UMG Recordings, et al. v. Roy, USDC-NH
Civil Action No. 1:08-cv-00090-JL
Expert Witness Report

Sergey L. Bratus, Ph.D.

May 30, 2009

1 Summary

I have reviewed *MediaSentry* materials[1] that purport to link the defendant with a computer or computers participating in a Gnutella-based file sharing network from which MediaSentry allegedly downloaded copyrighted songs.

In my opinion, these materials leave critical aspects of MediaSentry's evidence collection process undocumented. In my opinion, they express unwarranted assumptions regarding both software and network technologies involved, and attempt to create an illusion of evidence-supported certainty where it does not exist.

I have also reviewed the *Plaintiffs' responses to defendant's first set of interrogatories*. I do not agree with statements and representations made by the Plaintiffs in this document.

I have also reviewed the following reports:

1. *Declaration and Expert Report* by Dr. Doug Jacobson from January 29, 2009.

2. *Expert witness report* by Dr. J.A. Pouwelse in UMG Recording Inc., et al. v.Lindor[2]

[1] `RoyMNH0054-- RoyMNH0996`, as provided to the defendant's counsel

[2] Available from http://www.ilrweb.com/viewILRPDF.asp?filename=umg_lindor_080215ExpertWitnessReportPouwelse

3. *Declaration* of Jason E. Street in Arista Records, LLC, et al. v.Does 1–11[3]

4. *Expert witness report* by Dr. Yongdae Kim in Capitol v. Thomas[4]

5. *Deposition of Expert Witness* Dr. Douglas Jacobson in the UMG Recording Inc., et al. v.Lindor case held on February, 2007[5]

I disagree with the representation and opinions expressed in Dr. Jacobson's report. In my opinion, Dr. Jacobson's report contains many factually erroneous statements, oversimplifications, and misleading statements, as well as assumptions made without any supporting evidence.

I agree with Dr. Pouwelse, Dr. Kim, and Mr. Street's strong criticism of Dr. Jacobson's statements and opinions and share their doubts of MediaSentry's evidence collection and handling.

In my opinion, there are many problems with linking the defendant with computer or computers that MediaSentry allegedly accessed to download files, and there is even less ground to claim that the Defendant "engaged in distribution" of these files.

The problems include but are not limited to the following.

2 Lack of documentation and review of MediaSentry methods and procedures

MediaSentry materials purport to be logs of downloads performed by MediaSentry through the Bearshare peer-to-peer network.

The materials are computer data, obviously generated by custom software, running on a MediaSentry-controlled computer(s), and its accuracy is subject to any flaws, or "bugs" in this software and any misconfigurations of the computer(s).

[3]Available from http://www.ilrweb.com/viewILRPDF.asp?filename=arista_does1-11_070806DeclarationJaysonStreet

[4]Available from http://beckermanlegal.com/pdf/?file=/Lawyer_Copyright_Internet_Law/virgin_thomas_090303DeftsExpertWitnessReport.pdf

[5]Available from http://beckermanlegal.com/pdf/?file=/Lawyer_Copyright_Internet_Law/umg_lindor_070308JacobsonDepositionMinUScript.pdf

2.1 Lack of software validation.

Validation of software used for forensic purposes is highly desirable. Peer review has uncovered serious flaws in forensic software produced by even the most reputable computer security experts (even years after their use by professionals). Courts have ordered that the source code of software used to derive evidence be analyzed in order to ensure its reliability and trustworthiness.[6]

To the best of my knowledge, and according to the statements by the experts quoted above **MediaSentry's software or configuration procedures have never been validated by independent computer science experts**.

2.2 Lack of computer configuration validation; accuracy of timestamps.

Correct configuration of a networked computer on which software is running is critical to the accuracy of forensic data produced by the software.

For example, for data that includes **timestamps** it is important to ensure that the computer's clock is set to the correct time. Commodity computer hardware does not in fact provide guarantees that the computer clock will be set correctly, or, once set, will maintain correct time and not "drift". Clock drifts (accumulated deviations from the correct "wall clock" time at the location) can grow quite large on misconfigured systems.

