

A graphic with the text "Broadband Testing" in white, bold, sans-serif font. The background is a dark blue rectangle containing stylized, semi-transparent icons of a laptop, a server tower, and a network switch connected by lines, suggesting a network or data center environment.

Broadband Testing

Advestigo Audio and Video Verification Technology in the Context of P2P Monitoring Services

A Broadband-Testing Report

First published May 2008 (V1.0)
Second release September 2008 (V 2.0)

Published by Broadband-Testing
A division of Connexio-Informatica 2007, Andorra La Vella

Tel : +376 633010
E-mail : info@broadband-testing.co.uk
Internet : [HTTP://www.broadband-testing.co.uk](http://www.broadband-testing.co.uk)

©2008 Broadband-Testing

All rights reserved. No part of this publication may be reproduced, photocopied, stored on a retrieval system, or transmitted without the express written consent of the authors.

Please note that access to or use of this Report is conditioned on the following:

1. The information in this Report is subject to change by Broadband-Testing without notice.
2. The information in this Report, at publication date, is believed by Broadband-Testing to be accurate and reliable, but is not guaranteed. All use of and reliance on this Report are at your sole risk. Broadband-Testing is not liable or responsible for any damages, losses or expenses arising from any error or omission in this Report.
3. *NO WARRANTIES, EXPRESS OR IMPLIED ARE GIVEN BY Broadband-Testing. ALL IMPLIED WARRANTIES, INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT ARE DISCLAIMED AND EXCLUDED BY Broadband-Testing. IN NO EVENT SHALL Broadband-Testing BE LIABLE FOR ANY CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, OR FOR ANY LOSS OF PROFIT, REVENUE, DATA, COMPUTER PROGRAMS, OR OTHER ASSETS, EVEN IF ADVISED OF THE POSSIBILITY THEREOF.*
4. This Report does not constitute an endorsement, recommendation or guarantee of any of the products (hardware or software) tested or the hardware and software used in testing the products. The testing does not guarantee that there are no errors or defects in the products, or that the products will meet your expectations, requirements, needs or specifications, or that they will operate without interruption.
5. This Report does not imply any endorsement, sponsorship, affiliation or verification by or with any companies mentioned in this report.
6. All trademarks, service marks, and trade names used in this Report are the trademarks, service marks, and trade names of their respective owners, and no endorsement of, sponsorship of, affiliation with, or involvement in, any of the testing, this Report or Broadband-Testing is implied, nor should it be inferred.

TABLE OF CONTENTS

TABLE OF CONTENTSIII

BROADBAND-TESTING.....IV

EXECUTIVE SUMMARY 1

INTRODUCTION: THE P2P CONTENT PIRACY ISSUE..... 2

ADVESTIGO: A PRODUCT AND SERVICE OVERVIEW 4

THE P2P TEST 6

SUMMARY & CONCLUSIONS 9

TABLE OF FIGURES

Figure 1 – Advestigo Summary Scorecard By Number Of Videos.....6

Figure 2 – Advestigo Summary Scorecard By Relative Weighting7

Figure 3 – Advestigo Log Comparisons.....8

Figure 4 – Advestigo Detailed Scorecard.....8

Broadband-Testing

Broadband-Testing is Europe's foremost independent network testing facility and consultancy organisation for broadband and network infrastructure products.

Based in the south of France, Broadband-Testing offers extensive labs, demo and conference facilities. From this base, Broadband-Testing provides a range of specialist IT, networking and development services to vendors and end-user organisations throughout Europe, SEAP and the United States.

Broadband-Testing is an associate of the following:

NSS Labs (specialising in security product testing)

Limbo Creatives (bespoke software development)

Broadband-Testing Laboratories are available to vendors and end-users for fully independent testing of networking, communications and security hardware and software.

Broadband-Testing Laboratories operates an **Approval** scheme which enables products to be short-listed for purchase by end-users, based on their successful approval.

Output from the labs, including detailed research reports, articles and white papers on the latest network-related technologies, are made available free of charge on our web site at [HTTP://www.broadband-testing.co.uk](http://www.broadband-testing.co.uk)

The conference centre in Moux in the south of France is the ideal location for sales training, general seminars and product launches, and Broadband-Testing can also provide technical writing services for sales, marketing and technical documentation, as well as documentation and test-house facilities for product development.

Broadband-Testing Consultancy Services offers a range of network consultancy services including network design, strategy planning, Internet connectivity and product development assistance.



