

Tracking data over Bit Torrent

Dan Morrill, MSEC IAM

February 2006

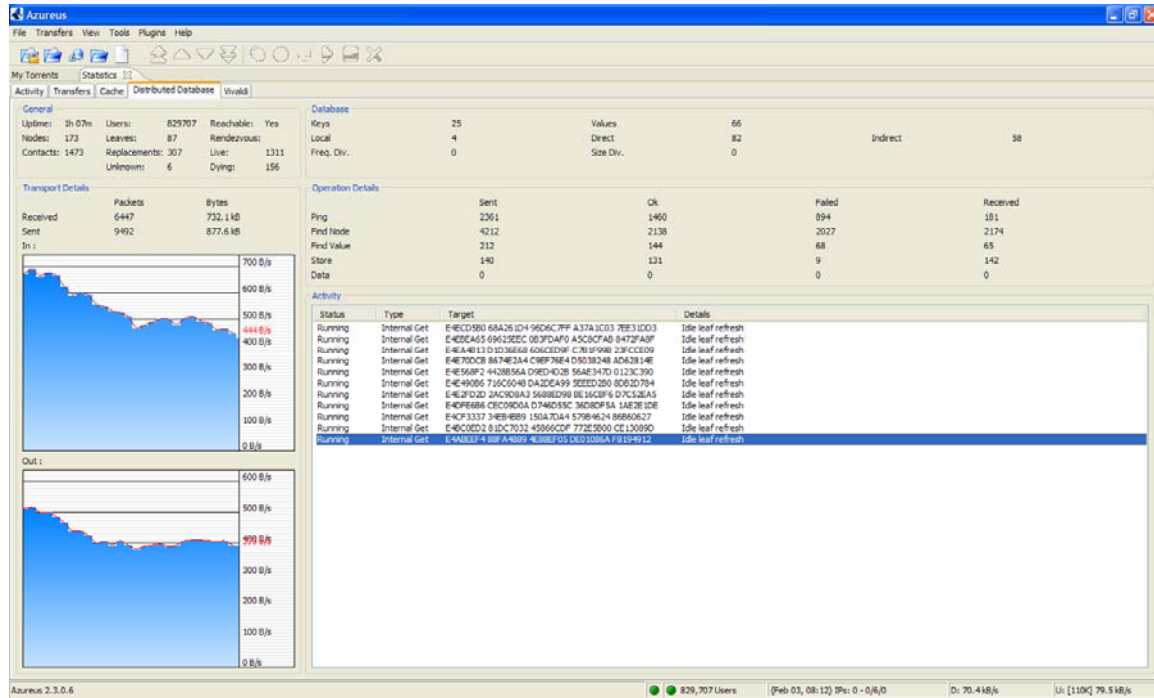
Bit Torrent has a reputation of being difficult to find out who is downloading movies, games, documentation, and other information. This is not necessarily true in all cases; any Peer-to-Peer system at some point relies on IPv4 and TCP/IP to make its connections. Because of that, the sender and the receiver can be well known to anyone who is using a program or programs that have robust logging, and other programs that help geolocate where those IP addresses are physically located. Anyone who produces or protects data that is confidential or otherwise protected by statute or law should have an understanding of bit torrent networks, how they work, and how they route. Bit torrent is a reliable and efficient way of sending very large data chunks from point A to a distributed points B. It was originally developed to move Linux Distributions from various points without choking the sender and spreading the download across multiple downloader's and senders of the same data point.

While the original intent of the system was to reduce the costs of shipping very large files across the internet, it is also an excellent tool for sharing other data types. Movies, games, and entire sound tracks are common fare across the bit torrent networks. Finding out if the company's data is on the bit torrent networks, and then being able to figure out where the data is going is an important part of any tracking and monitoring program that a company has. Once the data is in Bit Torrent, usually the only action that a company can take is to issue a take down notice, and this will work in the United States. However, in other countries, with other Intellectual Property Protection laws, take down notices may or may not work. This paper is not about sending take down notices, but discovering whom by IP and geographic location is downloading a particular file or files across the Bit Torrent networks. Your company's legal counsel can advise the company as to what actions are best for that company to take.