Clock time synchronization in computers across networks is an important practical and research problem and should not be taken for granted. Dedicated network protocols such as the Network Time Protocol (NTP)[7] are used to synchronize computer system time with special *time servers* trusted to have the accurate time (maintained, e.g., by the US NIST). Network security professionals stress the importance of correct network time synchronization.[8]

The problem of time synchronization is far from trivial. An MIT's Media Lab 1999 survey of NTP network time servers concluded that "only 28% of the Internet based stratum 1 clocks actually appears to be useful", and over a third had deviations of over 10 seconds, and some deviated by hours,

[6] http://www.dwi.com/new-jersey/state-v-chun

[7] http://www.ntp.org

[8] http://www.linuxdevcenter.com/pub/a/linux/2003/01/02/ntp.html

days, and even years.[9] Even though time network keeping practices have improved over the years, the issue still attracts attention of researchers and practitioners.[10]

MediaSentry's materials contain timestamps, and the Plaintiffs and Dr. Jacobson tacitly assume that these computer clock timestamps correspond to correct "wall clock" time. However, these materials contain no evidence of any time-synchronization procedures or configuration, such as NTP service configuration files and system logs with messages generated by NTP time synchronization software. Nor do they specify the reference source of computer time used for the timestamps.

Without such information, in my opinion, any time-related statements derived from software should be deemed unreliable. Because the **only** link to defendant is based on an inquiry to Comcast of what subscriber was using an IP address at a very specific time, timestamp validation should be essential to any of the Plaintiffs' assertions. Moreover, failure to describe the time synchronization measures suggests, in my opinion, lack of awareness of this crucial technical problem.

2.3 Invalid traceroute log?

In my opinion, the so-called "Tracing route" log (`RoyMNH0977`) should and does raise serious doubts with regard to the validity of the MediaSentry software or network configuration. **Since all of Plaintiffs' claims are based on the assumption that MediaSentry's software and computer configuration are trustworthy and free of errors, and this log clearly represents a failure of the MediaSentry software to perform the operation it claims to describe, the reliability and validity of the MediaSentry method should be questioned.**

Judging by the contents of this log, which starts with "Tracing route to " (IP address from the Complaint), the MediaSentry software has attempted to perform a standard operation known as "traceroute", in which a computer system attempts to discover which intermediary network hosts (computers, routers, or other networking devices) are likely to carry packets to the specified IP address. This operation is very common, and most commodity Oper-

[9] Nelson Minar, *A Survey of the NTP Network*, http://www.eecis.udel.edu/~mills/database/reports/ntp-survey99-minar.pdf

[10] E.g., http://www.ntpsurvey.arauc.br/globecom-ntp-paper.pdf was the fifth such survey since the original one quoted above.

ating Systems, such as Windows NT and XP, Mac OS X and most flavors of Linux include tools for performing it. Users are often instructed to perform it when troubleshooting network connectivity problems.[11]

The result of this operation is normally a numbered list of IP addresses and/or host names that will forward the packets on its path to its destination, and the times that packets spend in transit to and from these hosts. The sequence of the IP addresses is the *route* that packets take through the Internet. This route is *traced* by sending specially crafted packets to which each host on the current path responds in turn, and by interpreting and timing the responses.

The "traceroute" procedure is typically used to confirm that packets indeed travel to the intended destination via an expected path, and that the path is not "broken", suspicious, hijacked, or tampered with.[12] **In particular, packet path tampering or hijacking is the most effective way for perpetrators of illegal activities on the Internet to shift the blame for these activities to an innocent user**. Thus it is essential to establish that it was not happening.

It is apparent from the log that the operation has failed for the MediaSentry software, as the log shows neither the addresses nor names of the intermediary hosts nor realistic timings of packet round-trips between them and the MediaSentry computer. **The fact that this standard operation has failed suggests flaws, or "bugs", in either the MediaSentry software, or in its system or network configurations, or both.**

Thus – even assuming that the tracerouting was performed at the same time that the rest of the data was generated – the MediaSentry materials present no evidence as to how packets travelled between MediaSentry and the remote computer it was interacting with, in particular *whether any path deviations, tampering, or hijacking on this travel path might have taken place.*

This observation is extremely important in the light of Dr. Kim's explanations of spoofing and hijacking attacks on the Internet, in Sections 2.5, 2.6, and 3.2 of his expert witness report. Given the lack of data regarding the packets' path, any combination of such attacks might have occurred undetected by and unbeknownst to MediaSentry.

Dr. Jacobson states in item 16) of his report that he had re-

[11] E.g., http://www.exit109.com/~jeremy/news/providers/traceroute.html

[12] Dr. Kim in Section 2.5 of his report addresses well-known ways of hijacking or tampering with packet paths.

viewed the "MediaSentry Trace", but makes no mention of this conspicuous tracing problem in his report.