Executive summary

- With the Advestigo software we were able to quickly scan the Internet globally and identify potential copies of original feature films or musical tracks sourced from Peer to Peer (P2P) networks.
- Not only did the software discover these copies but it was also able to rate them very precisely, according to how accurate a copy of the film or the piece of music it was, based on video and/or audio content.
- As part of commercial services, the results are made available through a user-friendly secure web access giving both permanently refreshed data and easily sorted and manageable reports. Given the nature of media content and its related industry, the ability to appeal to "non IT savvy" individuals is of real value.
- In this specific 'Content Verification' testing, the results were clearly displayed in an Excel spreadsheet – the only user interfacing required in the whole process - meaning the software and service is suitable for any type of customer, IT aware or otherwise.
- The service and application is completely scalable. Being based on computational power, it is a simple case of matching the search requirements to the correct level of hardware. For this project we used a standard, Intel-based server platform that you could buy anywhere, so we were not talking "supercomputer" requirements here.
- The alternative to using the Advestigo software – some kind of manual search process – is not in any sense a realistic alternative. Within the context of our tests, we concluded that what might take a company months or years in determining copyright issues without Advestigo, is reduced to days or weeks with the product.
- Automated monitoring services, but not encompassing content verification capabilities, have become more popular over the last few years. However, the increasing volume of fake files in circulation makes it compulsory to provide a sound and automated 'Verified Monitoring Service' so that it can be a safe basis for litigation or mass deterrence policies.
- A key point about the Advestigo solution is that it takes what, beneath, is a complex series of computations and algorithms, and turns it into a simple reporting process from a user perspective.
- The bottom line here is that we have a solution to a problem that is potentially costing media and content companies tens of millions of pounds and beyond, so from an ROI (Return On Investment) perspective, the Advestigo approach makes complete sense.

introduction: The P2P Content piracy issueⁱ

For every valid use of IT there is equally an illegal use.

Nowhere is this more evident than in the world of Peer 2 Peer (P2P) networking, especially where illegal copying of music and video or film content is concerned. For content owners, infringement copyright is a major issue and one that is directly linked with the P2P user community. The question is, how practical is it to implement filtering similar to keyword or IP blocklists in current P2P software, or are there alternative solutions to protecting or identifying copyright infringement?

Where It All Started

Napster was seen as the starting point for many. However, its architecture - integrated with search performed centrally at the Napster server and then downloading on a decentralised basis directly between two users - made it easy to monitor.

Peer-to-peer networks have now evolved from this original model to various generations of technology. After the centralised Napster-like tools, there was a second generation of P2P protocols largely decentralised, like Gnutella (in its first version). The philosophy behind this decentralisation was to avoid incurring a court decision condemning a central server block, so bringing down the whole network. But decentralised networks are less efficient, due the erratic path taken by the messages between the peers. A third generation of P2P has appeared, correcting this point with a hierarchy of servers. The top-most server contains no infringing information per se, while the lowest tier is so replicated that it is impossible for a court decision to be made against them.

So this third generation is far more efficient and robust than its predecessor. Several evolutions appeared, the most representative P2P protocols of this period being FastTrack (Kazaa) and EDonkey. Now, some P2P software is able to bridge these several protocols together (one of the first was MLDonkey, where P2P nodes play both the role of both the server and the client). Many of the node software variants are able to manage a protocol like EMule, EDonkey, AMule, EMule+ and MLDonkey), for the EDONKEY2000 protocol.

It was therefore natural to have P2P nodes that are able to manage several P2P protocols, facilitating the diffusion of infringing content across typical Internet user borders, given that the majority tend to run one specific software variant. They have their particular user habits using one specific protocol, while a minority use software that bridges several protocols. The files coming from one protocol are then allowed to be available in another, increasing the diffusion effect.

Moving to the fourth generation of protocols, these are mainly oriented towards discretion, being anonymous to both the user and the cryptography of the contents. Another new tendency appears even stronger, this being the P2P streaming of live events.

A significant loss of revenue from online content piracy is damaging the motion picture, music, video game and information industries; P2P networks contribute heavily to this loss. Meantime, the steady increase in bandwidth and the ease with which media can be digitised for convenient distribution is only making the situation worse for the content developers and owners.

With in excess of a hundred proprietary and open source software clients in use in the P2P world, copyright infringement is now a very complex matter to trace. Some kind of manual search is simply not an option; meaning that some kind of automated content fingerprinting methodology is the only realistic solution.