Azureus

Azureus is a free java based Bit Torrent client with an excellent interface, good log files, and an excellent statistics generating tool set. As well, Azureus has a very good logging system that can be used for later analysis by writing a custom parser for the lines, and then using a database to collate and show trends in the data. The log files if signed and hashed can be used in total as a way of showing intent if the company or organization is determined to go to court over this issue. Azureus has a good number of plug ins that also allow for visualization of the data that is traversing between various points on the network. Using Azureus

stats can help find out what is happening on the network in regards to your data. To get to Azureus stats, click on Tools, Statistics. The screen below shows the Distributed database screen. The data will normally look like this:



The highlighted line shows a code set (randomized) that represents the hash value of an IP address on the internet. The Azureus log files then provide the IP address that can correspond to that hash value as shown below

```
[8:56:28] Obtaining external address
[8:56:28]   Contacts to search = 0
[8:56:28] Initial external address: /66.15.68.234:6881
[8:56:29] Imported contact 3D289978...[dht.aelitis.com/85.31.105.2:6881,V12]
[8:56:29] Imported contact F9EA0294...[/24.7.59.181:7000,V8]
[8:56:29] Imported contact F420770A...[/24.73.27.106:52231,V12]
[8:56:29] Imported contact F4DBB09F...[/24.188.222.5:6881,V12]
[8:56:29] Imported contact E53FA168...[/219.161.211.62:65535,V12]
[8:56:29] Imported contact F42F8999...[/81.235.144.165:6881,V8]
```

```
[9:57:27] DHT:ip=/66.15.68.234:6881,net=0,prot=V12,reach=true
[9:57:27]
router:nodes=165,leaves=83,contacts=1382,replacement=282,live=1253,unknown=5,failing=124
[9:57:27]
Transport:ping:1950,1247,674,213,store:210,202,8,103,node:2890,1470,1366,2052,value:95,61,34,75,stats:0,0,0,0,data:24,0,0,24,incoming:2470,alien:1981,72,2
```

19,0,91,packsent:7615,packrecv:5454,bytesent:775566,byterecv:559707,timeout:0,sendq:0,recvq:0
 [9:57:27] Control:dht=870536,
 Database:keys=29,vals=42,loc=0,dir=46,ind=109,div_f=0,div_s=0

For each file downloaded, Azureus also has a plug in module that allows for partial country to IP address resolution as shown below. To get this module, go to http://azureus.sourceforge.net/plugin_list.php

CC	Flag	IP	Client	T	Pieces	%	Down S...	Up Speed	State
gb		212.84.121.184	Azureus 2.3.0.6	L		100.0%	0 B/s	0 B/s	Fully established
pl		193.0.121.66	Azureus 2.3.0.6	L		100.0%	0 B/s	0 B/s	Fully established
fr		86.195.167.145	BitComet 0.60	R		100.0%	0 B/s	0 B/s	Fully established
us		24.183.2.98	Mainline 4.2.2	L		97.3%	0 B/s	0 B/s	Fully established
us		68.239.181.233	Azureus 2.3.0.6	R		83.9%	4.6 kB/s	6.4 kB/s	Fully established
gb		81.109.224.64	Azureus 2.3.0.6	L		81.3%	72 B/s	0 B/s	Fully established
gb		86.130.130.148	BitComet 0.60	R		70.7%	409 B/s	0 B/s	Fully established
gb		82.28.237.240	Azureus 2.3.0.6	L		67.7%	667 B/s	0 B/s	Fully established
dk		83.91.112.227	BitLord 1.1	R		59.6%	0 B/s	0 B/s	Fully established
gb		88.111.16.204	Azureus 2.3.0.6	L		59.6%	2 B/s	0 B/s	Fully established
ca		24.70.190.192	Azureus 2.3.0.6	L		53.2%	0 B/s	0 B/s	Fully established
yu		212.200.207.115	BitLord 1.1	L		45.6%	0 B/s	0 B/s	Fully established
ca		216.36.145.142	Shadow 5.8.11	L		42.2%	3.2 kB/s	0 B/s	Fully established
us		138.88.143.136	Mainline 4.2.2	R		35.2%	3.0 kB/s	3.2 kB/s	Fully established
gb		212.159.19.44	BitLord 1.1	R		34.0%	0 B/s	0 B/s	Fully established
ca		142.59.210.149	µTorrent 1.4.0	R		32.6%	0 B/s	0 B/s	Fully established
ae		217.164.65.206	Shareaza 2.2.1.0	R		29.3%	0 B/s	0 B/s	Fully established
gb		80.177.208.85	Mainline 4.4.0	R		26.7%	3.4 kB/s	3.2 kB/s	Fully established
gb		82.44.75.133	Azureus 2.3.0.6	R		23.9%	0 B/s	0 B/s	Fully established
ca		70.48.62.231	BitTornado 0.3.10	L		13.8%	4.7 kB/s	25.6 kB/s	Fully established
mx		200.56.182.62	BitComet 0.60	L		10.9%	0 B/s	0 B/s	Fully established
us		70.92.238.131	Azureus 2.3.0.6	L		9.6%	0 B/s	0 B/s	Fully established
ca		70.48.210.215	BitComet 0.61	L		6.0%	0 B/s	0 B/s	Fully established
se		213.113.220.82	Azureus 2.3.0.6	L		3.7%	0 B/s	0 B/s	Fully established
us		71.193.152.190	Mainline 4.2.0	R		2.4%	1.9 kB/s	0 B/s	Fully established

