuario monitorizado.

6.3.2. Ejecución de distintos algoritmos de Machine Learning sobre las bases de datos

Se han aplicado distintos algoritmos de Machine Learning con varios tipos de dataset para intentar extraer características que permitan identificar al usuario monitorizado dentro de una muestra. Las muestras empleadas son las siguientes:

- Dataset-1: Con la base de datos compuesta por 100 registros al que se le ha añadido una variable binomial que indica cuál es el usuario controlado. Se indica con el valor 1 los registros obtenidos del usuario control y con valor 0 el resto de los registros.
- Dataset-2: Con la suma de las dos bases de datos se le ha añadido una variable binomial que indica cuál es el usuario controlado. Se indica con el valor 1 los registros obtenidos del usuario control y con valor 0 el resto de los registros.

Ejecución de algoritmos con todas las variables para Dataset-1 con cross-validation de 10 grupos:

Ejecución de algoritmos sin variable de ISP, país y dirección IP para Dataset-1 con cross-validation de 10 grupos:

Si se intenta identificar al usuario con respecto al resto de usuarios monitorizados en el Dataset-1, se puede llegar a obtener un 100 % de precisión con varios algoritmos. Esto se puede deber a la poca diversidad de datos, como se comentará más adelante.

Ejecución de algoritmos con todas las variables para Dataset-2 con cross-validation de 10 grupos:

Ejecución de algoritmos sin variable de ISP, país y dirección IP para Dataset-2 con cross-validation de 10 grupos:

Tabla 6.1: Resultados algoritmos sobre Dataset-1

Algoritmo	**J48**	**LibSVM**	**RandomForest**
Clasificados Correctamente	100	100	99
Clasificados Incorrectamente	0	0	1
Kappa	1	1	0.9749
Error medio absoluto	0	0	0.0142

Tabla 6.2: Resultados algoritmos sobre Dataset-1 con eliminación de variables

Algoritmo	**J48**	**LibSVM**	**RandomForest**
Clasificados Correctamente	98	100	99
Clasificados Incorrectamente	2	0	1
Kappa	0.9493	1	0.9749
Error medio absoluto	0.0144	0	0.017

Tabla 6.3: Resultados algoritmos sobre Dataset-2

Algoritmo	**J48**	**LibSVM**	**RandomForest**
Clasificados Correctamente	99.847	97.1623	99.773
Clasificados Incorrectamente	0.1513	2.8377	0.227
Kappa	0.9718	0	0.9572
Error medio absoluto	0.0035	0.0284	0.0051

Tabla 6.4: Resultados algoritmos sobre Dataset-2 con eliminación de variables

Algoritmo	**J48**	**LibSVM**	**RandomForest**
Clasificados Correctamente	97.1623	97.1623	99.773
Clasificados Incorrectamente	2.8377	2.8377	0.227
Kappa	0.9718	0	0.9572
Error medio absoluto	0.0284	0.0284	0.0051

Los resultados obtenidos permiten llegar a identificar cerca del 99 %. De nuevo, esto se puede deber a la poca diversidad de datos. Se ha realizado clustering del Dataset-1 que está compuesto por pocas direcciones IP para intentar extraer características que con los algoritmos de clasificación no hayan visto (pues en los otros casos solo se ha validado con la variable binomial que indica si es el usuario controlado o no). Los resultados obtenidos han sido que el 50 % de los registros se han agrupado correctamente con un Log likelihood de -5.11461.

6.4. Conclusiones

En base a los resultados obtenidos, los casos de prueba pudieron ser clasificados con un 100 % de precisión. Esta precisión, a todos exagerada, se debe con casi toda seguridad al pequeño tamaño y sesgos de la muestra. Por ejemplo, hay que tener en cuenta que con variables como son el ISP o el país se podía identificar al usuario monitorizado.

Aún así, del estudio llevado a cabo del protocolo y la red BitTorrent se han identificado puntos interesantes a la hora de crear un clasificador.