2.4 MediaSentry logs show insufficient understanding of forensic challenges.

MediaSentry materials are clearly generated by custom software. It is reasonable to assume that the choice of words and the overall format of log messages was made by designers or developers of the software and reflects their understanding of forensic challenges involved in proper handling of computer data used as evidence.

In my opinion, some of these choices are questionable when applied to data intended to serve as evidence. For example,

1. presentation of data should be unambiguous, and
2. once captured, the data should be immediately "checksummed" (mathematically fingerprinted) so that any later accidental damage or tampering while the data is stored can be detected.[13]

Ambiguous presentation of evidence. However, in the MediaSentry log `RoyMNH0964--RoyMNH0972` containing representations of TCP/IP network packets, which consist of bytes, interpreted according to the IP and TCP protocol specifications, many *different* bytes are presented as *the same printed character* (black circle or white space). This presentation creates an ambiguity, which precludes an expert from being able to check whether the purported source and destination printed by the software before each packet representation is actually correct. If the packet logging format was meant to be interpreted by experts in printed form, why was the inevitably obfuscated "one character for byte" format chosen?

Highly desirable evidence handling step omitted. Furthermore, in my opinion, MediaSentry software misses an essential step in forensic clarity of its logs.

Namely, the analysis log in `RoyMNH0973--RoyMNH0976` produced by this software details, for each allegedly downloaded file, the "File Name", "First

[13] Recommended as a "rule of thumb" by many forensic training courses, including those taught by SANS Institute (`http://www.sans.org/`) and many others.

Packet Received", "Last Packet Received", "First Download Packet Received", "Last Download Packet Received", and "Bytes Completed". Once the download is complete, the software claims to perform "Copying" and "Logging" of the file, and proceeds to the next file.

However, none of this logged information can reliably identify the **actual contents** of the file. In fact, many files disseminated through p2p networks do not contain what their name suggests. The length of a file does not help to identify its contents either, because some, most or all of the file's contents can be replaced without affecting its length.

The only reliable way to measure the contents of a file and to be able to subsequently tell that it had not been damaged or tampered with is fingerprinting the file. For this reason, taking mathematical fingerprints of acquired data is a standard step in computer forensics to help establish the chain of custody. Given the malleability of computer data, without this step it would be very hard to argue that the data initially captured by a forensic procedure is the same data that was later examined by another expert (and had not been changed, accidentally replaced, or tampered with in the meantime).

However, this MediaSentry software's log does not report fingerprinting files after they are downloaded. A fingerprint computed and recorded at the point when the file has been acquired would provide a way to verify at a later time whether the file has been copied correctly, and not damaged, or tampered with, by repeating the fingerprinting computation and making sure that the earlier and later fingerprints are identical.[14]

However, the log suggests that this rule-of-thumb computer forensics step was not performed.[15]

2.5 Misnomers in logs.

The logs refer to IP address and port combinations such as "75.68.28.28:6346" as *user* ("Initiating analysis of user ...", "Log for user at address ...").

[14] Provided that one trusts the software to correctly implement the mathematical fingerprinting algorithm in question.

[15] A separate log `RoyMNH0978--RoyMNH0988` contains only filenames and SHA1 fingerprints, but it is not clear whether and when these fingerprints were computed by the software, or simply taken from the Gnutella packets in which they are available. Other SHA1 fingerprints appear in the logs, but it is again not clear when, how, and by whom they were computed.

However, as Dr. Kim, Dr. Pouwelse, and Mr. Street, as well as many other authorities point out, an IP address does not identify a user (see also explanations in Section 4). The addition of a port number to an IP address may or may not help to identify a program that will be handling the data on its way to being processed (not even necessarily on the same computer that has the IP) but does not help to identify the user at this or any other computer.[16]

In my opinion, use of this language in these MediaSentry logs, produced specifically to serve as evidence, underscores the lack of attention to the accuracy required for forensic use of this data.

According to the statements of the Plaintiffs and of Dr. Jacobson, MediaSentry provided Comcast with the IP address and the timestamp (see above for the discussion of its accuracy), and requested Comcast to identify a Comcast customer solely by this information. I now turn to the analysis of the problems with this identification.

3 Lack of documentation and review of Comcast user identification methods and procedures

Linking an IP address to a customer is not, in fact, a trivial process for an ISP, as evidenced by occasional inability of ISPs to make such identification. For example, in Plaintiffs' exhibit `RoyMNH0014`, the Comcast's July 23, 2007 responses to the Plaintiffs' subpoena, for the IP address ending in `.196.13`, Comcast states that it is unable to identify the customer.