This allows the downloader to find out what is happening on the Bit Torrent networks, but does not provide an adequate map of geo-location that can be used for management presentations. A program called Geo Spider from oreware.com (<http://www.oreware.com>) can be put on the same box that is running Azureus (or any other P2P client) to track by Geo location and IP address where people are coming from. In our test file, the geo location shows that USA and Europe were the most popular download points.

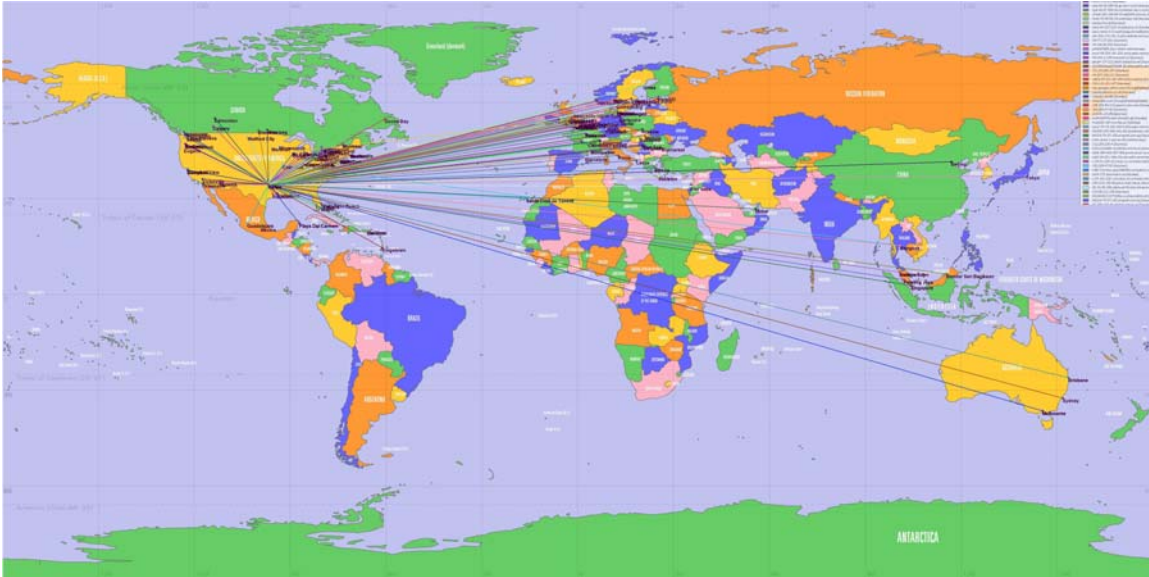


Figure 1 - Geo Spider graphic showing who is downloading from the test system

The combination of Azureus Log files, tracking mechanisms, and geo mapping of who is downloading the file can give an in-depth view of what is happening with any particular file across the Bit Torrent Network. This can help aid in tracking down who is downloading the file and allow by ISP to find out who has downloaded the file. Azureus in its logs keeps an IP track of everyone who is attaching to the file that is being downloaded as shown below. These entries can be stripped using a simple regex (regular expression), then dumped into a database for automated Whois information at the ISP or company level.

```
[13:26:49] {0:11:0} Sent [BT_PIECE data for #1817: 1146880->1163263]
message to R: 83.91.112.227: 3666 [BitLord 1.1]
[13:26:49] {0:11:0} Sent [BT_PIECE data for #1817: 1163264->1179647]
message to R: 83.91.112.227: 3666 [BitLord 1.1]
```

Trackers

Bit Torrent relies on trackers to determine where the file can be located, and who all is sharing the file. Many tracker web sites keep statistics on what they are sharing. For example, Mininova does a top 10 list of the most downloaded files from their systems viewpoint. Trackers can provide a lot of information about who is downloading files, what are the popular files being downloaded, or how often a file has been downloaded.