Primero, el puerto del Peer puede ser un indicativo muy fuerte, puesto que durante la preparación del cliente a monitorizar, se tuvo que abrir un puerto del router y asignar ese puerto a una IP fija dentro de la red local del router. Este puerto es elegido por el propio usuario y es necesario que se encuentre abierto para que pueda compartir archivos. Por tanto, ese puerto, que podría ser en un principio cualquiera, puede ser una forma de identificar al usuario aunque cambie de dirección IP. No obstante, para la red BitTorrent se suele emplear en muchas ocasiones el puerto 6881, en la muestra de 2543 registros se observa que hay una cantidad muy diversa de puertos empleados y que estos se repiten en las direcciones IP repetidas. Igualmente, en el caso del cliente monitorizado, aunque la dirección IP cambie el puerto se mantiene. También se ha observado que hay algunos clientes que van seleccionando los puertos de conexión a partir de un rango. En el caso de que uno de los puertos esté en uso, el cliente puede escoger el siguiente puerto abierto, es decir, si el puerto 6881 está ocupado, el cliente probará con el puerto 6882 y así sucesivamente hasta encontrar un puerto libre. Esta información es valiosa, pues puede haber casos en que los puertos no coincidan pero se encuentren dentro de un mismo rango.

Con respecto al atributo Extended_bytes, el protocolo BitTorrent soporta extensiones a su protocolo y la forma de comunicarlo es enviando a través de este campo una numeración especial que indica al otro cliente qué extensiones soporta. En el caso del usuario monitorizado la secuencia de números enviada es la 0000000000100005, que es la más habitual en los clientes más populares de BitTorrent. Aun así, en la muestra obtenida de 3614 datos se observa que hay algunos que solo envían 0s y otras variaciones; esta cadena puede ser un indicativo interesante para clientes que no envíen la cadena habitual y puede ser una forma de identificarlos a lo largo del tiempo.

El atributo de peer_id, que en el análisis de datos se separó en tres variables (cliente, versión del cliente y número aleatorio), aporta la información del cliente que está empleando y la versión de dicho cliente, siendo éste otro indicio que permite reducir la muestra e identificar a los usuarios o grupo de usuarios a través de este parámetros. El valor que identifica al cliente es fijo por cada aplicación. En cambio, el valor que puede cambiar es la versión del cliente, puesto que depende de la versión empleada por el usuario. De esta forma, si el usuario actualiza el cliente, este valor cambiará a su vez. Teniendo en cuenta esta información, se puede crear una regla que identifique los tipos de cliente y la versiones que emplean asumiendo que el tipo de versión puede cambiar. Con respecto al valor aleatorio existen diversos comportamiento, hay clientes que por

cada conexión con un cliente emplean un valor diferente, otros que por cada Torrent tienen un valor diferente pero se mantiene aunque la conexión con el cliente se cierre y no es modificado hasta que el usuario reinicie la aplicación y por último, hay algunos que mantienen el mismo valor para todos los Torrents. Conociendo el comportamiento de los clientes, se puede llegar a identificar a un usuario a lo largo de los días si, por ejemplo, no se desconecta. Esto es muy útil porque si se encuentran usuarios que a lo largo del tiempo mantienen este identificador se puede identificar sin ninguna duda que es el mismo usuario.

Con todos estos indicativos, se pueden crear unas reglas que filtren la muestra obtenida para reducir los usuarios y llegar a identificar si no es a un único usuario, al menos a un grupo de ellos. Es necesario indicar que el cliente de monitorización y la plataforma se centrarán en aquellos usuarios que tienen el perfil de compartidores de archivos, es decir aquellos que "permanentemente" están compartiendo un archivo. En el caso de los descargadores el modus operandi es el de localizar y descargar un archivo, y una vez descargan el archivo lo eliminan inmediatamente del cliente y no comparten con otros usuarios. A estos usuarios es más difícil seguirles la pista, por no decir que es inviable pues el tiempo de vida para monitorizarlos es muy reducido, además de ser necesario el tener un larga lista de archivos Torrent de material ilícito para conocer si vuelven a descargarse otro archivo.

7

Conclusiones

En esta sección se describirán las conclusiones resultantes en la elaboración de este proyecto, asimismo se detallarán los objetivos alcanzados, las limitaciones encontradas y las posibles futuras líneas de trabajo.