The key to fingerprinting on P2P networks is in the use of hash codes. When users perform actions such as running searches, attempting downloads or sharing files, P2P networks exchange a small amount of information. This data includes file name and hash code. While files names can, of course, be the same for different files, the hash code is used as a unique file identifier, ensuring that the file requested is the same one received, whether from a single user, or thousands of users. But, while this is a common methodology, most P2P networks use different hashing algorithms. While small – 100 characters or less - hash codes have three critical characteristics that can be identified for file infringement identifications.

So, at least there are real means of identifying copyright infringement. The question is, just how do you monitor the WWW for such infringements? Advestigo believes it has the answer, so we will now investigate what the company has to offer and just how successful a solution it provides.

Advestigo: A Product and Service overview

Advestigo is in business to provide solutions to the problems of copyright protection and creating new content distribution models.

Currently Advestigo offers a range of services such as online surveillance of unlawful content distribution, statistical analysis of piracy for specific sectors like the music or motion picture industry, or even the installation of end-to-end 'graduated response' solutions. The solutions are designed to protect the rights of authors, producers, editors and distributors of protected content, companies that need to protect or measure brand image and identity, or agencies that research and track unlawful online content distribution.

Advestigo's view is that the digital content industry can learn much from statistics that measure content popularity, the voluntary or involuntary diffusion of logos or design elements, the capacity and geography of Internet radio audiences, and the unauthorised reproduction of information available online. Here we are primarily concerned with the latter. The company has therefore developed a range of services to measure, monitor and combat online piracy (known as AdvestiSEARCH™).

In addition to monitoring the flow of illicit content, Advestigo can also pinpoint the machines that are illegally distributing or exchanging files, so an absolute identification can be made.

Contrary to the first generation of P2P Monitoring Services, Advestigo puts special emphasis to automatically verifying the nature of the content being traded, regardless of the name and attributes attached to the file. This allows delivering 'Verified Monitoring Services' that can serve as a robust and sound basis for further litigation or deterrence policies.

This combination of services and applications enable Advestigo to perform multiple operations, including:

- **O**btaining detailed statistics concerning piracy activity on P2P networks.
- **M**onitoring the unlawful transfer of protected digital content over P2P networks.
- **A**utomatically detecting unauthorised copying or reprocessing of protected content. This can lead to direct action or, by using documented evidence to build a case, to action in a court of law.
- **C**reating detailed statistics of the popularity of music works over peer-to-peer networks.

Client-specific results from Monitoring Services are made available through a secure website, updated in real-time and comprising a number of data filtering and sorting tools.

From a marketing statistics perspective, Advestigo can also help. Content owners can track general piracy activity on P2P networks and assess its development. Advestigo is able to inform about the top files downloaded by protocol, rights owner, and ISP.

Information about total download volumes, nature of P2P protocols used and P2P clients is also available. More music is exchanged over P2P networks than anywhere else on the Internet but, despite more than 800 million copyrighted music tracks being exchanged per month, there are currently no significant marketing statistics for P2P networks. Advestigo can make these statistics available through a secure website.

Advestigo Digital Fingerprint Analysis

The Advestigo 'Verified Monitoring Services' are based on proprietary, patented technology that analyses digital fingerprints, automatically identifying files which are partial or complete copies of protected content.

It has been developed to handle text, images, video and audio files. Importantly, it takes into account file variations caused by reformatting, compression, Hertzian diffusion, reframing, dilution, adding information, and other common factors. It can use existing markers such as watermarking, electronic signatures and encapsulation to identify files, but these are not essential for digital fingerprint analysis.

The software has been designed to accurately identify modified copies of original contents, whatever the type of modification, such as change of format, resizing, cropping or blurring. Importantly, the identification process is non-invasive, so the fingerprinting does not require any modification of the original content.

By automating these processes, Advestigo has developed fully automated services that it claims are capable of efficiently processing large, continuous streams of files exchanged over online networks, even when files are poor reproductions of an original. Here we put those claims to the test.

Dissemination of media contents can occur at several stages of the media product life, such as before official release, during theatrical release, or during DVD release life, all of which can impact on the profitability of that media. Hence the importance in providing a solution for the detection and monitoring of the illicit diffusion of music tracks, movies or TV series on peer-to-peer networks.

Advestigo claims it can help determine at exactly what stage of the hacking process the supervised digital content is, what are fakes (misleading files with the same filename as the original content but which content is very different) and what are real, and collate these results into easily analysed statistics that faithfully reflect the real situation of copyright infringements.