10 Most downloaded torrents (02/03/2006)

Downloads	Name	Size	Seeds	Leechers
873,935	Doctor Who 2005 1x12 Bad Wolf WS PDTV XviD-FoV [eztv efnat]	350.83 MB	2	1
332,538	Top Gear - [07x05] - 2005 12 11 avi	348.71 MB	180	130

284,884	Stargate Atlantis S02E15 HDTV XviD-TvD [eztv]	349.63 MB	45	57
261,422	Top Gear - [07x06] - 2005 12 27 avi	348.37 MB	201	97
257,762	CSI Miami S04E11 HDTV XviD-LOL [eztv]	349.17 MB	11	21
199,361	The Simpsons S17E09 PDTV XviD-LOL [eztv]	174.69 MB	35	26
193,563	UK - Fifth Gear - [08x09] - 2005 12 05 avi	349.37 MB	27	27
176,763	Top Gear - [07x04] - 2005 12 04 avi	348.64 MB	132	132
173,359	Family Guy S05E08 PDTV XviD-LOL [eztv]	171.12 MB	14	12
170,653	The Daily Show 07 28 05 Maggie Gyllenhaal DSR XviD-STFU [eztv]	173.03 MB	0	1

Trackers expose a large amount of data about themselves, and their users. The development of private trackers and the use of private trackers (that only allow members who have signed up) like black cats games is an interesting response to the issues of being caught. For every process that allows someone to quantify the data on the Bit Torrent networks, there is usually a countermeasure.

When tracking downloads and other data via BT, even if the person is using the TOR network, or other protection mechanisms, there is an initial handshake between the clients that can be monitored, and then mapped or otherwise used to find out more information about the user. While downloader's do use some protection mechanisms, the methods are not fool proof when dealing with and IPv4 network. As well, some well meaning web sites put up top 10 uploaders, and top 10 downloader's. Because of this kind of information sharing, it is usually painless to find out who is downloading what, and by what geographic distribution downloads are happening.

Downloader's Protection mechanisms

Downloader's have a number of protection mechanisms that they can use to ensure that they are not being tracked by known IP address ranges, these systems are Protowall and Block List manager.

Warning and Caveat, these tools are difficult to remove once they are installed, they are also very difficult to install, and might require a complete wipe and reload of the system that they are installed in. While the developers of this software have made every effort to make them easy to use, this is not always the case. In the case of Block List Manager, if you use an IP address that is in their block list, no traffic will originate from your computer requiring the user to explicitly allow that IP address range in Block List Manager. There are always inherent risks when installing any software on any computer. The installer assumes all liability and responsibility for installing software on there systems.

Nor is this construed to mean that these products are endorsed or suitable for use by anyone.

Protowall is a small driver level firewall provided by Blue tack at <http://www.bluetack.co.uk>. If you choose to visit this site, and you are coming from an IP address that has been designated as block able, then the user will get the following screen, such as the demo screen showing that Google is not allowed access to their web site.