Additionally, the same response appears to contain another fully redacted line (the 5th line in the table of IP addresses), which may represent another identification failure.

Thus, out of 20 (or 21) IP addresses, Comcast has failed to identify one (or two), which suggests a high failure rate of 10-20% for Comcast's customer identification procedure.[17]

[16] Except when the computer system is specifically configured by its administrators to aid with such identification and runs appropriate identification service software.

[17] In order to evaluate this failure rate, statistics experts would need a much larger sample.

Dr. Pouwelse and Dr. Kim both point out that **failure rates for customer identification procedures by ISPs are unknown and evidence of identification errors exists**. Furthermore, in item C) of his expert witness report Dr. Pouwelse refers to examples of outright customer misidentification by ISPs.

Some technical problems that may prevent accurate customer identification are common for all ISPs that allocate dynamic addresses to customers using the DHCP protocol; other problems differ between DSL and cable providers, and are in fact harder for cable providers to solve.

3.1 DHCP timestamp accuracy.

As described above, the accuracy of the ISP records of assigning dynamic IP addresses to customer systems depends on the accuracy of the computer clock of the ISP's system(s) running DHCP server software.

A "drifting" computer clock on a DHCP server system will lead to inaccurate IP assignment logs and therefore to customer misidentification, while not interfering noticeably with the customers' connections. Dr. Pouwelse mentions "simple clock skew of a DHCP server" as a possible reason for customer misidentification.

The Comcast subpoena response provides no information on their time synchronization procedures.

3.2 Cable modem spoofing, customer identity hijacking.

Comcast's responses provide no details on how it identifies its customers and what, at the time of their response to the subpoena, constituted the "proof" of customer identify.

The question of what constitutes proof of an identity (and therefore what needs to be "stolen" or replicated to successfully impersonate an individual to a computer system or an organization) is a non-trivial one. Sometimes it is enough for a malicious impersonator to know an answer to a simple question to succeed.

The standard way of teaching about the perils of authenticating an identity (i.e., how an identity can be hijacked) is the famous "Something you

have, something you know, something you are" principle.[18] By way of example, many banks use "something only you could know" such as a Social Security Number or a mother's maiden name to authenticate customers calling in; yet, this knowledge can be obtained from various public sources, and is therefore widely abused by identity thieves. We identify law enforcement officers by special badges, "something they have" (and, we assume, no one else is likely to have). Forms of biometric identification such as fingerprint or iris scans, are examples of "something (only) you are".[19]

It is a question, then, what constitutes de-facto proof of a customer's identity to Comcast, and what was used by Comcast to point at the defendant. Internet sources quoted below suggest that what is used by Comcast is merely the MAC address of the customer's cable modem, which, as I will show, can be easily spoofed.

In many well-known network attacks, knowing or controlling a MAC address or an IP address of an ISP's customer was enough to successfully pass the ISP's authentication measures and hijack the customer's identity.

In my opinion, Comcast's procedures need to be studied further before it could be concluded that the Defendant's identity could not have been trivially hijacked, especially while the Defendant's computer(s) and cable modem were disconnected.

Let us consider the following theory, based on the assumption that Comcast identifies customers by their cable modem's MAC address, as an FAQ http://www.dslreports.com/faq/13104 at *DSLReports.com* a popular ISP ratings and discussion site, suggests:

> Comcast Cable Internet uses the MAC number of the cable modem to identify the user to the system. This means that no password or login is required. To tell Comcast that a particular modem is for your account, the Comcast system must be told about your modem, so that the modem can be served by the cable system. The cable system will download a configuration file to your modem based on the class of service you are subscribed to. See Cable Modem Provisioning FAQ."

[18] Explained, e.g., at http://www.cs.cornell.edu/Courses/cs513/2005fa/NNLauthPeople.html

[19] Experts warn that biometric identification is easier to fake than usually believed. See, e.g., the opinion of the world-recognized expert Bruce Schneier, http://www.schneier.com/essay-019.html, or respected technological news outlet ZDNet's report http://www.zdnetasia.com/techguide/security/0,39044901,39376855,00.htm

Assuming this, a hijacker in possession of a modified cable modem that is capable of having its MAC address changed, need only know a legitimate subscriber modem's MAC address to impersonate him or her. When any of victim's network-connected equipment is in use, this method would likely fail and cause network instability for both the victim and the hijacker. However, when the victim's equipment is not connected, the impersonation will succeed, because the individual in possession of a cable modem with the right MAC would appear to Comcast as the legitimate customer.