The P2P Test

For the P2P testing, we ran two series of tests, the first in May 2008 entirely focused on illicit copies of movies and the second performed in late August 2008 dedicated to benchmarking the technology on musical tracks.

Movies

As the motion picture section of the test goes, we created seven queries, relating to seven specific movie titles that we wanted to identify. These were all well-known releases, covering a number of genres from horror, through action and family, to animated.

From this list we then searched across the P2P networks for suspect titles, resulting in the detection of over 400 movies in total being offered over P2P networks. Important to stress that these 400 variants represented more than 99.9% of the total number of uploads corresponding to these movies during the test period, hence being a fully comprehensive picture of the variety of copies available. We then downloaded a total of more than 300, the remainder proving being unavailable (see below). We then generated statistical analysis of these films. This process involved transcoding the films to Advestigo’s format, which then meant we could carry out fingerprint computation of these suspect files.

The Advestigo software then carries out a one-to-one comparison of each suspect file with the original matching file, meaning over 300 direct comparisons in this particular test. The comparison creates a series of data files which can then be uploaded into a custom Excel spreadsheet developed by Advestigo in order to make the final analysis as easy as possible from what is known as the “Table of Truth”!

The spreadsheet has both summary modes and detailed examination of the comparison process, complete with hash codes.

Score Benchmark Peer To Peer
 This Table Is Calculated Automatically

By Number Of Videos

Search	Number Of Suspects		Unavailable For Download		Not Processed	True Positives		False Positives		False Negatives
	NA	NR	NP	TP		TN	FP	FN		
Film 1	28	1	2	1	3	21	0	0		
Film 2	38	0	0	0	7	31	0	0		
Film 3	41	1	0	0	10	30	0	0		
Film 4	31	0	6	1	0	24	0	0		
Film 5	61	14	3	0	18	26	0	0		
Film 6	25	2	0	3	18	2	0	0		
Film 7	181	86	0	2	91	1	0	1		
Total	405	104	11	7	147	135	0	1		

Figure 1 – Advestigo Summary Scorecard By Number Of Videos

The first summary table breaks down the seven films we investigated by number of suspects identified, then by a number of different rules. Non-downloaded (non-telechargé) files are where the software attempted to download these over an eight day period and

failed, perhaps because of too few sources, or being fake files, for example. Sometimes these are pushed out by protection agencies as decoy files – typically a hash code with a large number of sources. After eight days, Advestigo rules that this now too long a period to be a viable option for a standard internet user so the download attempt is stopped.

Non-readable (non lisible) files are those where it is not possible to see the content of a file after downloading, with standard tools. Currently Advestigo uses four different readers, including Windows Media Player and Apple QuickTime.

Non-processed (non traité) files are those unable to be converted by the system, perhaps because it could not be converted to specific, valid formats – for example, where the frame rate is not the same in the container versus the content; a metadata issue in other words. This can be an issue where software for hacking video streams is used.

True positive and negatives (absolute copies and/or absolutely not) are identified along with false positives and negatives. The complete lack of false positives (quite critical in terms of not improperly assessing an illegal behaviour) and almost lack of false negatives (only one in total) is very reassuring in terms of the ability of the Advestigo software to accurately appraise the suspect content.

By Relative Weighting		Total Weighting	Unavailable For Download	Unreadable	Not Processed	True Positives	True Negatives	False Positives	False Negatives	OK
Search		NA	NR	NP	TP	TN	FP	FN		
Film 1	5963	0.2%	4.7%	1.3%	27.1%	66.8%	0.0%	0.0%		98.7%
Film 2	19290	0.0%	0.0%	0.0%	15.2%	84.8%	0.0%	0.0%		100.0%
Film 3	7780	0.1%	0.0%	0.0%	30.1%	69.8%	0.0%	0.0%		100.0%
Film 4	15309	0.0%	22.2%	0.9%	0.0%	76.9%	0.0%	0.0%		99.1%
Film 5	13368	0.3%	5.6%	0.0%	18.1%	76.0%	0.0%	0.0%		100.0%
Film 6	3147	5.9%	0.0%	0.6%	91.2%	2.4%	0.0%	0.0%		99.4%
Film 7	1386724	2.9%	0.0%	0.8%	94.9%	0.1%	0.0%	1.3%		97.9%

Figure 2 – Advestigo Summary Scorecard By Relative Weighting

The suspect films are also reported upon in a similar way by relative weighting – a good way to quickly see which content is being copied extensively in relation to other content. Equally, the suspect files can be studied in more detail, by hash code and at what hour they were sourced and compared.

