Computer Laboratory
University of Cambridge
CAMBRIDGE
CB3 0FD
2nd October 2012

Mike O'Connor
Chief Executive
Consumer Focus
Fleetbank House
Salisbury Square
London
EC4Y 8JX

Dear Mike,

I am writing to give you my reflections on the meeting we both attended on the 19th September where Ms Marianne Grant, Senior Vice President of the Motion Picture Association of America (MPAA), gave us a presentation regarding an automated monitoring system that is being used by a company that contracts their services to MPAA members to detect unauthorised file sharing of copyright material.

The system the MPAA described to us detects the 'uploaders' of material, those people who have material on their machines which they transmit to others, the 'downloaders'. The automated system plays the part of a downloader and gathers evidence about the uploaders it interacts with. The operation of the system is very similar to the approach I outlined in the expert report I produced for Consumer Focus, which you published on 26 July with the title: "Online traceability: Who did that?"

The initial step is that there is a search for material that could be infringing – Ms Grant described this in terms of using keywords in searches, where the keywords would typically be some or all of the title of a movie. When a new item is found, a complete copy of the movie is downloaded and a determination is made as to whether or not it is unlawfully shared copyright material. If it is an infringement then all the details of the item are recorded and the cryptographic hashes of the content are calculated. If the same item is encountered again it will be possible to accurately determine that it is an infringement by performing a partial download and comparing the hashes of the data that has been fetched.

The system described to us avoids the trap of using advertised hash values from the file transfer protocol – they calculate the hashes themselves from the material which is actually fetched. This is the right way to do it. However, I have a very slight concern in that the system is still using SHA-1 as the hash function. This hash function has been shown to be slightly weaker than it should be, and it is generally considered to be prudent to avoid continuing to use it. That said, none of the weaknesses identified so far is realistically capable of being exploited to subvert the way in which the hash function is being used in this system – and so this is currently a theoretical rather than a practical concern.

On the plus side, the system described to us keeps a complete record (as a PCAP file) of the entirety of the file sharing traffic that has taken place with each particular uploader – and this would provide valuable forensic evidence in the event of a dispute about many aspects of the

system's operation. I did not recommend this record keeping in my expert report to you because of the cost and complexity – but clearly the operators of the system the MPAA described consider the trade-off to be worthwhile.

Although we were told that identification of copyright material owned by MPAA members (e.g. films) involved a manual process, Mr Kiaron Whitehead, General Counsel of the British Recorded Music Industry (BPI), who also attended the meeting told us that for music it was common to use automated identification systems – doubtless based on the type of signal processing technology that is used in products such as those marketed by Audible Magic. Unfortunately, recent events have shown that fully automated systems can make patently incorrect decisions, and you might have read of the blocking of streamed video of the Mars lander, the Hugo awards and part of the Democratic National Convention. Therefore, I would be concerned to learn that automated systems were not supplemented by manual checks.

Although there was a lot of valuable detail in what we saw, the presentation did not provide any real details of the 'hygiene' arrangements that the system employs. By that I mean, for example, exactly how the system ensured that the timestamps it reported were accurate and whether there were regular checks on automated parts of the system to ensure that they were still functioning correctly. In my expert report I listed a number of such issues – such as keeping logs of software versions, measures for keeping systems secure and so on. If the MPAA were, at a minimum, to adopt these recommendations and have their contractor publish appropriate details about the day-to-day operation of their system then this would go some way to improving public confidence in the reliability of the system.

You will recall that in in my "Online Traceability" report I specifically advised that Ofcom should not view 'secret' designs as being capable of providing reliable results. It is essential that the designs of monitoring systems can be independently reviewed and that the public should have the opportunity to understand how they work and why they are capable of precisely identifying the IP address of an unauthorised uploader. A handful of details, such as the monitoring system's IP address and the criteria for deciding upon keywords, must necessarily remain commercially confidential – but almost every other aspect of the system should, in my view, be published.

Should a dispute arise as to whether the system had correctly identified an IP address as having been involved in the unauthorised file sharing of copyright material then it would be extremely helpful for full details of the theory of operation of the system (along the lines of the presentation made to us) to be immediately available along with copies of the 'hygiene' arrangement logs that showed that there were no known problems at the relevant times.