7.1. Objetivos alcanzados

Se puede considerar que se han alcanzado los objetivos planteados para este proyecto. De esta forma, se ha conseguido crear una plataforma web que servirá a la Guardia Civil en sus lucha contra el abuso infantil en las redes P2P, en concreto en la red BitTorrent. Como se ha comentado anteriormente, dicha plataforma debía responder a unas características propias marcadas por el problema que enfrentan las FFCCSE en España convirtiendo esta plataforma en una herramienta distinta a otras disponibles. La plataforma está preparada para ser una herramienta que permita a los investigadores de la GC identificar a usuarios que están compartiendo este tipo de material para llevar a cabo procesos judiciales. A su vez, se ha podido crear un cliente P2P que es capaz de monitorizar a usuarios que compartan determinados archivos. El cliente implementado está limitado a los protocolos estándar de BitTorrent, pero está diseñado para ser ampliado. Asimismo, se han extraído y analizado datos de usuarios compartiendo archivos lícitos para encontrar características que puedan identificar a un usuario a lo largo del tiempo. De esta forma, se han podido identificar que características de los clientes monitorizados pueden ayudar a dicha identificación.

7.2. Metodología empleada

En este proyecto se ha conseguido crear una plataforma y un cliente que estuvieran muy optimizados de acuerdo con las limitaciones de recursos del servidor donde vaya a ser instalado. Asimismo se ha tenido en cuenta que tanto el cliente como la plataforma Web no sea estática, sino que en el futuro vayan a ser ampliados. Estas características resultan imprescindibles para

que la herramienta sea útil para la Guardia Civil. Esto se ha conseguido con un estudio de las tecnologías actuales, como es el caso de un diseño asíncrono para el cliente BitTorrent. Además, se ha considerado relevante añadir un valor de seguridad mediante la codificación de tests de las funcionalidades del cliente para asegurar que los datos que se obtienen son realmente los que se reciben, evitando así que sean corrompidos durante el proceso de monitorización. Esto resulta necesario pues puede ocurrir que en un futuro estos datos sean empleados en un proceso judicial. Igualmente, se ha considerado relevante tratar de encontrar formas de viabilidad de reconocimiento automático de usuarios dado que la cantidad de datos que pueden generarse resulta abrumadora para los profesionales que tratan con estos. El uso de algoritmos de Machine Learning se ha considerado la mejor forma para abordar esta cuestión.

7.3. Limitaciones encontradas

Cabe mencionar algunas limitaciones encontradas en la realización de este trabajo. Debido a las restricciones legales referidas a la pornografía infantil, no se ha podido analizar datos reales de usuarios que comparten dicho material y poder adquirir una visión más realista de este problema en la red BitTorrent. Igualmente, se han encontrado limitaciones de tiempo y de recursos disponibles que podrían haber mejorado el resultado final del proyecto, aportando una visión más pormenorizada de la problemática en cuestión. Por otro lado, a la hora de monitorizar los archivos seleccionados, se ha detectado poca variedad de los registros obtenidos, lo cual ha obstaculizado el proceso de creación de un clasificador que permitiera identificar a un usuario en esta red. En concreto no se han obtenido datos de usuarios que emplearan direcciones IP de España, pues la herramienta estará limitada al uso en territorio español. Esto es debido a las limitaciones del protocolo estándar de BitTorrent, el cual no proporciona un listado completo de los usuarios que comparten los archivos.

7.4. Futuras líneas de trabajo

Desde el comienzo del planteamiento de este trabajo, se ha concebido este proyecto como un cometido a largo plazo. Por ello, los siguientes pasos de este proyecto son:

- Ampliar el cliente BitTorrent para que incorpore nuevos protocolos como son el Distributed Hash Table (DHT), que permite consultar a los peers para conocer más usuarios de la red que comparten el archivo en cuestión. Asimismo, añadir otros protocolos como uTP y PEX para el intercambio de archivos.
- Obtener más datos para poder crear unas reglas que permitan identificar a los usuarios de esta red, con estas reglas incorporarlo al sistema creado para que realice estas búsquedas automáticamente y así ayudar a los investigadores.
- Ampliar la integración de otros clientes P2P en el sistema, para que dispongan de un sistema unificado para trabajar. Además, se propone la investigación o ampliación del sistema para que haya retroalimentación de todos los clientes P2P y poder encontrar más usuarios.
- Investigar las nuevas vías de anonimato de BitTorrent mediante I2P y la red Tor.