Advestigo P2P Analysis

Date	Time	Operation	Suspect	Candidate	Audio	Video
03/20/08	13:54:03	compare	045D768255944026945B1E29ADD089A5.avi	P2PMASTER_fast_tokyo_.avi	0.83	1.00
03/20/08	14:01:53	compare	8FB931EFC728396B62194C3FA574ED0A.avi	P2PMASTER_fast_tokyo_.avi	0.00	0.00
03/20/08	14:10:55	compare	A9CF32950F0F01D5E639C5C0D2781334.avi	P2PMASTER_fast_tokyo_.avi	0.00	0.00
03/20/08	14:13:17	compare	7B9F403468CD821C38885E7777153C1C.avi	P2PMASTER_ne_dis_personne.avi	0.00	0.00
03/20/08	14:22:18	compare	788046F960F64C9DD9928F7FF4BE67E9.avi	P2PMASTER_ne_dis_personne.avi	0.00	0.00
03/20/08	14:33:41	compare	A6D980477B99D60BEC36E8CB8B1325A5.avi	P2PMASTER_ne_dis_personne.avi	0.00	0.00
03/20/08	14:41:35	compare	398EEB1E7A78E016389C2C4902C2089B.avi	P2PMASTER_ne_dis_personne.avi	0.00	0.00
03/20/08	14:50:03	compare	85B8767C027A9A2F56BFA88CA5EAA7BB.avi	P2PMASTER_ne_dis_personne.avi	1.00	1.00
03/20/08	14:53:15	compare	901E8F9EECEBE9880452B8221CD4B0E3.avi	P2PMASTER_taxi4_.avi	0.00	0.00
03/20/08	15:01:06	compare	07870CBA878033B2FB6D08AC7BE9313B.avi	P2PMASTER_taxi4_.avi	0.00	0.00
03/20/08	15:14:50	compare	E46C17EAEBF4D245D2BFD79B7519CC02.avi	P2PMASTER_taxi4_.avi	0.00	0.97
03/20/08	15:23:41	compare	02AD20F5E596BD8A9F6638DFD7A18CA5.avi	P2PMASTER_taxi4_.avi	0.00	0.00
03/20/08	15:26:47	compare	13FEAB50964C5DBF6BCB914F76AC7D27.avi	P2PMASTER_taxi4_.avi	0.00	0.00
03/20/08	15:28:44	compare	C90CA09AF593E3426F6D422D5DE4D3D1.avi	P2PMASTER_age_glance_2_.avi	0.00	0.00
03/20/08	15:38:01	compare	2DBABF68E524957D52A9B9657EF78EC7.avi	P2PMASTER_age_glance_2_.avi	0.00	0.00
03/20/08	15:40:45	compare	012F3E2589083A92FBD9B8A395638BDF.avi	P2PMASTER_age_glance_2_.avi	0.00	0.00
03/20/08	15:48:44	compare	95A01ECCBA603B80FC5775CAEDDFEBC2.avi	P2PMASTER_hannibal_lecter.avi	0.00	0.00
03/20/08	15:51:06	compare	7EDACEDC9FD60652753110A38A652309.avi	P2PMASTER_hannibal_lecter.avi	0.00	0.00
03/20/08	16:04:06	compare	9AABE96E64F0DA0A16A6731593DD1BB0.avi	P2PMASTER_hannibal_lecter.avi	0.93	1.00
03/20/08	16:06:49	compare	B16B97C297F2EFF173934372CDD8ADCC.avi	P2PMASTER_hannibal_lecter.avi	0.29	1.00
03/20/08	16:11:14	compare	3EC5A81E4DEBC7B961C28C2E86476639.avi	P2PMASTER_hannibal_lecter.avi	0.00	0.00
03/20/08	16:24:21	compare	24DA6CC82D2F2948150936A20DCAB1C7.avi	P2PMASTER_hannibal_lecter.avi	0.94	1.00
03/20/08	16:33:46	compare	7BF62DB2A5B715EFD727699B7B665E502.avi	P2PMASTER_casino_royale_.avi	0.25	0.94
03/20/08	16:46:22	compare	4A1E1248043FE27D141F9590B65720F0.avi	P2PMASTER_casino_royale_.avi	0.89	1.00
03/20/08	17:04:42	compare	902FEA1EDBCECB425F8FD7134A63FEA5.avi	P2PMASTER_national_treas_.avi	0.00	1.00
03/20/08	17:27:14	compare	5091A63A3946160BD38CAD87C8DF3F15.avi	P2PMASTER_national_treas_.avi	0.00	1.00
03/20/08	17:38:39	compare	7B7234238CA549A28260248307C35BA9.avi	P2PMASTER_national_treas_.avi	0.29	1.00
20/03/2008	17:38:39.39	compare	P2PMASTER_national_treas_.avi	D075AE7C8F9D2DA29CDA87743C0BD2BC.avi	SUSP_NOT_FOUND	
03/20/08	17:57:52	compare	22AE86537D3481B6CE1AE167C166657E.avi	P2PMASTER_national_treas_.avi	0.60	1.00