This scenario is not purely theoretical. There exist cable modems that allow reprogramming or replacement of their firmware to modify their MAC addresses and other parameters. For legitimate purposes, one can purchase cable modems with custom firmware already installed[20] or reprogram a modem using a development kit. The potential security risks of such modifications have been discussed publicly by the security practitioner community since at least 2004 (e.g, *Cable modem hackers conquer the co-ax*, by Kevin Poulsen, http://www.securityfocus.com/news/7977).

Also, from FAQ at http://www.tcniso.net/Nav/Tutorials/Questions/:

> **Is it possible to change the MAC address of a cable modem?** Yes, the MAC address of a cable modem is usually written on the Flash memory used to store the modem's firmware. This data can be often be [sic] changed in many ways and varies by cable modem model. The methods include using a RS-232 V2 board (effective on Surfboard models SB3100, SB4100, and SB4200) to boot shell enabled firmware that allow you to execute the 'factdef ' command. You can also use a E-JTAG cable such as BlackcatUSB to manually reprogram the Flash data which is effective on Surfboard models SB5100 and SB5101. Additionally, cable modems that have been modified with SIGMA enhanced firmware allow you to change the MAC address from the modem's diagnostic HTTP menu.

Experts note that unlike DSL providers, which can distinguish between the subscribing customers by their individual phone lines (cf. Dr. Kim's report, Section 2.5), cable providers have many customers share the same physical "circuit":

[20] http://www.tcniso.net/shop/product.php?productid=5

> ...The topography of cable modem networks typically puts between 500 and 1,000 homes in a neighborhood on the same circuit, their Internet traffic all mingled on the same co-ax cable...

Assuming the above theory, a hijacker in the same neighborhood would be indistinguishable to the ISP, unless additional security measures were implemented.

Finally, there are many ways in which a malicious spyware program infecting a victim's PC can find out a MAC address of a cable modem connected to the victim's computer. For example, malware can easily obtain it from the cable modem itself, by loading its own "diagnostic page". This method is described, e.g., in http://www.tcniso.net/Nav/Tutorials/Questions/:

> The Terayon TJ-715 and TJ-715x has a secret page located at: http://192.168.100.1/diagnostics_page.html; the password is: icu4at!

Note that the data considered "secret" by the manufacturer (possibly in order to protect customer identity from hijacking) is in fact easily obtained via a Google search by someone who knows the technological basics involved.

3.3 Misleading representation of "registration".

In item 22) of his expert witness report and elsewhere Dr. Jacobson uses the term **registered** to describe the relationship between the ISP and a customer's *computer*. In my opinion, this choice of term (not typically used in DHCP protocol descriptions, which rather talk about "leasing" an IP) is misleading, because it suggests technical certainty and deliberacy by an individual, neither of which exist.

According to the above DSLReports.com FAQ, a Comcast customer must call in to register the MAC address of his or her **cable modem**, which is printed on the modem's label. The customer is not asked for and need not report the MAC address of his or her **computer** (more precisely, of the computer's Network Interface Card (NIC), a component found inside the computer and enabling its network connectivity).

Even though the MAC address of a customer *computer's NIC*, or a connected network appliance's NIC, such as of a popular Linksys, Netgear or D-Link router, may actually be used by the modem to obtain an IP address from the ISP as a part of the dynamic IP address requests, the customer does

not need to take any affirmative steps to "register" his or her **computer(s)** or router MAC addresses.

Plaintiffs' exhibits I reviewed do not contain **any** of the defendant's computer(s) MAC addresses. In my opinion, this suggests that Comcast does not use customer computers' MAC addresses for the so-called "registration" as Dr. Jacobson's report suggests.

4 Plaintiffs do not distinguish between the IP address, computer(s), program(s), and the Defendant

Although the Complaint, the Interrogatory responses, and Dr. Jacobson's report suggest that accurate identification of the individual ("user") responsible for running a file sharing program (and thus "distributing" or "offering" files for download) has been made by an IP address with which a MediaSentry system allegedly communicated, no such determination can in fact be made with currently existing technology, for a number of reasons.

