Global Information Assurance Certification Paper

A Forensic Primer for Usenet Evidence

GCFA Gold Certification

Author: Mark Lachniet, mark@lachniet.com

Adviser: Charles Hornat (mrcorp@yahoo.com)

Accepted: January 23rd 2006

Outline

1. Introduction

This document is intended to provide an overview of the Usenet on the Internet, including the NNTP protocol and types of evidence of Usenet abuse that may be present on permanent storage devices such as hard disks and flash drives. A cursory review of the Usenet shows that the Usenet is frequently used as a means of anonymous transmitting and receiving digital content including pirated software, intellectual property such as movies and music, and possibly even child pornography. Due to this fact, investigating abuses of the Usenet system has become of interest to the private and public sectors alike. Indeed, Usenet abuse has apparently become a focus of the Recording Industry Association of America (RIAA). In October, 2007 a number of parties including "Arista Records, Atlantic Recording, BMG Music, Capitol Records, Caroline Records, Elektra Entertainment Group, Interscope Records, LaFace Records, Maverick Recording, Sony BMG Music Entertainment, UMG Recordings, Virgin Records America, Warner Bros. Records and Zomba Recording" filed a lawsuit against a well-known Usenet provider - Usenet.com[1]. Due to the prevalence of (and difficulties in investigating) Usenet abuse, this paper has been created.

2. Usenet Overview

Usenet, as accurately described at Wikipedia.org[2], can be thought of as "a global, decentralized, distributed Internet discussion system" that utilizes the NNTP protocol for communication. While originally used via the UUCP protocol and over modem connections, most modern NNTP traffic takes place over the Internet. The NNTP protocol is defined in RFC 977[3] and RFC 1036[4]. According to RFC 977, the NNTP is a "protocol for the distribution, inquiry, retrieval, and posting of news articles using a reliable stream-based transmission of news." In essence, Usenet is a "store and forward" system very similar to standard SMTP e-mail. However, whereas standard SMTP contains a destination of one or more unique e-mail addresses, NNTP specifies a destination newsgroup (such as alt.binaries.warez), which may be subscribed to by any number of NNTP servers and clients in an intercommunicating network. Thus, a single message, while being initially posted through a single server, it may subsequently be copied to hundreds of other servers, and viewed from these other servers by their subscribers.

Just as with SMTP e-mail, it is possible for binary attachments (e.g. ZIP archives, MP3 music files, DiVX and XViD movie files, GIF and JPG images, etc.) to be attached to these messages. While some textual information posted in NNTP groups may be of interest to an investigator, it is binary attachments, and in particular inappropriate binary attachments, that will be the focus of this document. Indeed, some high-profile intellectual property breaches have already been attributed to the Usenet system[5]. To demonstrate the prevalence of Usenet binaries, consider which Usenet newsgroups are most active. Newsadmin.com, a web site that tracks Usenet usage statistics, reported that the following newsgroups

were the most active binary newsgroups for the sampling period of February 25th, 2008:

#	Newsgroup	Unique Access	Percent
1	alt.binaries.movies.divx	194,538	13.7
2	alt.binaries.dvdr	129,020	9.1
3	alt.binaries.dvd.erotica.classics	65,578	4.6
4	alt.binaries.tv	58,355	4.1
5	alt.binaries.erotica	48,376	3.4
6	alt.binaries.warez.ibm-pc.0-day	40,853	2.9
7	alt.binaries.ijsklontje	40,556	2.9
8	alt.binaries.pictures.erotica.orientals	26,509	1.9
9	alt.binaries.pictures.erotica.older-women	20,836	1.5
10	alt.binaries.dvd.erotica	19,431	1.4
11	alt.binaries.erotica.divx	16,001	1.1
12	alt.binaries.erotica.collections.rars	15,941	1.1
13	alt.binaries.erotica.pornstars.90s	14,138	1.0
14	alt.binaries.pictures.bluebird.reposts	12,604	0.9
15	alt.binaries.multimedia.erotica	12,540	0.9

From here, we can clearly see the popularity of newsgroups specializing in pirated movies, pornography, and pirated software on the servers monitored by the site. Some newsgroups, such as alt.bin.pictures.child.pornography may be explicitly targeted towards the abuse of underage minors. For these reasons, a review of how the NNTP protocol and Usenet function, and how to investigate such abuse is of use to the forensic community and law enforcement, and is presented in this paper.

3. Usenet Message Example

First, it is of use to understand what a Usenet message looks like in its raw format. The following is an example of the headers of an actual NNTP formatted message posted to the Usenet:

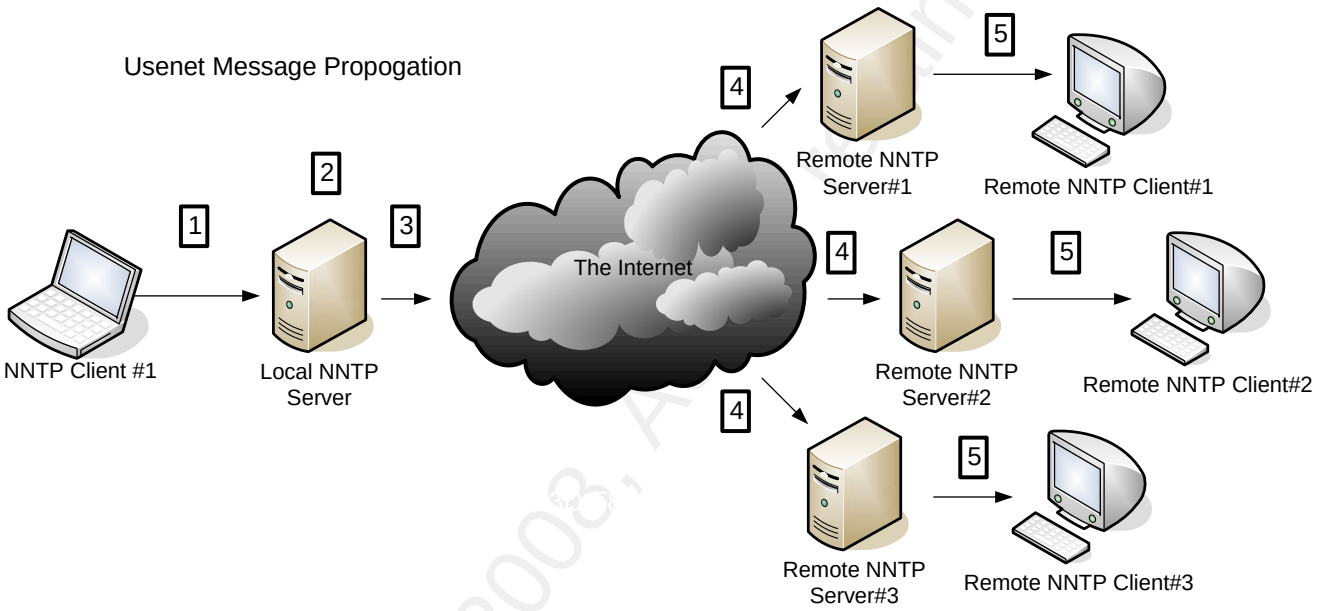
```
Path:
border1.nntp.dca.giganews.com!nntp.giganews.com!feed2.newsreader.com!newsreader.com!npeer.de.kpn-
eurorings.net!news.tele.dk!news.tele.dk!small.news.tele.dk!news.astraweb.com!newsrouter-
eu.astraweb.com!eweka!hq-
usenetpeers.eweka.nl!81.171.88.219.MISMATCH!newsreader30.eweka.nl!not-for-mail
From: "Apollo" <obscured@email.whatever>
Subject: were can i download the series?
Newsgroups: alt.binaries.battlestar-galactica
Date: Sun, 1 Jan 2006 16:47:29 +0100
Lines: 7
Message-ID: <43b7f8e3$0$1030$c807b3c@newsreader30.eweka.nl>
Organization: Eweka Internet Services
NNTP-Posting-Host: Eweka Internet Services
X-Trace: Posted by Eweka Internet Services, http://www.eweka.nl
X-Complaints-To: abuse@n-o-s-p-a-m.eweka.nl
Xref: number1.nntp.dca.giganews.com alt.binaries.battlestar-galactica:824361
```

In this example, you can see some similarities and differences from e-mail. For example, there are headers common to both, such as Subject and Date. However, some key differences exist. For example, rather than a To field, we have a Newsgroups field. A complete treatment of message header fields can be found in RFC 1036, which details the following mandatory fields: From, Date, Newsgroups, Subject, and Path. In addition, several optional fields are described including: Reply-To, Sender, Follow-up To, Expires, References, Control, Distribution, Organization, Keywords, Summary, Approved, Lines, and Xref.

As with SMTP mail, it is possible to easily spoof some of these fields. For example, the From: field is taken from whatever the user types into their NNTP client and is obviously of dubious accuracy. Other fields, such as the Message-ID field, should be more difficult to spoof, and would require administrator access to a participating (and presumably trusted) NNTP server. The Message-ID field shows you which news server originally received the message. For the Usenet to work, this field has to be globally unique. Indeed, this is how messages are propagated between Usenet servers without creating redundant copies of the message, as a server will not accept a NNTP posting where the Message ID matches one already stored by the server. Other fields such as the Path can be useful in determining which servers a particular message was routed through. In the above example, it is possible to see that the message originated from a client connected to the newsreader30.eweeka.nl NNTP host, where it was assigned a unique message ID. From there, it appears to have been routed through additional hosts including newsroutereu.astraweb.com, news.astraweb.com, small.news.tele.dk, news.tele.dk, npeer.de.kpneurorings.net, newsreader.com, feed2.newsreader.com, nntp.giganews.com, and finally border1.nntp.dca.giganews.com where it was retrieved. Each of these servers might have log or file data of interest to an investigator.