Figure 3 – Advestigo Log Comparisons

So why does the relative quality of a suspect copy vary? The reason why the Advestigo software has to extensively compare suspect files against the original is because those “copies” can vary enormously in terms of accuracy of video and audio content copying.

For example, copies made directly from a DVD release are far more likely to be accurate, high quality copies than those hacked from cinema releases. And then there are deliberate fake files to add to the computation. Advestigo has to pull its suspect files from many different sources and copy types, hence the scoring system it has introduced to quickly identify the level of copy accuracy of a suspect film.

Reference	Suspect	Hits	Hits %	Downloaded	Readable	Transcoded	Thumbnail OK	Positive	Expected	Comparison Completed	Audio Scoring	Video Scoring	Label
fast_tokyo	045D768255944026945B1E29ADD089A5.avi	1093	18.3%	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	0.83	1.00	TP	
fast_tokyo	E0100D727CD616817B07C239346AC2D6.avi	498	8.4%	TRUE	TRUE	FALSE	TRUE	TRUE	0.91	1.00	TP		
fast_tokyo	4D03EC20C3912064883087181FFA209E.avi	475	8.0%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	4D269785169D709C462147D44A415DF9.avi	472	7.9%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	D817C5E6C1D59B6F65E170059DC34096.avi	432	7.2%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	D795B14F09D4953D144FAA07988C3FDE.avi	363	6.1%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	71ED4D3C5145B94DBCE80B0B9FAED59.avi	319	5.3%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	F64DA1DB4393946CBAEDB976D4C17948.avi	269	4.5%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	8FB931EFC728396B62194C3FA574ED0A.avi	245	4.1%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	BD6C4ACD1CF0711048CC2DBF3D473BEB.avi	218	3.7%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	E8EC2543BBC021AB4FAF4AFB1E289E9D.avi	212	3.6%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	3568E609DD44F9A0FB6C3BA2E4764EE3.avi	195	3.3%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	E1AE1E140F494E9C85B668E94CEFD2D5.avi	193	3.2%	TRUE	FALSE			FALSE			NR		
fast_tokyo	4839F83E4250B5DA22ADD0A405625726.avi	168	2.8%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	5277E89D062F6A71A4B2C1922AE3E729.avi	129	2.2%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	8F0A3B5B4AC15AC9B887DA48A0BA1B48.avi	128	2.1%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	2C24177693A18A2FAD0590119D18CC0.avi	86	1.4%	TRUE	FALSE			FALSE			NR		
fast_tokyo	CA8CBF39E218141ED4EE51595C3327CD.avi	79	1.3%	TRUE	TRUE	FALSE		FALSE			NP		
fast_tokyo	3602D31A3DFCEDDEDB63CCF7F8CF1AE.avi	62	1.0%	TRUE	TRUE	FALSE		FALSE	0.00	0.00	TN		
fast_tokyo	A9CF32950F0F01D5E639C5C0D2781334.avi	54	0.9%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	C7309C63B0641C17AB2B360E4C66E537.avi	48	0.8%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		
fast_tokyo	3D0546C7297A8C95E0A289DA90A98FC0.avi	45	0.8%	TRUE	TRUE	FALSE	FALSE	TRUE	0.00	0.00	TN		

Figure 4 – Advestigo Detailed Scorecard

Scores are therefore an absolute rating of the quality of recognition. Anything above “0” means there is some recognition – the exact percentage related to the amount of degradation, so the closer to “1” the better quality the copy, “1” being a very good

match, but not necessarily an exact 100% replica. For example, it may be an exact replica of part of the movie, or near-exact versions in part or completion.