4.1 Impossibility to positively identify a program on a remote computer.

It is technically impossible to determine the actual identity of a program receiving and sending packets on a remote commodity PC or Mac computer without the use of special hardware technology such a a Trusted Platform Module, together with appropriate support by the remote computer's BIOS, device drivers, and operating system. The problem of determining the identity of programs running on remote computers in a trustworthy manner is known in the Trusted Computing community as *remote attestation*, and is still considered to be a hard engineering problem. No existing commodity systems so far have managed to provide satisfactory solutions to it.

For example, the remote computer might appear to be running a particular version of a webserver, a remote access server such as the Secure Shell (SSH), or a particular version of file sharing software. In fact, the program actually running on the computer may

- lie about its particular version (also a defensive tactic recommended by

some security experts to confuse potential attackers and have them try exploits that do not actually match the target);

- transparently forward all or some received packets to a different remote computer, and relay its responses back;

- selectively multiplex packets and connections between several different remote computers and programs while creating the appearance of all of them residing on the computer in question.

In particular, proxying a well-described protocol such as Gnutella (the basis of Bearshare and Limewire) should be within reach of a moderately proficient programmer. Such a proxy can be easily packaged with malware. Some of the well-known steps to reveal the existence of such a proxy would have been examining statistics such as those of packet paths and round-trip times (as explained in Section 2.3), the IP header TTL field, and, possibly, various OS and application fingerprinting methods. None are mentioned in MediaSentry materials.

In my opinion, it is impossible, on the basis of such data as provided by MediaSentry, to conclude with certainty — as Dr. Jacobson does in item 19) of his report — that a remote computer has been running a Gnutella-based file sharing program such as a Bearshare or Limewire client program.

Dr. Pouwelse describes a six-step procedure that would produce solid evidence that a remote computer is actually running file sharing software to make files available for download (with knowledge and intent of the individual in control of the computer). Quoting from his report:

> Due to the complexity of file sharing applications, limited observation powers, rampant deception, high pollution levels, and multi-peer downloading it is nearly impossible to obtain solid evidence and detailed checks are therefore required.
>
> I believe that the following 6-step test takes the necessary precautions when trying to establish if a computer is making copyrighted works available for download:
>
> 1. Collect filenames by searching the network using keywords.
> 2. Filter out polluted files by checking the actual content.

3. Establish that a specific file can be downloaded from a certain computer. File sharing applications often talk to numerous other computers at once. Sufficient hygiene precautions should be taken by blocking traffic from all possible other computers.
4. Investigate if the computer is possibly hijacked or the Internet connection is shared with others. Check if a computer is cracked, for instance, running an open proxy or a hacked Microsoft Internet connection sharing application. A measurement is needed to establish if there is no significant difference in traceroute timings, SYN responses, and KaZaA[21] protocol rendezvous times.
5. Track this computer for several days if it does not go offline for reliable IP address translation by an ISP.
6. Establish that no IP address spoofing, BGP hijacking, or other tampering with IP addresses has taken place.

I support Dr. Pouwelse's assessment and conclusions, and note that the MediaSentry materials suggest that most of these precautionary steps were **ignored** by MediaSentry, and that **the only attempted precautionary step (tracing route) has failed**.

In my opinion, it is highly likely that a malicious hijacker would take steps to disguise his or her actual file sharing software, therefore alternative explanations involving malice and deception **must** be considered, and additional tests performed, before any assumptions about software running on the other end of an Internet connection can be relied on. To quote Dr. Kim:

> ...[if] peer-to-peer networks are nothing more than havens for illegal activity, we would expect anyone committing copyright infringement to want to mask their trail when using these systems. ... Malicious users have great incentives to attempt to either hide or displace blame for their actions onto a third party.

Dr. Pouwelse concurs:

> ...When we can only observe this computer through the Internet we are severely limited in our observational power. ... the Internet

[21] This applies to any file sharing protocol, such as Gnutella. – Bratus

> and P2P are dark places where people commit fraud and abuse. All obtained information must be treated with suspicion.

In my opinion, **hijacking someone else's identity and thus shifting the blame for illegal transactions to someone else is the most efficient way to be anonymous on the Internet.** Various kinds of cyber-identity theft have become a lucrative (criminal) business on the Internet.[22]

In my opinion, the Plaintiffs' and Dr. Jacobson's analysis and representation do not exhibit appropriate precautions and fail to consider alternative explanations for their data, which, in my opinion, is rather incomplete and unreliable.