Providing a version of the PCAP file that was recorded (doubtless with the exact IP address of the monitoring system redacted to avoid compromising future monitoring) would also be of invaluable assistance to anyone who wished to resolve the dispute by carefully examining the evidence. In my view, the MPAA would be well advised to require their members to put arrangements in place for all of these types of disclosure, and Consumer Focus should be strongly encouraging this.

There were two actions that the system took which seemed to me to be rather pointless and potentially confusing. The first was that it would 'ping' the uploader's machine to determine if it was alive. Since many people will configure their systems not to respond to pings it seems likely that this 'liveness' test will make any difference as to whether or not their file

sharing activity is monitored – and the subsequent file sharing traffic provides considerably more accurate evidence of liveness.

The second action was that a 'traceroute' was performed to the uploader's machine. I assume that this is an attempt to show that inter-provider routing is working in the way that might be expected. However, there are considerable limitations as to what a traceroute can demonstrate and as I discuss in #3.5.3 of my PhD thesis[1] it is possible to 'extend' the traceroute in an entirely misleading manner to apparently demonstrate that a different ISP is hosting the destination host. It would be far more appropriate, in the most unlikely event of a dispute about the bulk ownership of IP address blocks by ISPs, to consult public records of IP address allocations and routing announcements (such as those archived by RIPE).

Once the system has identified unauthorised file sharing of copyright material then it generates a report using an XML-based protocol described at http://acns.net and ships it to the appropriate ISP. It is then for the ISP to correctly identify the account holder to whom they had allocated the reported IP address at the relevant time.

As I explained at length in my expert report, if the ISP finds that their records show that a reported IP address was not allocated to any account then some part of the detection system (or the ISP allocation record keeping) is not working correctly. That malfunction may well be systemic and if so then all the other records in the same batch will also be wrong. Depending on the nature of the error, this may lead to incorrect identification of ISP customer accounts. Therefore, when errors occur, an investigation will need to be mounted to determine the source of the error and correct it. In the meantime, all reports from the same batch must be considered to be unreliable and not acted upon. In my report I called this a "Doctrine of Perfection".

Unfortunately, the ACNS protocol does not contain any mechanisms to give the sort of feedback required to report that an error has become apparent nor is there any mechanism to tell ISPs that previous reports are now, at least pending the conclusion of an investigation, to be considered unreliable. In other words – no provision has been made within this protocol to ensure that if errors occur then all potential misidentifications can be identified in a prompt and automated manner. I consider this a significant failing.

Of course I accept that the system design is intended to be foolproof, and that if my comments above about hygiene are taken on board then faulty components will be rapidly identified and fixed – nevertheless, it is in my view extremely unwise to assume that a system such as this will be operated without any errors ever occurring. Hence the design and implementation of the system should include the mechanisms for providing feedback and for dealing with errors – no matter how rare they may be – and at present that it not the case.

I hope that my comments will be of assistance to you. You need not consider this letter to be in any way confidential – and you may freely share it with whomever you wish. Should you do so, it is doubtless helpful that I should disclose that Consumer Focus reimbursed my return train fare to attend the meeting, for which many thanks.

It will also be helpful to mention that when Consumer Focus intervened in the High Court case "Golden Eye (International) Ltd & Anor v Telefonica UK Ltd [2012] EWHC 723 (Ch)

[1] http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-653.html

(26 March 2012)" I was paid by Consumer Focus to produce an expert witness report. Subsequently, I was also paid to create a report on online traceability which Consumer Focus submitted to the 2012 Ofcom consultation on the "Initial Obligations Code" for the Digital Economy Act 2010. This latter report was published on the Consumer Focus website, as I mentioned above.[2] These reports contain details of my qualifications and expertise.

Dr Richard Clayton

[2] http://www.consumerfocus.org.uk/publications/online-traceability-who-did-that-technical-expert-report-on-collecting-robust-evidence-of-copyright-infringement-through-peer-to-peer-filesharing