Éstas son unas de las tantas posibilidades que permite este proyecto.

7.5. Conclusiones finales

Como se ha visto, la pornografía infantil y su distribución a través de redes P2P supone un grave problema social hoy en día. Tal y como ocurre en distintos ámbitos de la criminalidad, las nuevas tecnologías han facilitado dichas conductas, y en muchos casos han dado lugar a modalidades nuevas de delincuentes. Por ello resulta imperativo comprometerse con la ardua tarea que supone combatir esta peligrosa cuestión. En este sentido, la colaboración entre el mundo académico y las fuerzas de seguridad del estado se convierte en una obligada necesidad. La realización de este proyecto ha permitido corroborar que con una adecuada inversión de recursos, la investigación puede arrojar luz sobre retos técnicos que a priori parecen difíciles de afrontar. No obstante, también queda patente que aún falta mucho por investigar y que cada día, y a medida que aparecen nuevas tecnologías, se complica esta tarea.

8

Bibliografía

- European Commission (2010). *Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on combating the sexual abuse, sexual exploitation of children and child pornography, repealing Framework Decision 2004/68/JHA*. Brussels.

- Carr, J. (2004). *Child abuse, child pornography and the internet*. NCH. London

- Chothia, T., Cova, M., Novakovic, C. & González Toro, C. (2012). The Unbearable Lightness of Monitoring: Direct Monitoring in BitTorrentIn Keromytis, A.D. and Di Pietro, R. (Eds.) *Security and Privacy in Communication Networks. 8th International ICST Conference, SecureComm 2012. Padua, Italy. Revised Selected Papers*. (pp. 187-202). Springer: Ney York.

- Cisco (2014, June). Cisco Visual *Networking Index: Forecast and Methodology*, 2013–2018

- de la Cruz, I. P., Aller, C. F., Garcia, S. S., & Gallardo, J. C. (2010, June). A careful design for a tool to detect child pornography in P2P networks. *Technology and Society (ISTAS), 2010 IEEE International Symposium*. 227-233. IEEE.

- GAO (2003). Combating child pornography: Federal agencies coordinate law enforcement efforts, but an opportunity exists for further enhancement. Washington, DC: Government Accountability Office.

- INSAFE - INHOPE (2014). *Annual Report 2013*.

- Jefatura del Estado. (24 de noviembre de 1995). Artículo 189 del Código Penal. España. Jefatura del Estado. (06 de diciembre de 2015). Artículo 588 ter k de la Ley de Enjuiciamiento Criminal. España.

- Liberatore, M., Levine, B. N., & Shields, C. (2010, November). Strengthening forensic investigations of child pornography on p2p networks. In *Proceedings of the 6th International Conference* (p. 19). ACM.

- Liberatore, M., Erdely, R., Kerle, T., Levine, B. N., & Shields, C. (2010). Forensic investigation of peer-to-peer file sharing networks. *Digital investigation*, 7, S95-S103.
- Piatek, M., Kohno, T., & Krishnamurthy, A. (2008). Challenges and Directions for Monitoring P2P File Sharing Networks — or — Why My Printer Received a DMCA Takedown Notice. *Proceedings of the USENIX Workshop on Hot Topics in Security*. San Jose, CA, USA.
- Prichard, J., Watters, P. A., & Spiranovic, C. (2011). Internet subcultures and pathways to the use of child pornography. *Computer Law & Security Review*, 27(6), 585-600.
- Steel, C. (2009). Child pornography in peer-to-peer networks. *Child Abuse & Neglect*, 33(8), 560-568
- Wolak, J., Finkelhor D. & Mitchell, K.J. (2012). *Trends in Arrests for Child Pornography Possession: The Third National Juvenile Online Victimization Study (NJOV-3)*. Crimes against Children Research Center: Durham, NH.
- Wolak, J., Liberatore, M. & Levine, B.N. (2013). Measuring a year of child pornography trafficking by U.S. computers on a peer-to-peer network. *Child Abuse & Neglect*.