4.2 Remote computers and software misidentified by companies looking for copyright violations online

To further illustrate the points made above, in a paper titled "Challenges and Directions for Monitoring P2P File Sharing Networks - or - Why My Printer Received a DMCA Takedown Notice" (dmca.cs.washington.edu/dmca_hotsec08.pdf), a group of researchers from the University of Washington conducted a series of highly successful experiments in framing innocent arbitrary network-connected computers and devices for illegal content sharing, attracting a flood of takedown notices.

Quoting from the paper:

> ...we find that it is possible for a malicious user (or buggy software) to implicate (frame) seemingly any network endpoint in the sharing of copyrighted materials. We have applied these techniques to frame networked printers, a wireless (non-NAT) access point, and an innocent desktop computer, all of which have since received DMCA takedown notices but none of which actually participated in any P2P network.

[22] E.g., the pioneering UCSD research into the drivers of cyber-crime, in which researchers analyzed communications of criminals: http://www.cs.ucsd.edu/~savage/papers/CCS07.pdf

5 Factually erroneous and misleading statements in the report

In my opinion, the "Description of Technologies Involved" part of Dr. Jacobson's report contains multiple factually erroneous and misleading statements regarding the nature and function of Internet addresses, ISP logging, and other technical topics.

Since many of these statements have appeared almost verbatim in other Plaintiff's documents, I refer to Mr. Street's treatment of these statements in his report, in particular:

Item in Mr. Street's report	**Item and paragraph in Dr. Jacobson's report**
21	15, middle of par. 2
22	14, end of par. 6
24	15, end of par. 7
25	15, end of par. 2

Additionally, Dr. Pouwelse offers analysis of other verbatim statements by Dr. Jacobson. I refer to Dr. Pouwelse's treatment of these statements in his report, in particular:

Lines in Dr. Pouwelse's report	**Item and paragraph in Dr. Jacobson's report**
100–113	14, start of par. 2
114–125	15, middle of par. 2

The criticism of Dr. Kim's report, Section 3.3 applies to the same statements.

Finally, a number of statements by Dr. Jacobson that have not been repeated verbatim in previous documents have implications also treated in Mr. Street's report:

Item in Mr. Street's report	Item and paragraph in Dr. Jacobson's report
26, 27	19, implies that MediaSentry can reliably identify remote file sharing programs
28	22, implies that an ISP can accurately identify the customer controlling a computer that had a given IP address at a certain time

6 Unintentional File Sharing

Files are often shared on peer-to-peer networks unintentionally and without user knowledge. Malicious programs can offer files for download without user knowledge.

Users are easily tricked into downloading and installing malware that is offered for free and advertized as "toolbars", "screensavers" and similar. The malware then infects their computers and joins peer-to-peer networks.

It is notable that the list of files allegedly retrieved by MediaSentry software from a computer engaged in file sharing contains a program named "YSB toolBar.exe" (appears on `RoyMNH0988`), as well as evidence that suggests that the program has been installed on the computer.[23]

I suspect that "YSB toolbar" is spyware or malware. Information available regarding software by this name includes:

- http://www.pantheraproject.net/wiki/index.php?title=YSBToolbar – software **starts peer-to-peer software, in particular Gnutella-based file sharing software.**

  > "What does it do? We don't have a lot of information about it yet, but we do know that it restarts Shareaza after the user shuts it down, probably as a way to spread itself more efficiently on the network. It is probably installed by the user after it comes up as a result for a random query."

  Shareaza is a Windows p2p client which supports Gnutella.

[23] A file named "Shortcut to YSB toolBar.lnk" exists in the same folder; such "shortcut" files are created by software during its installation.

- http://forums.spybot.info/showthread.php?t=3594 – software does not allow itself to be removed and restores itself after removal attempts.
- http://www.nuker.com/container/details/downloader_ysb_toolbar.php –

  > Comes bundled and may be silently installed. Downloads and installs third party software including adware and internet search software.

- Symantec, a leading anti-virus anti-malware software vendor, has products that remove "YSB Toolbar" software and provides the following additional information[24]:
  - Software is normally installed via ActiveX.
  - The software can be distributed with Trojans.

  The former means that a computer can be infected unbeknownst to the user in the course of normal web browsing by a malicious website. The latter suggests that the software may come packaged with and be installed by other malware.

If "YSB toolBar.exe" really is the software described above, it is highly likely that the computer's owner has lost his or her control over the computer to malware and was not in control of the computer sharing files.