4. Usenet Message Propagation

In order to understand how NNTP messages propagate, a simplistic message flow is described below:



1. A client with access to a NNTP server posts a message. The message may be text, a binary attachment, or both. This communication is between the client and the server directly, and usually takes place over TCP port 119. In some cases, alternate ports and encryption may be used. Let us suppose that the client had posted their message to the newsgroup alt.binaries.battlestar-galactica.
2. The local NNTP server receives the message and assigns it a unique Message-ID field.
3. The local NNTP server communicates with its remote NNTP server peers (other Usenet servers). The servers copy messages between servers for newsgroups to

which both are subscribed..

4. If the peer servers do subscribe to the newsgroup, they look to see if they already have a copy of the message based on the unique Message-ID field. If they subscribe to the newsgroup but do not have the specific message, it is transmitted from the local NNTP server to remote NNTP server. If the server already has the message, or if it does not subscribe to that newsgroup, it ignores that message.
5. A second client, the remote NNTP client (the consumer) connects to their Usenet server. The client will query the server for a list of messages for the newsgroups to which they are subscribed, and a list of messages is presented. The remote NNTP client then selects the required messages, and downloads the content to their local workstation. NNTP client software will typically strip out the attached binary and save it to local storage.

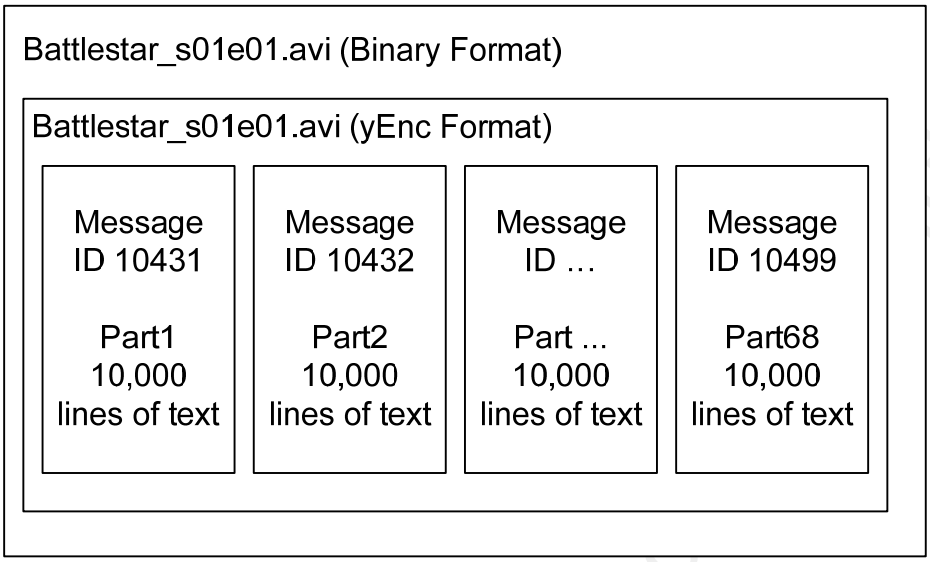
It should be noted that NNTP servers do not store message data indefinitely. Indeed, it is common for high-volume newsgroups to maintain only a few days of message data, due to the huge volume of traffic on some Usenet groups. This is commonly referred to as "retention" and is used (and marketed) by providers as a metric for the quality and completeness of their service's servers. This also means that potential evidence may be volatile, and quickly removed from servers, which may force a forensic investigator to work quickly or risk the loss of evidence.

5. Usenet Binary Attachments

As noted above, it is possible to attach a binary file to a NNTP message. As with SMTP e-mail, there are a number of ways to do this. In all cases, a binary file is first converted to an ASCII message that can be transmitted. The primary protocol used for this conversion with Usenet is the yEnc format[6], though other encoding formats such as BASE64, BinHex, UUencode, and Quoted Printable may be used. As such, a forensic analyst should include yEnc files (at a minimum) in their analysis when investigating Usenet abuse, but should be aware that other MIME types are possible.

The binary files themselves can be in any number of formats, including their native format (JPG, GIF, AVI, etc.) or archive files. While an analyst may be lucky enough to have a piece of evidence contained entirely in a single message in yEnc format, this will not occur for larger files due to a number of constraints. Specifically, NNTP articles have a maximum size of 10,000 lines of text, limiting the size of a possible single-message binary attachment. For this reason, it is not unusual to find a file broken across several different NNTP articles that must be reassembled by a client before the binary can be extracted. Most modern Usenet client programs will automatically find and reassemble the various parts of a file posting for the user, minimizing the need for a detailed understanding of how the system works.

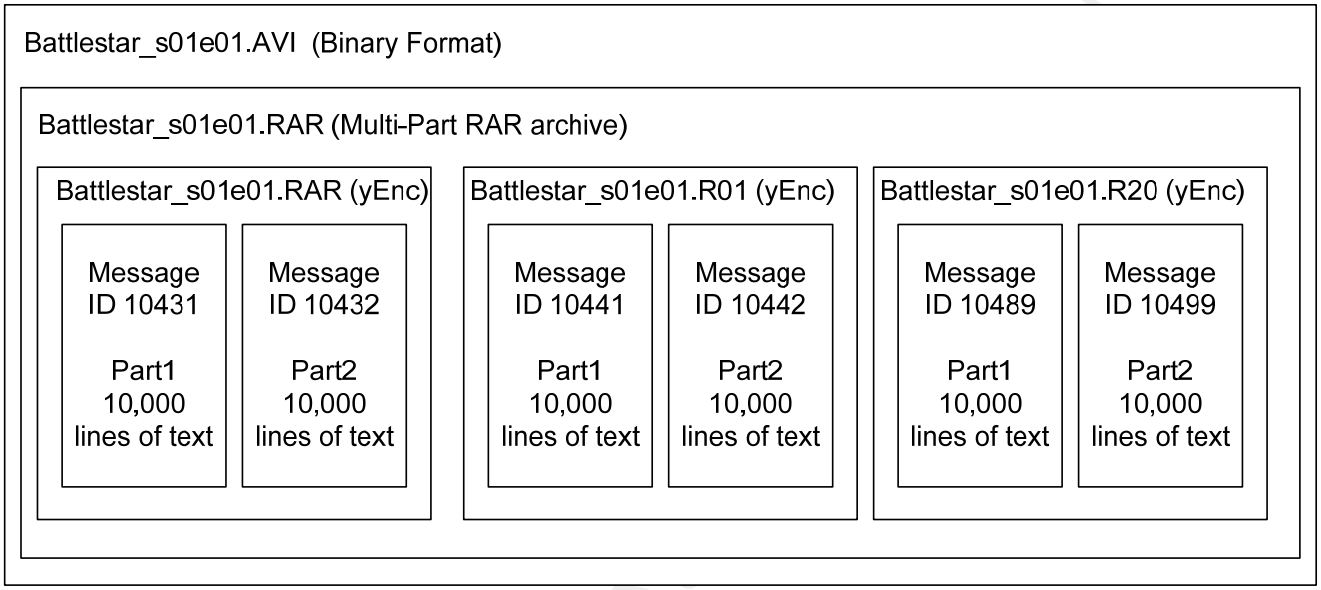
Consider the following diagram, where we will assume that a message was posted to a Usenet server containing a 300mb AVI file of an episode from the television show Battlestar Galactica:



In the above example, we can see that a single binary AVI file has been encoded by the client into yEnc format, and then posted in 68 unique 10,000 line messages to a Usenet server. In order for a remote NNTP client to receive this binary attachment, they would have to connect to a news server that had all 68 parts of this message, download the 68 parts in text format, reassemble them, and then convert the resulting yEnc text back into a binary.

Unfortunately, much of the content that is posted to binary newsgroups is too large to be posted in its native format (even in a large number of 10,000 line messages). For this reason, it is common to see multi-part archives being used. In particular the RAR compression protocol seems to be predominantly used, although other protocols such as ZIP may sometimes be observed. For general information on the RAR protocol and WinRAR program, refer to the rarlab.com web site[7]. When a multi-part RAR archive is created, one or more files are compressed and broken into a number of smaller compressed RAR-formatted files which can be independently downloaded and combined at the end destination. For example, consider the

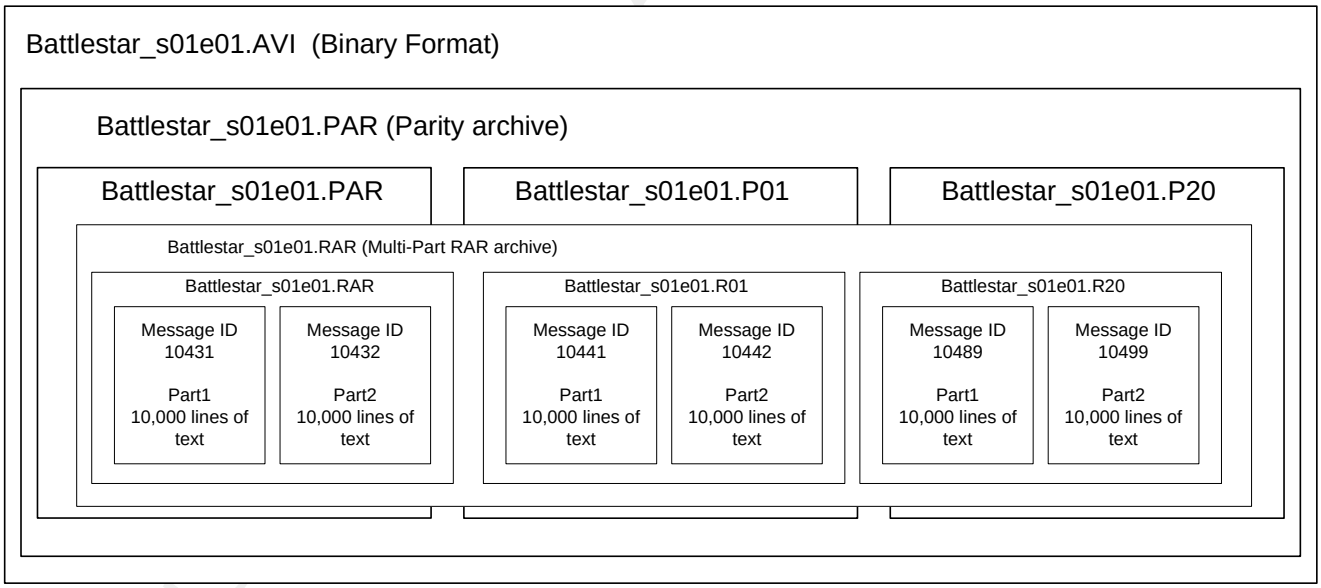
following example:



In the above example, we can see that the video file has been first converted to a compressed RAR archive, and then split into a number of RAR files (the first file suffix is .RAR, the second is .R01, and the last is .R20). These RAR files were then converted to a text format using yEnc, and then posted to the NNTP server in chunks of 10,000 lines each. In order for a remote NNTP client to receive this binary attachment, they would have to connect to a news server that had all 68 parts of this message, download the 68 parts in text format, convert the various messages into their 20 RAR files, and then use a RAR compression program to combine these 20 RAR files and extract the binaries. Thus, a forensic analyst should also pay careful attention to RAR files, and other files containing compressed data.

Unfortunately for the forensic analyst, the evidence may be further obscured (or in some cases helped) by the use of parity archive files such as PAR and PAR2. Due to the way in which the Usenet operates, it is not uncommon for a single NNTP

server to have missed some of postings related to a single file attachment due to communication failures, disk space issues, or incomplete postings from the originating side. To accommodate for this lack of reliability in NNTP servers, parity archives are used to allow for the successful recovery of a binary attachment even if some of the Usenet articles are corrupted or not available. These parity archives are extremely similar in concept to the disk-based RAID 5 system, in which one or more parity files can be used to recreate damaged or missing data. For example, an individual might create a number of PAR and post them along with RAR files, so that if one or more of the RAR files were corrupted or missing, enough could be recreated from the PAR files to extract the contents undamaged. Consider the following diagram:



In the above example, we can see that the binary attachment has been compressed into a multi-part RAR archived, and then processed by a PAR program into multiple parity archives, encoded into yEnc format, and then posted in a number of 10,000 line messages. . In order for a remote NNTP client to receive this binary

attachment, they would have to connect to a news server that had all 68 parts of this message, download the 68 parts in text format and extract them to binary PAR files, in addition to the RAR files. They would then have to use a PAR program such as QuickPar[8] to repair or recreate any of the missing RAR files. They would then use a RAR program to combine the multiple RAR files and extract the AVI binary. Due to this, parity archive files are of interest for a forensic analyst.

6. Usenet Evidence File Carving

As seen above, there are a variety of file types that may be of interest when investigating Usenet abuse. These file types should obviously be included in analysis of target media, including timeline analysis and disk carving. A brief summary of the file protocols, and references to further information follows:

yEnc Files

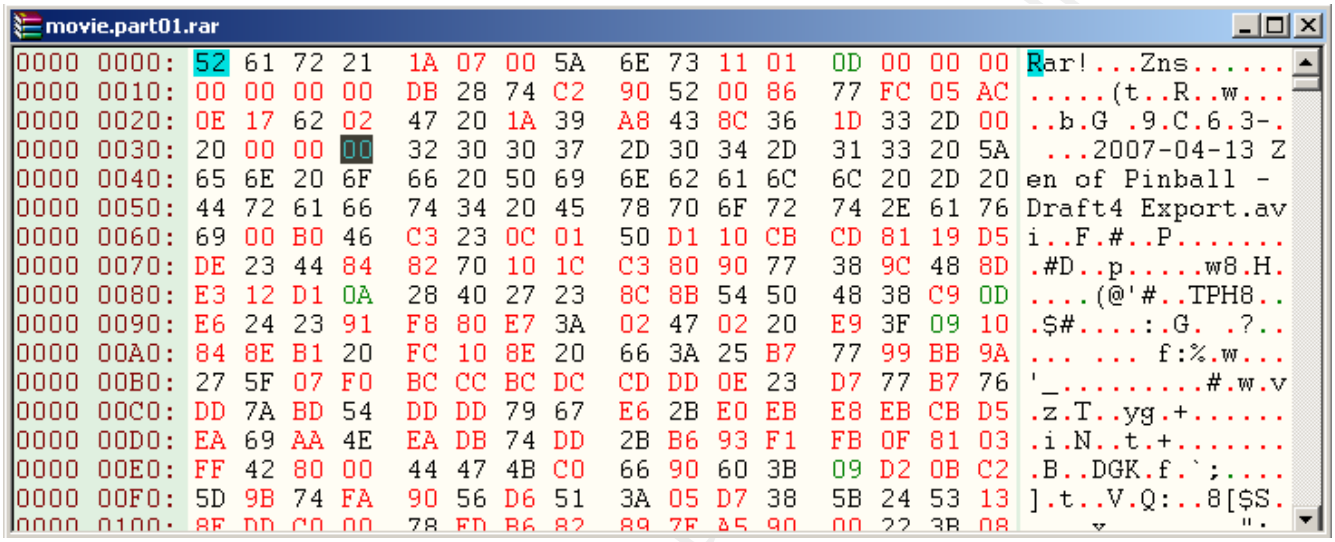
yEnc formatted files may be discovered in cache directories, slack space, etc. and may be identified and carved by their headers and footers. More information on the yEnc format may be found at http://www.yEnc.org. A Windows version of the yEnc program (to convert to and from the yEnc format) can be downloaded from http://www.yenc32.com/download.php. By observing a file in the yEnc format, it is possible to see that yEnc files may be identified by the following headers and footers, which could be fairly easily adapted to file carving programs:

```
=ybegin line=128 size=102210 name=underage_inappropriate.jpg
) )=J*:tpsp*+++*r*r**) *m*/-=n=n=n-   (a lot of text removed)
=yend size=102210 crc32=1E2501D1
```

RAR Files

Compressed RAR files may be discovered on the target media. More information on the RAR format can be found at the Wotsit.org web site[9]. The header of a RAR

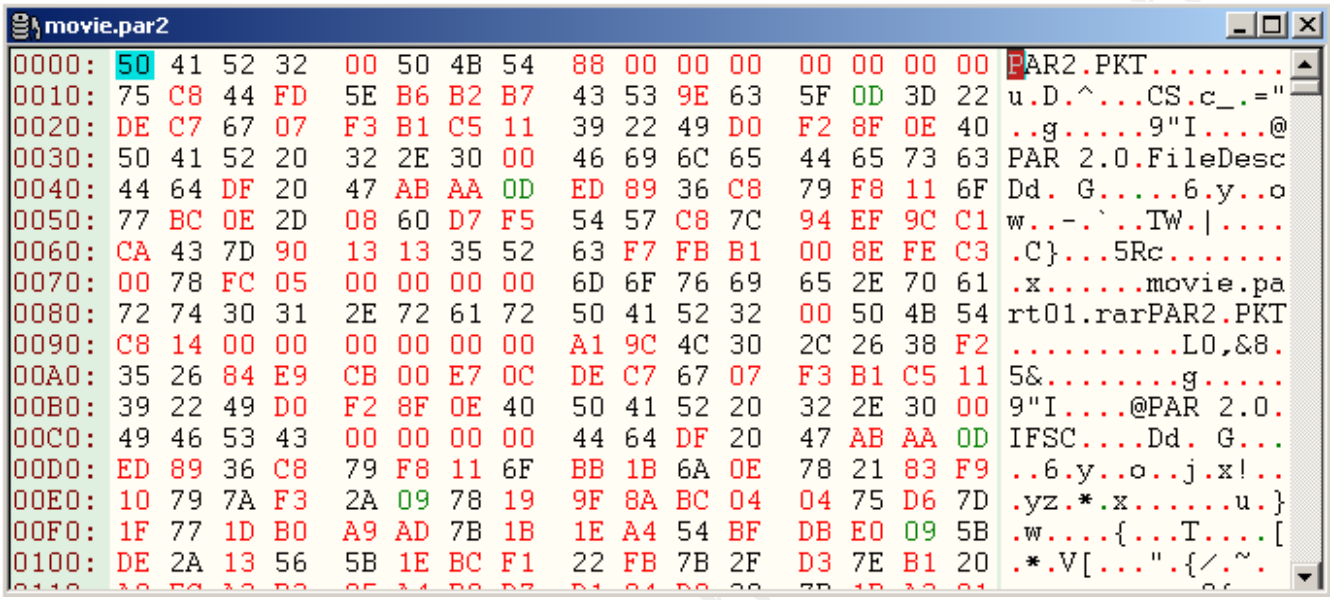
file can be identified by the header "Rar!" and appears to have the contents filenames near the top of the file as seen below:



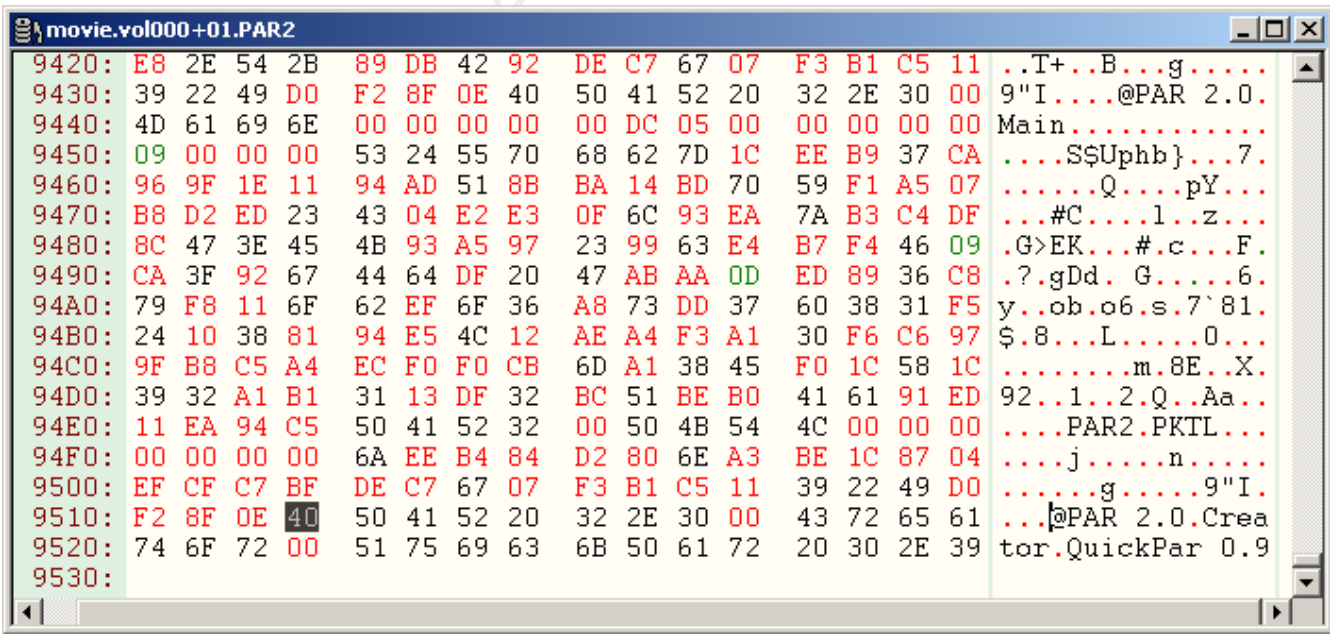
Unfortunately, the footer for the file will apparently vary depending upon the RAR file's content, which will most likely require a forensic analyst to carve the files well beyond the actual end of file if no filesystem metadata information is available. Analysis of a sample of RAR files by this author has shown that RAR files do not have a consistent footer that can easily be programmed into carving tools.

PAR and PAR2 Files

Parity files, as detailed in the 2.0 specifications found on Sourceforge[10] appear to also have a common header of 'PAR2 PKT' as seen below:



Unfortunately, it appears that the packet footer may vary depending upon the type of content in the PAR file and the file creator. That said, it was noted that the footer of the field does appear to contain a text string indicating which program created the parity archive. Consider the following file footer:



This footer, if consistent for a particular set of files, could be used to more accurately carve the parity archive data. Failing this, and without proper filesystem metadata to better refine the data carving work, it will probably be necessary to carve well beyond the file's end and manually manipulate the file(s) to recover the relevant data.

Example Foremost.conf file

The following is an example of a configuration file that can be used with the Foremost carving tool. While it can be fairly efficient in carving yEncoded files, as they have a file footer to identify the end of file, it is less efficient at RAR and PAR2 files, which must be carved to an arbitrary length and manually fixed before they can be recovered:

```
#
# Foremost configuration file
#
#----------------------------------------------------------------------
# Usenet    (NOTE THIS FORMAT HAS A BUILTIN EXTRACTION FUNCTION)
#----------------------------------------------------------------------
#
        yEnc    y       20000000  =ybegin =yend
        RAR     y       20000000  Rar!
        PAR2    y       20000000  \x50\x41\x52\x32\x00\x50\x4b\x54
#
```

7. Usenet Clients

One way in which Usenet activity can be readily identified is by analyzing the software packages installed on the target media. For example, there are a number of Usenet client programs for Windows, UNIX and Macintosh that could contain evidence of usage. By reviewing the target's registry, "Program Files" or equivalent directories, it may be possible to identify the use of popular NNTP clients such as Agent by Forte (http://www.forteinc.com/main/homepage.php), NewsReactor (http://www.daansystems.com/newsreactor/) or any number of other clients. Some of these programs will store information such as subscriber names, passwords, activity logs, and message headers on the local disk or in the registry. Obviously, the exact artifacts left by these clients will vary from program to program.

8. NZB Files

As accurately described on the Wikipedia.org[11] web site, a NZB file is "an XML-based file format for retrieving posts from NNTP (Usenet) servers. The format was conceived by the developers of the Newzbin.com Usenet Index. NZB is effective when used with search-capable websites. These websites create NZB files out of what is needed to be downloaded. Using this concept, headers would not be downloaded hence the NZB method is quicker and more bandwidth-efficient." The following is an example of an NZB file:

```
<?xml version="1.0"?>
<!DOCTYPE nzb
  PUBLIC "-//newzBin//DTD NZB 1.0//EN"
         "http://www.newzbin.com/DTD/nzb/nzb-1.0.dtd">
<nzb xmlns="http://www.newzbin.com/DTD/2003/nzb">
  <file subject="Linux Debian-3.1r5-i386-binary- CD1 Lucky Strike.par2 (1/1)"
        date="1175058765" poster="Yenc@power-post.org (Lucky Strike)">
    <groups>
      <group>alt.binaries.cd.image.linux</group>
    </groups>
    <segments>
      <segment bytes="423495" number="1">b5488$4609f95d$55b3de22$8282@news.usenext.de</segment>
      <segment bytes="423450" number="2">d9f8c$4609f96d$55b3de22$8282@news.usenext.de</segment>
    <!-- ... -->
    </segments>
  </file>
  <!-- ... -->
</nzb>
```

In the above, we can see all of the specific Usenet messages (noted by the unique Message-ID field and newsgroup) that are necessary to get the complete binary file. For file carving purposes, and a header of "<!DOCTYPE nzb" and footer of "</nzb>" can be used to identify NZB files. The existence of NZB files that reference clearly inappropriate binary postings could be indicative of intent. In addition, as the unique Message ID of particular segments of a file are given, one could then perform a keyword search of the target media to find these file parts and recover them.

9. Internet Browser Activity on Target Media

While most Usenet communications seem to take place over dedicated NNTP clients, it is not unusual to find evidence of Usenet abuse in Internet browser histories. For example, there are a number of web-based search engines that allow a user to search through Usenet groups looking for a particular type of content. While a complete inventory of all such Usenet search engines is beyond the scope of this paper, http://www.newzleech.com/, http://www.binsearch.info/ and numerous other web sites provided this capability as of February 2008. In addition to search engines, there are also a number of services that allow users to download Usenet binaries directly from a "front end" website (see http://motzarella.org and similar sites). Finally, there are reportedly some NNTP to e-mail (and vice versa) gateways that might be used to obtain content, but these seem to comprise a small portion of total usage. Due to the availability of web sites such as these, a complete review of Internet browsing history (including URLs in deleted and slack space) should be performed when investigating Usenet abuse.

10. Analyzing Network Activity

If an analyst has access to network activity logs (such as those provided by routers, firewalls and protocol analyzers like Wireshark) it may be possible to identify Usenet abuse from these sources. In particular, the NNTP protocol takes place over TCP port 119 in plain text. If activity is noted on this port, it is likely that the packet contents will be unencrypted and available for review. Some providers, such as GigaNews.com offer encrypted SSL tunnels to protect their users' NNTP traffic from this type of monitoring. NNTP over SSL can take place on any port (depending upon the provider) but may be found on TCP port 563, or in some cases 443.

11. Requesting Help from Usenet Providers

In some cases, it may be necessary to obtain records from Usenet providers when performing an investigation. Although this author has not gone through this process, a high-level overview of what might take place follows. First, a forensic analyst would have to identify a suspicious Usenet posting, for example one that contained child pornography, and identify the specific Message ID field of that message. By analyzing the Message ID, it should be possible to determine which Usenet server originally received the message, or failing that the next server upstream to receive the posting. Once the Usenet server that originally received the message has been identified, the analyst may be able to obtain a subpoena to request server and network logs from the Usenet service provider to determine which IP address posted the message in question. From there, a second subpoena to the ISP that owns the suspect IP address may turn up subscriber information such as the name and address of the person who was assigned that IP address at the time of the posting.

Unfortunately, due to the volatile nature of the logs involved, which are high in volume, and low in retention priority, it would be difficult to follow this trail of evidence in a timely fashion. This is further complicated by the fact that many service providers apparently market themselves intentionally as providing a high level of "privacy" to users, and that they intentionally do not keep records of activity at all. For these entities, it might be impossible to identify individuals who have posted more than a few days in the past. Finally, Usenet providers are established in multiple sovereign nations, and an investigation could require extensive international cooperation to follow the trail to a specific

individual.

12. Usenet Investigation Problems

As seen in this document, investigating Usenet abuse is far from trivial. There are a number of reasons why investigating Usenet evidence is difficult for a forensic analyst, and this is most likely why Usenet is such a popular tool for transmitting illegal content. Among the difficulties that a forensic analyst may face in a Usenet investigation are the following:

- Internet Anonymity. Due to the fact that it is trivially easy to change the identification strings (such as the From: header), it will most likely be necessary go to Usenet providers for logging information, which they may or may not have kept. These providers may be protected under the "safe harbor" laws, and argue that they simply allow access to content, and do not explicitly monitor what their users access or encourage illegal activity.