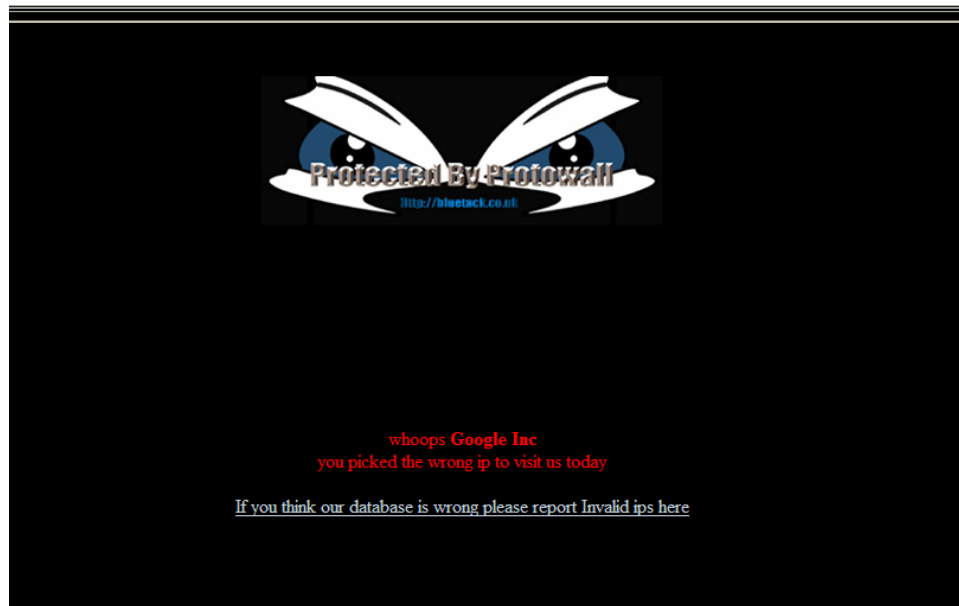


Figure 2 - Protowall demonstration block from Google cache

Protowall works as a low-level driver at the NIC (Network Interface Card) level looking at packet headers, extracting the IP address, and comparing it against a list managed and maintained by another program called Block List Manager (BLM). It is not a firewall; rather it simply blocks IP addresses that are on a known list. Most of the Anti-peer to peer companies (AP2P) is aware of this protection mechanism, so it is important for the person who is looking for data not to be coming from an IP address that has been identified as AP2P. Usually a commercial DSL or Broadband cable line is not blocked unless the downloader gets to aggressive in downloading the file. Then a program like Azureus will internally ban the IP address independent of Protowall or BLM.

Block List Manager

Block List Manager is a program that maintains a list of all IP addresses that are either government, AP2P, corporate, or IP addresses of suspected or known fake seeders or providers of corrupt data as shown below.

ED2K Corrupt Data Senders:62.0.175.170-62.0.175.170
ED2K Corrupt Data Senders:62.0.188.154-62.0.188.154
ED2K Corrupt Data Senders:62.0.191.104-62.0.191.104

Fake BitTorrent Seeders:64.168.30.40-64.168.30.40
Fake BitTorrent Seeders:64.217.229.102-64.217.229.102
Fake BitTorrent Seeders:65.49.132.215-65.49.132.215
Fake BitTorrent Seeders:66.68.84.184-66.68.84.184

BLM works with this list in conjunction with Protowall to ensure that the downloader is getting data from non-corporate, government, or AP2P companies. This process makes it easier to track down people who are downloading information, when the searcher is coming from a corporate, government or AP2P company. This is a protection mechanism for heavy use traders on the Peer-to-Peer (P2P) networks. The GUI for BLM is easy to use, and can be automated so that automatic downloading of new updated lists and then exported to your favorite P2P tool. This allows the file trader to not connect, or otherwise reject in conjunction with Protowall any IP address that is in this list. Using a DSL line or Broad Band line to circumnavigate BLM while searching for and downloading data is a best response for people who are downloading and are worried about being caught doing so. BLM coordinates a number of lists, and then exports a common list for your system. The sources are:

- Antip2p list
- cDonkey Blocklist
- Morpheus Blocklist
- IANA reserved ranges
- Gov and Mil IP ranges
- Ads spammers and bad pr0n
- Research companies
- Edu ranges
- For File Sharing apps only (takes out the splits to websites)
- webspiders and bots
- MDawson IP DB
- range testing
- trojan list by redzulu2003

The list output is very comprehensive, and in conjunction with Protowall very effective at identifying AP2P, and other address strings that are advertisers, spammers, and other people on the internet.