This is in fact a common situation, described in security publications: users do not realize that their folders with sensitive documents are shared on the Internet. Furthermore, this problem of unintended data sharing has become a very serious risk for US businesses. This is characteristic of Gnutella clients in particular, as described, for example, in InformationWeek's "Your Data And The P2P Peril" by John Foley:[25].

> What might have been a minor breach of IT policy at Pfizer last year cascaded into a serious security incident when the personal data of 17,000 employees and former employees leaked onto a

[24] http://www.symantec.com/security_response/writeup.jsp?docid=2005-090615-4926-99&tabid=1

[25] http://www.networkcomputing.com/channels/security/showArticle.jhtml?articleID=206904104

peer-to-peer network. Connecticut's state attorney general, concerned that state residents were at risk, launched an investigation. At least one former employee filed a lawsuit against the company.

It all started when the spouse of a Pfizer employee used file-sharing software on a company laptop, presumably to swap music or other content with other P2P users. Unknowingly, the laptop user also exposed 2,300 work files, including those containing sensitive Pfizer employee data–names, Social Security numbers, addresses, and bonus information resident on the laptop.

Pfizer isn't the only company to have its sensitive data exposed in this way. A former employee of ABN Amro Mortgage Group last year exposed spreadsheets with personal data on 5,000 customers from a home computer loaded with the BearShare file-sharing program. And last fall, a terrorist threat assessment of Chicago's transit system, completed by Booz Allen Hamilton under contract to the Federal Transit Administration, surfaced on a P2P network.

7 Disclosure.

My rate for analysis of data and preparation of the expert report in this case is $100 per hour, my rate for testimony at a deposition or trial is $200 per hour, and additional costs related to such analysis and testimony, including travel, are reimbursed as incurred.

8 Qualifications

- I am employed as a Research Assistant Professor at the Dartmouth College's Department of Computer Science. My research concentrates in Computer Security, system and network intrusion analysis and forensics.

- I also hold an appointment as a Principal Security Technology Adviser to Dartmouth's central IT organization, Peter Kiewit Computing Services.

- I am affiliated with the Dartmouth's Institute for Security, Technology, and Society (ISTS)[26], a leading research institution in computer security and privacy. From ISTS mission statement:

  > The Institute for Security, Technology, and Society (ISTS) at Dartmouth College is dedicated to pursuing research and education to advance information security and privacy throughout society. ISTS engages in interdisciplinary research, education and outreach programs that focus on information technology (IT) and its role in society, particularly the impact of IT in security and privacy broadly conceived. ISTS nurtures leaders and scholars, educates students and the community, and collaborates with its partners to develop and deploy IT, and to better understand how IT relates to socio-economic forces, cultural values and political influences.

- I provide research leadership for the Dartmouth's Computer Security Initiative, a program that educates undergraduate and graduate students in applied computer security, and employs them to assess and help secure Dartmouth College networks.

- I am a Subject Matter Expert for the Information Assurance and Technology Analysis Center (IATAC)[27], featured in its IA Newsletter vol. 8 no. 2.[28] IATAC operates operates under the DoD Scientific and Technical Information Program (STIP).[29]

- I taught and developed curriculum for innovative Computer Security and Advanced Operating Systems courses taught at Dartmouth.

- I presented the research results in Trusted Computing achieved at the Dartmouth PKI/Trust Lab at the TRUST 2008 conference, an influential forum for systems security research. The paper on which I am the first author received the prestigious **Best Paper** award.

- I participated as a Computer Security researcher in multiple projects funded by the US Department of Justice, Department of Defense, and

[26] http://www.ists.dartmouth.edu
[27] http://iac.dtic.mil/iatac/
[28] http://iac.dtic.mil/iatac/download/Vol8_No2.pdf
[29] http://iac.dtic.mil/iatac/history.html

the National Cyber Security Division (NCSD) of the Department of Homeland Security (DHS).

- I participate in the Trustworthy Cyber Infrastructure for the Power Grid (TCIP) project[30] sponsored by the National Science Foundations, DHS, and the US Department of Energy. My research into vulnerabilities of the power grid protocols has been presented at a leading security practitioner conference in 2008.
- I presented my research at many peer-reviewed academic and practitioner security conferences and workshops.
- My NSF biographical sketch is attached.

Sergey L. Bratus, Ph.D.
Research Assistant Professor
Dept. of Computer Science
Dartmouth College

Date:

[30] http://www.iti.illinois.edu/content/tcip-trustworthy-cyber-infrastructure-power-grid