- Volatility. Usenet data, particularly binary downloads, are extremely high-volume. Records, particularly on the Usenet servers, may be kept only for a very short period of time.

- Difficulty in Carving. As can be seen, most files without existing filesystem metadata are not easily carved due to the lack of a consistent file footer. While it might be possible to create carving routines that could intelligently analyze the header and content in these files and correctly estimate the ending of the file, this capability is certainly not present in any free or open source products that this author is aware of.

- Difficulty in Analysis. As seen in this document, Usenet binary evidence may often be found in nested and difficult-to-understand file structures. For example, evidence may include items such a movie file which has been converted to a multi-part RAR file, which has then been converted to a multi-part PAR file, which has then been converted to yEnc format. This places a particular burden on the forensic analyst to understand, piece together, and describe in a coherent manner.

13. Proposed Methodology

In order to provide some high-level guidelines of how to investigate Usenet abuse, the following methodology is proposed. Obviously, every forensic engagement is different, and will require a unique approach and the active involvement of an informed and capable analyst. That said, the following minimum steps are recommended as a starting point for the forensic analyst:

1. Perform data collection and establish a right to analyze the data. The first step to any forensic process should be to make sure that you have been given written permission to perform the analysis. One way to do this is a business contract that directs you to perform the work. In addition, it is ideal to collect data about the investigation, the type of data you are looking for, etc. so that you can narrow your search terms and work to as small of a scope as possible. The more focused your investigation is, the more quickly (and inexpensively) you can perform the investigation. Forensic work, like any work, will inevitably involve a cost/benefit analysis on how much work to perform versus how good the results must be. The greater the need for accuracy and completeness (if indeed that is possible) the more expensive and time consuming it will be.

2. Image the disk. As in any forensic engagement, it is preferable to image the target media in a secure manner, and without disturbing the primary source of evidence. Using a write blocker or software toolkit such as Helix is an ideal way to create a disk image. This will vary depending upon the analyst's preference and is better documented in other sources.

3. Mount as loopback and manually analyze. I recommend that before any further analysis is done, a preliminary manual investigation of the drive be performed. One easy way to do this is to mount the disk image as a loopback device so that it can be viewed as a locally attached disk. This can be done in UNIX by mounting the image with the '-t loop' flag or in Windows using a product such as Mount Image Pro. Once the drive is available for browsing without the risk of modifying the primary source of evidence, an analyst may:

   - Look for browser history using a tool such as the open source Pasco tool, or the commercial tool Net Analysis (http://www.digital-detective.co.uk/netanalysis.asp). Make sure to search through deleted and slack space if at all possible. Go through the browsing history and look for any sites that you don't recognize as legitimate, and review them to determine if they are relevant. Pay attention to search terms in URLs that may be indicative of inappropriate activity.

   - Look for installed programs that are indicative of Usenet use. Usenet clients, yEnc and yDecode, WinRAR, Parchive and other programs may be of interest.

   - Use a simple filesystem search (e.g. Windows search) to find undeleted files of interest. A preliminary search of undeleted space for media (e.g. AVI, MPEG, etc.) and Usenet files (e.g. NZB, PAR, PAR2, RAR, etc.) may quickly reveal evidence that would take several hours to find using disk carving tools. In

addition, keyword searches of files containing terms such as 'alt.binaries', 'warez', 'porn', and the like may be fruitful.

4. Analyze the Registry. There may be evidence of installed programs or Usenet history contained in the registry. By loading registry files in a viewer such as Access Data's Registry Viewer and performing keyword searches, one may be able to identify relevant evidence.

5. Create a timeline. The next step to take would be to create a timeline of filesystem activity from the existing metadata. This can be done using the 'mac-robber' and 'mactime' tools. Reviewing this timeline for evidence of software, files, etc. being created and deleted could quickly point the analyst to areas of interest. For example, an analyst may be able to note the creation of a number of temporary files, then the creation of PAR and RAR files, and finally the creation of an AVI or JPG file in a relatively short timeframe.

6. Perform keyword searches of the target media. In some cases, inappropriate material may be found in files with apparently innocent file names, or as fragments in slack space. The forensic analyst should create a "naughty words" list that includes phrases that are relevant both to the specific investigation (e.g. 'child porn') and to Usenet in general. A suggested list of Usenet-specific search terms might include:

   - Usenet
   - NNTP

- Alt.binaries
- warez
- <!DOCTYPE nzb
- =ybegin
- =yend
- Rar!
- PAR
- PAR2

7. Carve files of interest. A next step would be to configure a carving tool to parse through the entire disk image and carve files of interest. A first pass to look for the "end destination" binary files (e.g. AVI, JPG, GIF files) should be conducted first, as this will most likely be fairly quick, and reveal content of interest. Failing (or in addition to) this, an analyst should carve files that are indicative of Usenet abuse such as yEnc, RAR, PAR, ZIP, NZB, etc. Unfortunately, as previously noted, many of these files do not have a discrete file footer, and will probably have to be carved with a "maximum file size" option to ensure that the entire file is retrieved. This is likely to consume a good deal of time and disk space.

8. Review firewall and network activity logs. If possible, the analyst

should obtain any sources of log information that pertain to network activity. Examples of this may include router, firewall, and proxy server logs. In addition, there may be information in personal firewall products such as Zone Alarm, Norton Internet Security, and others that may have evidence of interest. Focus on network activity to TCP ports 119 and 563, as well as any IP addresses or DNS names associated with known Usenet Service providers.

9. Obtain information from service providers. During the investigation, the analyst may discover information that could positively point an investigator to a specific piece of evidence. For example, the Windows registry or disk may include a username and password for a Usenet service provider. Or, an analyst may discover a particular piece of evidence that they are interested in, and be able to identify a unique Message ID field, and subpoena information on the poster from the provider.

14. Example Usenet Forensic Engagement

The following is a contrived example of a forensic analysis process, as it might be performed by an outside consultant. It is not intended to be complete or comprehensive, but rather give an example of how one investigation might take place. The following are the analyst's notes, which include the specific commands used and notes on the results of these commands. In the real world, these notes would then be included in a formal deliverable document that summarizes and organizes the findings. As in the real world, the assessment must be performed on a budget, so not *all* avenues of possible evidence are examined (though a customer might request further analysis work).

For this example, a simplified Windows XP disk image was created in VMWARE and analyzed. However, for the interested forensic analyst, a loopback file image that can be used to validate the foremost.conf file (or other third party tools) can be downloaded from http://lachniet.com/flash.img as of March 8th, 2008. This image file contains a number of deleted yEnc, RAR and PAR2 files, and matches the later portion of this example engagement. The forensic analysis workstation is a VMWARE image that was distributed in the SANS 508 class in Las Vegas, Nevada in the fall of 2008, and includes a number of tools such as TSK, Foremost, etc.

Analyst Notes - Sansorvino Giacomo, Inc. investigation - by: Mark Lachniet

11/28/07

11:17am Met with customer and completed data collection form:

Forensic Analysis Service Data Collection Worksheet

Customer Contact Information

Organization Name	Sansorvino Giacomo, Inc.
Main IT Location Name	3141 West Rhombus Way
Street Address	
City, State & Zip	Lansing, MI 48840
Internet Domain Name	SANSGIAC.ORG

Primary Contact	Nelly Nervosa	
Title	Director of Human Resources	
Phone Number	517.555.1212	
Alternate Phone Number	517.666.1212 (Cell)	Pager / Cellular:
Fax Number	517.555.1000	
Email Address	nervnelly@sansgiac.org	
Preferred method of contact	phone	
Alternate Contact		
Title		
Phone Number		
Alternate Phone Number		Pager / Cellular:
Fax Number		
Email Address		
Preferred method of contact		

Days, hours and time zone of normal IT operations	8am-5pm, EST

Specific issues to consider when contacting above contacts…
Upcoming holidays will have spotty coverage due to vacation time being used.

Goals and Outcomes

Please describe the goals and outcomes of this investigation
In general, what is the history and purpose of this investigation? Mrs. Nervosa reports that an Intern, Pamella Narcington, overheard the CIO having a discussion on the phone that made her feel uncomfortable. The CIO, one Richard McCheney was heard on 11/27 or 11/28 to say something in his phone conversation to the effect that he "really enjoyed that inappropriate picture of the little boy" and laughing. Mrs. Narcington then reported this to Mrs. Nervosa. Based on this information, Mrs. Nervosa discussed the issue with legal counsel and determined that an investigation was necessary to determine if any inappropriate material (particularly illegal material such as child pornography) existed on the workstation. Mrs. Nervosa then contacted Analysts International to perform an investigation. Mrs. Nervosa has stated that she doesn't believe that Mr. McCheney is the kind of person to view child pornography but she wants to perform an investigation anyway. She wishes the investigation to be done without the knowledge of Mr. McCheney, and that any potential evidence be preserved in case the incident needs to be escalated to law enforcement.
What type of data would you like to have analyzed or recovered? Any images that may have inappropriate material, Internet browsing history, or other items uncovered during the source of the investigation.
Who is the target audience of the deliverable documents? Are they technical or non-technical? Both technical and non-technical. The report should be of sufficient detail for a third party to replicate the results given the same source information.
Do you wish to have a complete package of all collected data provided to you, including recovered files, disk images and working documents? If so, will you provide storage media for this data or do you wish to have it provided as part of this service? Yes – an external 1TB USB hard drive will be provided