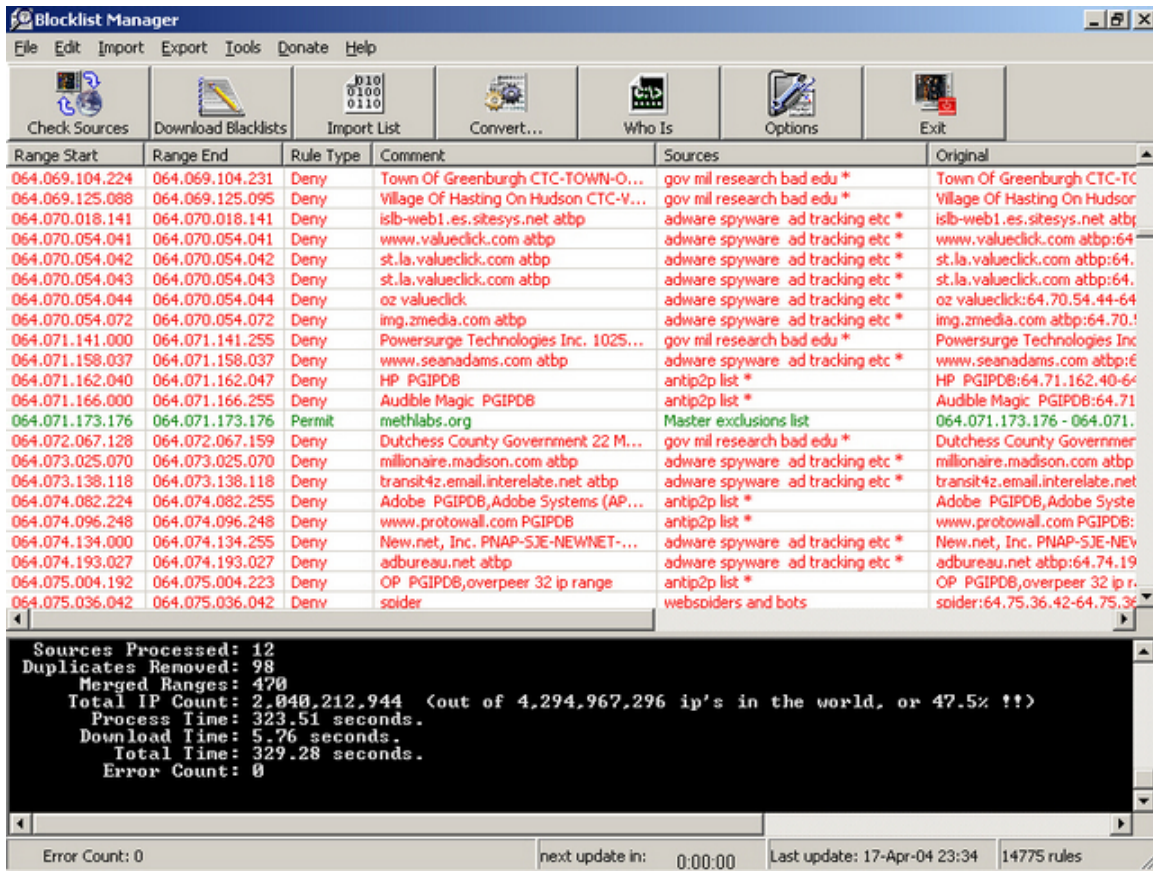


Figure 3 - BLM management Screen

Peer to Peer other programs

Geo Spider is flexible enough to work with any program that accesses the internet via TCP/IP. So using Kazza, Shareazza, E-donkey, or others will not affect the ability of the geo location software from finding out where geographically people are located. However, local logging is another issue. Many of the standard peer-to-peer clients have a logging function, but the function is usually not good enough to be used as evidence, or shows the function of network navigation without a lot of noise. E-Donkey however has a robust logging system that is not overly noisy that can be used for forensic and information purposes down the road. It is much the same as Azureus logs in terms of how it can be parsed and used.

Summary

Using commonly available tools, fewer than 100 dollars, and writing custom parsers it is cost effective and inexpensive to build your own suit of software that is capable of tracking a file across the internet to see who is downloading it. Although people on the peer-to-peer networks may use tools to obfuscate or otherwise keep people from seeing their data, the Achilles heel of the Bit Torrent networks (and indeed others) is that at some point they have to use IPv4 to get

on the internet. These connections can be recorded in log files for later parsing or analysis, or an in line program like Geo Spider from oreware can be used at the application layer to monitor all the connections going in and out of the system. These software packages can be used to build a profile of the data that is being transferred, by what country, and using a database and automated Whois routines, allow for by city, by ISP information. The use of a DSL or Broadband cable system is recommended so that people who have software such as Protowall and Block List Manager will not know you from someone else. While there are some lists that look for people who are deliberately trying to modify the data in the stream, these lists are not part of the full BLM IP load, until proven that they are sending corrupted data.

References:

<http://www.lickmytaint.com/>

<http://www.mininova.org>

<http://www.bluetack.co.uk>

http://azureus.sourceforge.net/plugin_list.php

<http://azureus.sourceforge.net>

<http://www.oreware.com>