As part of the computation, the Advestigo software generates large amounts of data in log and xml files, analysing movies chain by chain and frame by frame. In this way it is possible to very accurately identify true positives and negatives, taking into account all kinds of variables such as time distortion and cuts.

There are many elements to consider which are not immediately obvious. For example, within the P2P community, there is a big movement in taking standard content and modifying it to create what are described as new art forms. These may involve radical changes in the content, but they still need to be identified. So the complete evolution of the content can be monitored and identified.

Musical tracks

In a similar fashion we have selected 33 tracks belonging to different musical genres but all highly popular with illegal P2P users.

Score Benchmark Peer To Peer

This table is calculated automatically

By number of audios

Searched	Artist	Number of suspects		Not tested	True positives		True negatives		False positives		False negatives	
		NA	NR		TP	TN	FP	FN				
Bonus Track	black eyed peas	3	0	0	0	3	0	0	0	0	0	
Disco Club	black eyed peas	3	0	0	2	0	0	0	0	0	0	
Red Light	u2	2	0	0	0	0	0	0	0	0	0	
Ba Bump	black eyed peas	1	0	0	0	1	0	0	0	0	0	
They Don't Want Music	black eyed peas	3	0	0	1	0	0	0	0	0	0	
Gone Going	black eyed peas	8	0	0	4	0	0	0	0	0	0	
Stay The Night	carey mariah	4	0	0	2	1	0	0	0	0	0	
Mine Again	carey mariah	6	0	0	4	0	0	0	0	0	0	
Shining from heaven	Bob Sinclar	1	0	0	1	0	0	0	0	0	0	
Say Something	carey mariah	9	0	0	1	4	0	0	0	0	0	
Union	black eyed peas	4	0	0	2	2	0	0	0	0	0	
Can't Take That Away	carey mariah	4	0	0	3	0	0	0	0	0	0	
Get Your Number	carey mariah	13	0	0	10	0	0	0	0	0	0	
Be Here To Love Me	jones norah	8	0	0	7	0	0	0	0	0	0	
Don't Miss You At All	jones norah	6	0	0	5	0	0	0	0	0	0	
In The Morning	jones norah	9	0	0	6	0	0	0	0	0	0	
The First Time	u2	6	0	0	6	0	0	0	0	0	0	
All I Want Is You	u2	9	0	0	7	2	0	0	0	0	0	
Pump It	black eyed peas	9	0	0	6	0	0	0	0	0	0	
Like That	black eyed peas	8	0	0	3	3	0	0	0	0	0	
If You Wear That Velvet Dress	u2	2	0	0	2	0	0	0	0	0	0	
Surrender ?	u2	10	0	0	2	8	0	0	0	0	0	
Justify My Love	madonna	8	0	0	3	3	0	0	0	0	1	
Exit	u2	6	0	0	6	0	0	0	0	0	0	
Live To Tell	madonna	10	0	0	9	1	0	0	0	0	0	
The Fly	u2	10	0	0	9	1	0	0	0	0	0	
Two Hearts Beat As One	u2	6	0	0	4	1	0	0	0	0	1	
Do You Feel Loved ?	u2	4	0	0	4	0	0	0	0	0	0	
Red Hill Mining Town	u2	3	0	0	3	0	0	0	0	0	0	
Amora, amor	Bob Sinclar	2	0	0	2	0	0	0	0	0	0	
Heartbreaker	carey mariah	0	0	0	0	0	0	0	0	0	0	
Last Night A Dj Saved My Life	carey mariah	0	0	0	0	0	0	0	0	0	0	
Beautiful Stranger	madonna	11	0	0	11	0	0	0	0	0	0	
Total		188	0	31	0	125	30	0	0	0	2	

The suspects downloaded here represent between 95 and 100% of the hits measured on the P2P. Quite typically, one can notice that the average number of distinct files per

original to reach this threshold is significantly less than for movies, i.e. the variability of the copies in circulation is lesser.

If we look closer to what percentage of the total volume of downloads the 'missed' suspects represent, they come to less of a %.