What format do the reports need to be in? Is there a document template that you wish to have used? Reports should be presented in printed and PDF format. No specific document template is required.
When does the analysis need to be completed? Are there any "hard dates" that need to be accommodated? The report should be complete within 2 working weeks (When Mr. McCheney returns from vacation) if possible.
Who is authorized to have communication and information about the case? How many meetings do you anticipate requiring? To whom will the results be presented? All reports should be discussed with, and presented to, only Mrs. Nelly Nervosa.
Will the results be used for disciplinary or legal action? If for legal action, is the case civil or criminal? Do you believe it is likely that testimony will be required? Depending upon the findings, it is possible that the case may go to law enforcement. As such, preservation of all original evidence is important. Depending upon findings, it may be necessary for the forensic analyst to testify. Mrs. Nervosa acknowledges that any time required for testimony or requisite preparation will be billable under a separate contract, to be determined.
How has the data or media been handled up until this point? Was any prior investigation conducted on the media? If so, when was the analysis conducted, who conducted it, and what was done? Also, where and how has the media been stored? Please provide any supporting documentation possible. Mr. McCheney's workstation has not been touched in any way, and is currently locked in his office

Analysis Parameters

Description of Media to be Analyzed	
Will you be providing complete computer CPUs? If so, how many?	1
Will you be providing hard drives? If so, how many and what type are they (e.g. IDE-66, IDE-100, SATA, SCSI-2, SCSI-3, etc.)	IDE-100
Are any of the drives part of a RAID array?	No
Will you be providing any removable media such as CD-ROMS, DVD's, Flash memory (from cameras, PDAs and cell phones)? If so, what type?	No
Approximately how many Gigabytes of data, in total, is to be analyzed?	40
How many forensic copies of the target media do you require?	1
Can you provide an adequate quantity of identical (or nearly identical) media for making forensic copies of the target media? (e.g. extra hard drives of the same or similar model)	Yes (1TB drive for image)
Is the media in good working condition? Are there any known disk errors or other problems that should be considered?	Yes
Analysis Parameters	
Internet History Analysis	
Do you wish to have an analysis performed of Internet History (e.g. browser history for Internet Explorer and Firefox)? If so, what type of browser was used (if known)?	Yes
If an analysis of browsing history is desired, please list the type of Internet sites you would like to have analyzed (e.g. pornographic or inappropriate web sites, web mail, hacker sites, auction sites, etc.) below. If specific sites are known, please list them:	Pornographic
Electronic Mail Analysis	
Do you wish to recover and analyze electronic mail folders or history? What type of e-mail program was used (if known):	Yes – any other than corporate e-mail
Do you wish to have a comprehensive analysis of e-mail, including decoding and analysis of e-mail attachments performed? If so, is there a particular emphasis	As needed

for this analysis?	
Are there any specific e-mail addresses that you would like to focus on? If so, please list them below:	As needed
File Recovery Options:	
Do you wish to recover specific files or documents from the hard drive? If yes, please list the exact file names (if known) or file types (e.g. DOC, XLS, PPT, etc.) below:	JPG, GIF, other images
Are there specific software packages that you wish to have analyzed? For example, peer to peer file sharing, online chat applications, database packages, or other specific applications? If so, please list them below:	As needed
General Analysis Parameters:	
Is this investigation taking place after a security incident such as a hack?	Yes, see above
Is there a known date period that you are specifically interested in?	11/27/07 – 11/28/07
Are firewall logs available to support the investigation? If so, what type are they?	Yes
Are mail server logs available to support the investigation? If so, what type are they?	Yes
Are proxy server logs available to support the investigation? If so, what type are they?	No
Are operating system logs available to support the investigation? (e.g. Windows Event Logs from servers, UNIX history logs, etc.)?	Yes
Are there any other comments or requirements? If so, please list them below:	
The organization wishes for the investigation to be performed while Mr. McCheney is on vacation. Access to the workstation will be provided for imaging, and will be stored in a locked office until either the investigation is completed or Mr. McCheney returns from vacation.	

11:45am Was accompanied by Mr. Nervosa to Mr. McCheney's office. Reviewed the workstation. The machine is turned off. It is a Dell Dimension 5100, with a serial number of "S/N:451231AFDQ". A corporate asset tag of "QWE9231" is affixed to the rear. A single CD-ROM drive and floppy drive are in the machine.

11:50 Turned on the workstation, and immediately hit the DEL key to enter BIOS setup. Noted that the current boot order is "floppy, hard drive, CD" Configured the machine boot order to "CD, floppy, hard drive." Noted that the system date and time in BIOS was within approximately 5 minutes of other clocks including my cell phone.

11:55 Booted the HELIX 1.9 boot CD. This CD was burned from the ISO image "Helix_V1.9-07-13a-2007.iso". Verified that the system could boot successfully and obtain a DHCP address. The IP address of Mr. McCheney's workstation is 192.168.233.2. Used the "dmesg" command to verify that a single IDE hard drive (/dev/hda) is configured in the system with a single partition (/dev/hda1)

12:05 Attached forensic workstation to the local area network and got an IP address. The forensic workstation's IP address is 192.168.233.3 Configured a netcat listener on the forensic workstation using the command:

nc -l -p 1234 > mccheney_hda.img

12:10 Began imaging Mr. McCheney's workstation across the network using the command:

dd if=/dev/hda bs=512 conv=sync,noerror | nc 192.168.233.3

12:15 Verified that imaging was working by watching file size grow on the forensic workstation.

17:45 Verified that the imaging was complete on Mr. McCheney's workstation (command had completed without error in Helix). Hit Control-C on forensic analysis workstation to stop the netcat listener.

18:45 Rebooted Mr. McCheney's workstation and immediately entered the BIOS. Changed boot settings back to original. Removed Helix boot CD from hard drive. Turned off the computer. Left the office and verified that the door was locked behind me.

11/29/07

07:17am Copied mccheney_hda.img to Linux forensic workstation

07:52am Mounted the image as a loopback device using '*mount -t ntfs -o loop,ro,noexec,offset=32256 ./casename_hda.img /mnt/data*' and verified that I could see the filesystem

08:15am Used '*mac-robber /mnt/data2 > bodyfile.txt*' to create a body file

Used '*mactime -b bodyfile.txt > timeline.txt*' to create a formatted timeline

08:20am Copied timeline.txt to a Windows forensic workstation and opened in Wordpad.

Note: There is Internet browsing activity under the "Richard McCheney" profile

Note: Around Fri Nov 23 2007 17:10:38 - Program Files/NewsReactor/NewsReactor.exe activity (this is apparently a Usenet browsing program)

Note: Around Mon Nov 26 2007 12:40:39 - Program Files/WinRAR/ (this is apparently the WinRAR archiving program)

11/30/07

08:27am Mounted the disk image on a Windows workstation using Mount Image Pro. Used the NetAnalysis Deleted History Extractor to scan the disk for Internet history files. These files were saved to a temporary Index.dat

09:15am Opened the newly-created index.dat file with NetAnalysis and manually reviewed Internet History.

Note: There appears to be general browsing activity to sites

such as nytimes.com, golf.com, homebrewheaven.com, michiganbrewing.com, etc. but not sign of impropriety

12/01/07

07:44am Created a foremost configuration file for Usenet items as follows for carving:

```
#
# Foremost configuration file
#
#-----------------------------------------------------------------------
# Usenet    (NOTE THIS FORMAT HAS A BUILTIN EXTRACTION FUNCTION)
#-----------------------------------------------------------------------
#
        yEnc    y       20000000  =ybegin =yend
        RAR     y       20000000  Rar!
        PAR2    y       20000000  \x50\x41\x52\x32\x00\x50\x4b\x54
#
```

07:56am Ran foremost with the following:

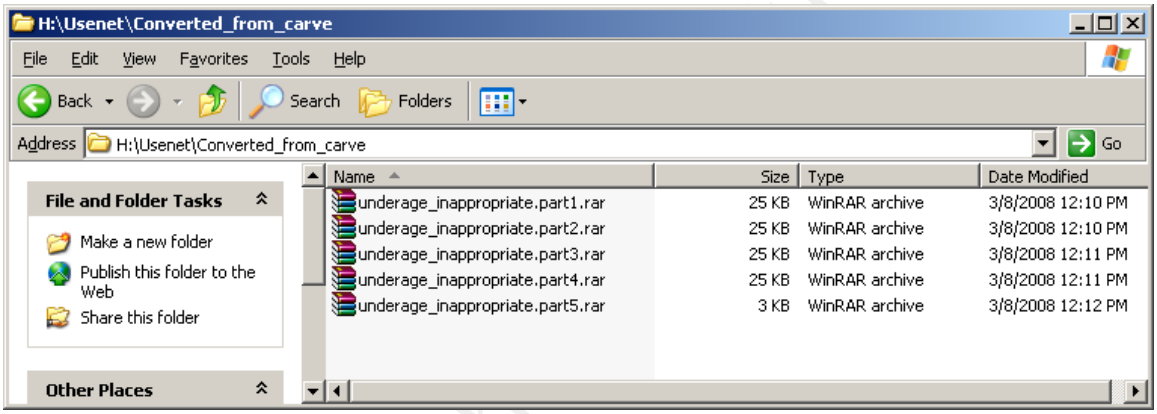
foremost -q -o /mnt/usb/Usenet/foremost_results -c ¥
/mnt/usb/Usenet/foremost_usenet.conf /mnt/usb/Usenet/ ¥
mccheney_hda.img

09:32am Changed to the foremost_results directory and validated findings. Reviewed audit.txt and see that a number of yEnc, RAR and PAR2 files were carved. Created MD5 sums of each of these files:

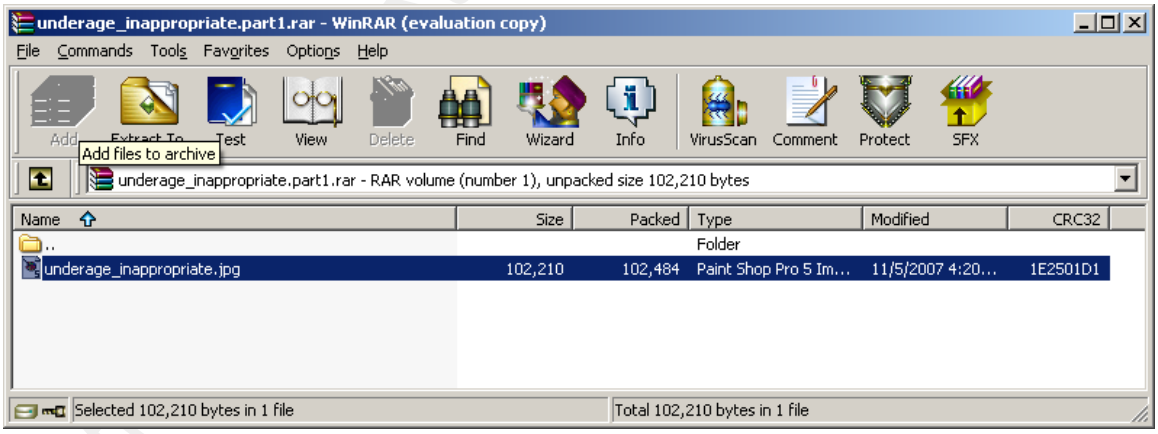
09:44am Attempted to recover the contents of the yEnc files

a. Opened yEnc32 on a Windows workstation

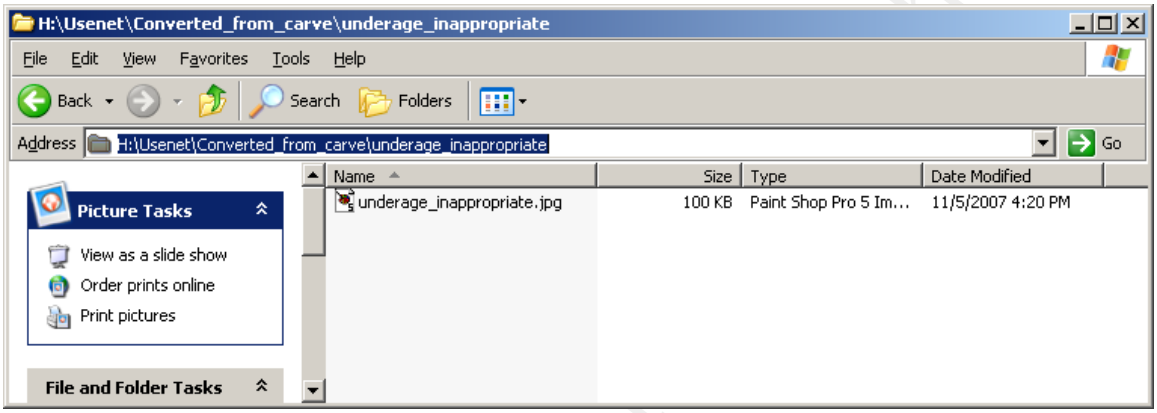
b. Loaded 00000299.yEnc and exported to H:¥Usenet¥Converted_from_carve (got an error "out of bounds" but a file was created anyway)
c. Loaded 00000403.yEnc, 00000507.yEnc, 00000611.yEnc, 00000671.yEnc with similar error message, but the files were created correctly:



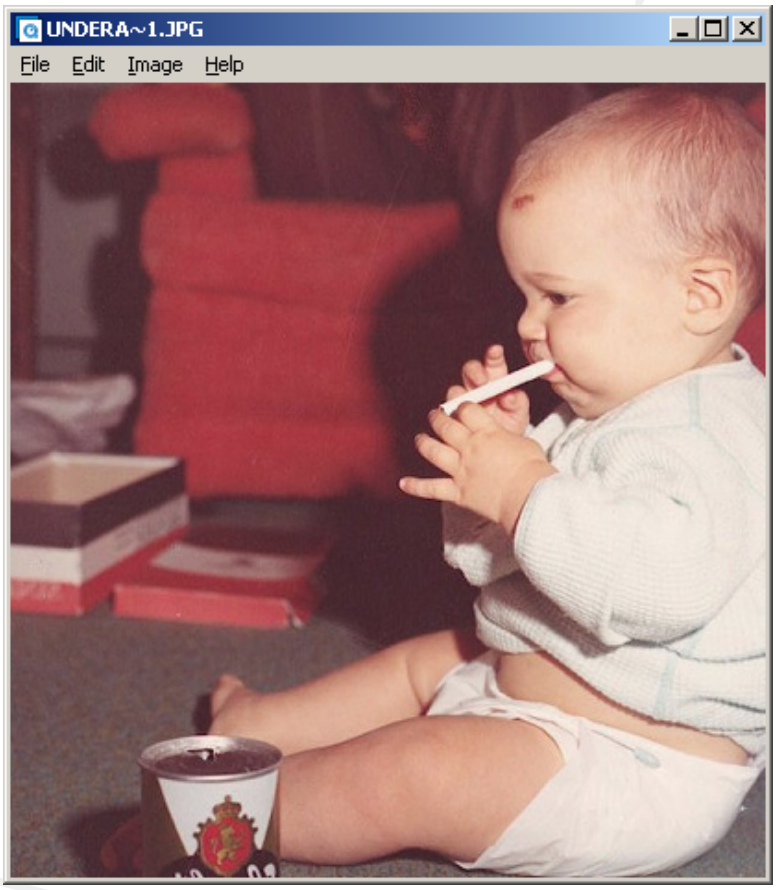
d. Loaded underage_inappropriate.part1.rar in Winrar:



e. Extracted underage_inappropriate.jpg to H:¥Usenet¥Converted_from_carve¥underage_inappropriate:



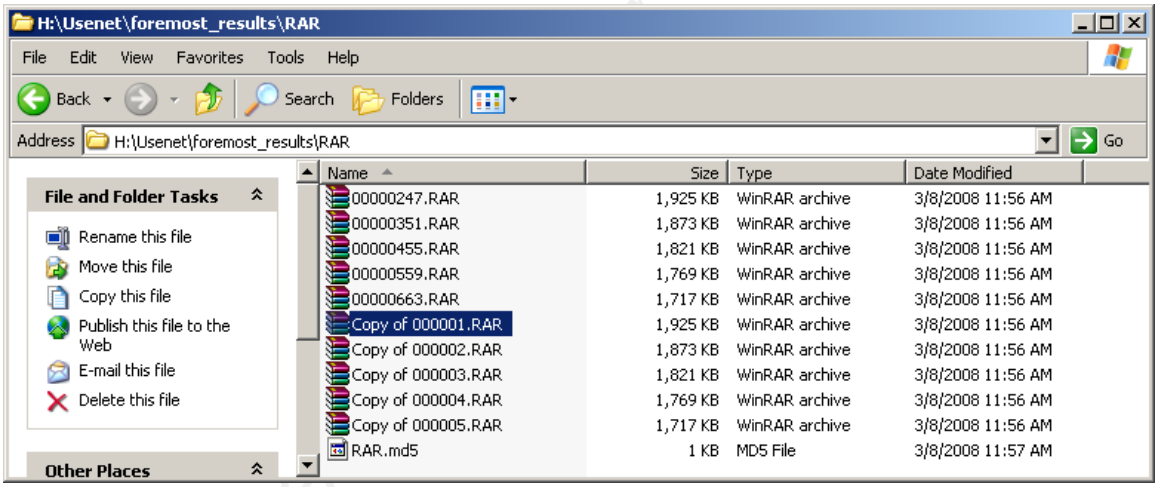
f. And opened the file in Microsoft Picture Viewer:



This may be the "inappropriate underage" image that Mr. McCheney was reportedly discussing.

01:36pm Attempted to analyze the contents of the RAR files:

Unfortunately, the RAR files will not open and allow recovery of their contents as are, probably because WinRAR expects them to have logical file names and be in the right order. The name that the file should have does not appear to exist within the carved file, so we will have to guess about the order. One logical guess would be to order them based on the order that they were created on the disk - hence we should number them from the earliest sectors forward:



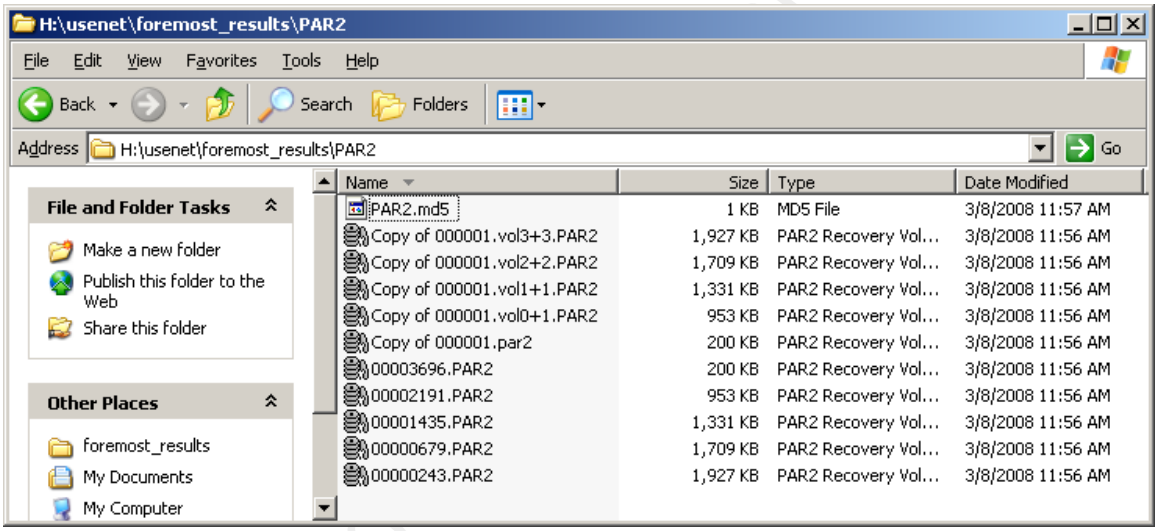
Hence in the above example "00000247.RAR" becomes "Copy of 000001.RAR", "00000351.RAR" becomes "Copy of 000002.RAR", "00000455.RAR" becomes "Copy of 000003.RAR", etc.