By relative weighing

Searched	Artist	Total number of hits	Not downloaded		Not tested	True positives		False negatives		False positives	False negatives	OK
			NA	NR		TP	TN	FP	FN			
Bonus Track	black eyed peas	323	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Disco Club	black eyed peas	254	0,0%	98,8%	0,0%	1,2%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Red Light	u2	2	0,0%	100,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Ba Bump	black eyed peas	1	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
They Don't Want Music	black eyed peas	647	0,0%	99,4%	0,0%	0,2%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Gone Going	black eyed peas	1114	0,0%	82,2%	0,0%	17,1%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Stay The Night	carey mariah	728	0,0%	97,0%	0,0%	0,4%	1,9%	0,0%	0,0%	0,0%	0,0%	100,0%
Mine Again	carey mariah	767	0,0%	98,4%	0,0%	0,7%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Shining from heaven	Bob Sinclar	98	0,0%	0,0%	0,0%	99,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Say Something	carey mariah	1706	0,0%	62,1%	0,0%	4,0%	32,8%	0,0%	0,0%	0,0%	0,0%	100,0%
Union	black eyed peas	472	0,0%	0,0%	0,0%	2,8%	96,2%	0,0%	0,0%	0,0%	0,0%	100,0%
Can't Take That Away	carey mariah	582	0,0%	94,2%	0,0%	4,3%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Get Your Number	carey mariah	879	0,0%	3,2%	0,0%	95,2%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Be Here To Love Me	jones norah	310	0,0%	51,6%	0,0%	46,8%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Don't Miss You At All	jones norah	527	0,0%	93,9%	0,0%	4,4%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
In The Morning	jones norah	777	0,0%	91,6%	0,0%	6,4%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
The First Time	u2	296	0,0%	0,0%	0,0%	97,6%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
All I Want Is You	u2	848	0,0%	0,0%	0,0%	91,5%	5,8%	0,0%	0,0%	0,0%	0,0%	100,0%
Pump It	black eyed peas	497	0,0%	55,5%	0,0%	41,6%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Like That	black eyed peas	698	0,0%	44,7%	0,0%	49,0%	3,3%	0,0%	0,0%	0,0%	0,0%	100,0%
If You Wear That Velvet Dress	u2	65	0,0%	0,0%	0,0%	96,9%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Surrender ?	u2	608	0,0%	0,0%	0,0%	10,0%	86,3%	0,0%	0,0%	0,0%	0,0%	100,0%
Justify My Love	madonna	1042	0,0%	52,8%	0,0%	37,2%	5,2%	0,0%	1,0%	99,0%	0,0%	100,0%
Exit	u2	77	0,0%	0,0%	0,0%	96,1%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Live To Tell	madonna	446	0,0%	0,0%	0,0%	95,1%	0,7%	0,0%	0,0%	0,0%	0,0%	100,0%
The Fly	u2	231	0,0%	0,0%	0,0%	80,5%	15,2%	0,0%	0,0%	0,0%	0,0%	100,0%
Two Hearts Beat As One	u2	157	0,0%	0,0%	0,0%	93,6%	1,3%	0,0%	0,6%	99,4%	0,0%	100,0%
Do You Feel Loved ?	u2	67	0,0%	0,0%	0,0%	95,5%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Red Hill Mining Town	u2	134	0,0%	0,0%	0,0%	95,5%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Amora, amor	Bob Sinclar	107	0,0%	0,0%	0,0%	95,3%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Heartbreaker	carey mariah	594	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Last Night A Dj Saved My Life	carey mariah	239	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Beautiful Stranger	madonna	1010	0,0%	0,0%	0,0%	95,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
Total		16303										

This means that, while analyzing more than 95% of the available P2P traffic generated by as many as 33 different originals, more than 99% of the illicit exchanges of files would be properly identified with ZERO FALSE POSITIVES.

Summary & Conclusions

What Advestigo is achieving here, in identifying very accurate levels of quality of copies of movies or musical tracks distributed in the P2P environment is far from trivial.

The underlying complexities and scale of task become quickly clear on using the software and service. With the Advestigo solution we were able to quickly scan the Internet globally and identify potential copies of musical billboard hits or original feature films sourced from Peer to Peer (P2P) servers. Not only did the software discover these copies but it was also able to rate them very precisely, according to how accurate a copy of the film it was, based on video and audio content.

Importantly, the service and application is completely scalable. Being based on computational power, it is a simple case of matching the search requirements to the correct level of hardware. The alternative to using the Advestigo software – some kind of manual search process – is not in any sense a realistic alternative. Within the context of our tests, we concluded that what might take a company months or years in determining copyright issues without Advestigo, is reduced to days or weeks with the product.

So, here we have a solution to a problem that is potentially costing media and content companies tens of millions of pounds and beyond, so from an ROI perspective, the Advestigo approach makes complete sense.