Opened the first RAR file and extract the image as above. We may have simply gotten lucky with our choice of filenames in this case, but it did work.

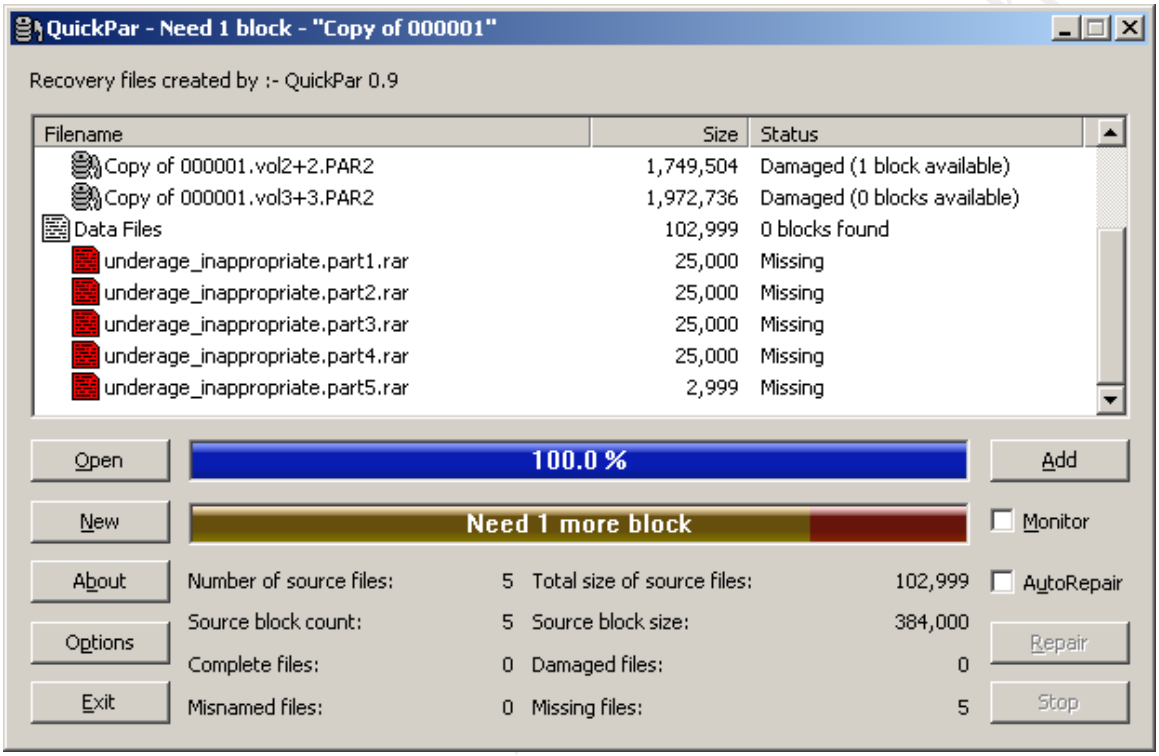
03:22pm Attempted to analyze the contents of the PAR2 files:

Attempted to rename the PAR2 files based on the type of file naming

suffixes used by Parchive. To determine how files were named, I manually created a set of PAR2 files using Parchive on a separate computer, and noted that usually there is a naming system that appends .vol0+1, vol1+1, etc. to the end of a file name. I then attempted to rename our carved files as I did with the RAR files where the earliest sectors get the first numbers, such:



Unfortunately, if you then load "Copy of 000001.par2" the process fails:



From these PAR files, we can now see that they were intended to repair a set of RAR archive files named underage_inappropriate.partX.rar. We might try looking inside fo the RAR files for file name information, and then try to rename them with the file names that the PAR files were expecting. Due to the fact that we appear to have successfully extracted the file these PAR2 parity archives were intended for, we will hold off on further analysis of this at this time, but the customer will be made aware that further evidence could possible be recovered with more research, if necessary.

12/02/07

08:22am Did a carve using the full foremost.conf file for all known image

types and archives. This will be time consuming, but may find other images. Invoked with the following syntax:

foremost -q -o /mnt/usb/Usenet/foremost_results -c ¥
/mnt/usb/Usenet/foremos.conf /mnt/usb/Usenet/mccheney_hda.img

11:43pm Loaded the results of the carve onto an external USB hard drive and manually browsed through the results in a thumbnail view. The previously noted image was found, as well as a large number of images from various programs and Internet web sites, but no obvious pornography or inappropriate content was discovered.

12/03/07

08:52am Began collecting accumulated data and working on deliverable document

05:44pm Draft deliverable document done. Submitting for peer review to second analyst

12/05/07

08:02am Peer review completed. Finishing deliverable, printing and binding

12/07/07

01:30pm Conducted deliverable meeting with customer. Discussed all findings and shortcomings of the project (including the fact that we did not recover all of the PAR2 files). Customer stated that they were comfortable with the results and did not wish to pursue the case further. Customer stated that they did not feel the need to take this to law enforcement.

04:30pm Case complete. Copy work product to a protected directory for future reference.

15. Conclusions

As should be obvious from even this high-level treatment of the subject, investigating Usenet abuse is technically difficult and time consuming. However, due to ease of use and relative anonymity (compared to peer-to-peer file sharing, for example) the questionable use of the NNTP protocol and Usenet service providers by tech-savvy individuals is only likely to increase over time. This problem is further compounded by the questionable marketing of "private" and "secure" international NNTP access by the providers themselves, who may argue that they are exempted by "safe harbor" provisions in their jurisdictions. In the opinion of this author, current forensic tools and education sources do not adequately deal with the issue of Usenet abuse. What is clearly needed is further awareness of the issue, perhaps through professional organizations such as the High Tech Crime Investigation Association (HTCIA), as well as improved commercial and open source tools to facilitate analysis. This author welcomes comments and corrections on this document, and can be reached at mark_at_lachniet_dot_com.

16. Note on the public flash image

As was noted in *Section 14* of this document, a small loopback image containing the RAR, PAR2, yEnc and JPG files used in the fictitious analysts note was created and is available at http://lachniet.com/flash.img as of March 8th, 2008. This flash image is provided so that forensic analysts can validate the findings of this paper, as well as their own carving tools. The following details how this flash image was created:

1) Copied the underage_inappropriate.jpg file to a temp directory

2) Create a RAR file with a maximum size of 25k per part, creating 5 RAR files

3) Use yEnc32 to turn each of these RAR files into a yEncode .ntx file

4) Used Parchive to create PAR files of the 5 RAR files (turned into 4 PAR files)

5) Note: In this approach the yEncoded files don't include all the levels of nesting (it is NOT a par with a rar with a image, the yEncoded files are RAR files).

6) On a Linux box created a blank loopback file with:

a. *dd if=/dev/zero of=/mnt/usb/Usenet/flash.img count=4096*

b. *mkfs.msdos /mnt/usb/Usenet/flash.img*

c. *mount /mnt/usb/Usenet/flash.img /mnt/test*

7) Copied all the files to /mnt/test

8) Created a readme file

9) Deleted all but the readme file

10) Unmounted the file

17. References

[1] Butler, S. (2007). *Labels Sue Usenet Service*. Retrieved February 26th, 2008, from http://www.billboard.biz/bbbiz/content_display/industry/e3i66abf6954df1d43fbdf1692e0860d269

[2] Anonymous (2008) *Usenet*. Retrieved February 26th, 2008 from http://en.wikipedia.org/wiki/Usenet

[3] Kantor, B., Lapsley, P (1986) *Network Working Group Request For Comments: 977*. Retrieved February 26th, 2008 from http://www.ietf.org/rfc/rfc0977.txt

[4] Horton, M., Adams, R. (1987) *Standard for Interchange of USENET Messages*. Retrieved February 26th, 2008 from http://www.w3.org/Protocols/rfc1036/rfc1036.html.

[5] McCandles, D. (2005). *Want the Sith DVD? Go to Usenet*. Retrieved February 26th, 2008 from http://www.wired.com/entertainment/music/news/2005/05/67588

[6] Anonymous (2001). *What is yEnc?* Retrieved February 26th, 2008 from http://www.yenc.org/whatis.htm

[7] Anonymous (2008). *WinRAR Arvhiver Home Page*. Retrieved February 26th, 2008 from http://www.rarlab.com/

[8] Anonymous (2008). *QuickPar – About PAR2*. Retrieved February 26th, 2008 from http://www.quickpar.org.uk/AboutPAR2.htm

[9] Roshall, E. (2008) *RAR version 2.02 - Technical information*. Retrieved February 26th, 2008 from http://www.wotsit.org/download.asp?f=rar202&sc=257343632

[10] Anonymous (2008) *Parity Volume Set Specification 2.0*. Retrieved February 26th, 2008 from http://parchive.sourceforge.net/docs/specifications/parity-volume-spec/article-spec.html

[11] Anonymous (2008) *NZB*. Retrieved February 26th, 2008 from

http://en.wikipedia.org/wiki/NZB