May 24th, 10:30 AM

Forensic Analysis of Ares Galaxy Peer-To-Peer Network

Frank Kolenbrander
Politieacademie, The Netherlands, frank.kolenbrander@politieacademie.nl

Nhien-An Le-Khac
School of Computer Science & Informatics, University College Dublin, Ireland, an.lekhac@ucd.ie

Tahar Kechadi
Centre for Cyber Crime Investigation, University College Dublin, Ireland, tahar.kechadi@ucd.ie

FORENSIC ANALYSIS OF ARES GALAXY PEER-TO-PEER NETWORK

Frank Kolenbrander
Politieacademie
7334 AC Apeldoorn,The Netherlands
Frank.Kolenbrander@politieacademie.nl
Nhien-An Le-Khac, M-Tahar Kechadi,
University College Dublin
Dublin 4, Ireland
{an.lekhac, tahar.kechadi}@ucd.ie

ABSTRACT

Child Abuse Material (CAM) is widely available on P2P networks. Over the last decade several tools were made for 24/7 monitoring of peer-to-peer (P2P) networks to discover suspects that use these networks for downloading and distribution of CAM. For some countries the amount of cases generated by these tools is so great that Law Enforcement (LE) just cannot handle them all. This is not only leading to backlogs and prioritizing of cases but also leading to discussions about the possibility of disrupting these networks and sending warning messages to potential CAM offenders. Recently, investigators are reporting that they are creating more serious cases on Ares Galaxy (Ares) than on other open P2P networks. Little has been done on automatic prioritization of cases with the information obtained from data that is available on P2P networks. Cases are mostly selected based on the highest number of CAM, while studies indicate that the abusers are most likely to be found not within that top user list. What kind of information can we use to prioritize cases in another way? Is it possible to disturb the network by sending warning messages and sharing fake material? Although the past years have seen a lot of successful CAM cases, generated in several countries, there is still little known about the Ares network. Although Ares network is open source, the protocol is not documented and the program does not come with serious documentation or support. In this paper, we present first of all a forensic analysis of using of Ares network in relation with the distribution of CAM. We then describe forensic artifacts found on an Ares computer involved in CAM.

Keywords: P2P network forensics, Ares Galaxy network, Child Abuse Material, forensic artifacts, registry dencryption

1. INTRODUCTION

Ares started as a P2P [1] Gnutella client in 2002, but after 6 months [2] they started with their own new network. This Ares network used a system where a client (*clientnode*) in normal user mode could be promoted to a client that also acts as a server called *supernode*, the same methodology and name that was used by Fasttrack [3], a closed source P2P network. A server (*supernode*) keeps records of connected *clientnodes* and their shared files and this information is used by the *supernodes* for handling searches of connected

nodes. Systems like this are called decentralized networks [4] because the servers (*supernodes*) are not owned by a centralized company or person. This makes it harder to hold them responsible for the contents that are being shared on the network, unlike a centralized system, like Napster, that was easily held responsible and forbidden [5].

Even though the Ares network does not keep statistics it is reasonable to assume that the network consists of hundreds of Ares *supernodes* and a couple of million Ares *clientnodes*. The *clientnode* hold a list of available *supernodes* and connects to a maximum of five of them. Each *supernode* can handle a maximum of 400 *clientnodes*.

Ares is used all over the world for the exchange of all kinds of files, including copyrighted material, like music and movies. Ares can be moreover used for accessing existing chatrooms and users can start their own public or even private chatrooms. These chatrooms look like the traditional IRC channels.

In order to make it more difficult for users to be tracked by for example the music and movie industry, countermeasures are used in Ares network. In fact, this network does not provide statistics about the number of users and the numbers of their shared files. Searching is automatically limited by the system by countermeasures in both the normal user mode and the *supernode* mode. On the used computer, users can easily wipe the history of previous searches and information about shared files is stored in encrypted data-files. Other information such as the last time connected, first time used, time online are encrypted before they are saved into the registry.

For a very long time, Ares was not disturbed by fake files, over the past few years this has changed, but the developers wrote detecting modules and succeeded in limiting the effects. The software is probably not owned by a company but maintained by private developers and therefore more difficult to approach by investigators. This all makes Ares a network that is very suitable for people that want to exchange CAM. On the discussion page at the SourceForge community [6] a developer offered his assistance to block CAM on Ares but it doesn't look like his offer was accepted. One of the forum members responded: *"Ares uses a decentralized protocol for it is searches and downloads and developers can't be held responsible for what users share on the network."*

As mentioned above, Ares does not support statistics, which makes is difficult to estimate how many users involved in CAM, how many CAM files are shared on this network. On top of this, there are a lot of clients on Ares that share fake files, files from which the real content has nothing to do with the indication that the filename, title and other meta-information would suggest. Some of them are just files with other content shared with deliberately wrong filenames, title and other meta- information. Sometimes just as a joke or maybe by mistake but also often likely to discourage the exchange of copyrighted or illegal material. There are also clients that deliberately respond to searches, by reporting that the files are found, and then generate fake files on the fly with the search phrase in a part of the filename. A lot of these generated files are zip files, containing .wmv (windows media file). These .wmv files redirect the user to websites that spread malware. The author has also seen generated fake files containing advertisements, for example for ringtones.

In this paper, we present a forensic analysis of using Ares network world-wide in relation with the distribution of CAM. We also moreover contribute a comprehensive description of traces on a computer running

Ares. The rest of this paper is structured as follows: Section 2 shows the background of this research including related work in this domain. We present the analysis the use of Ares network in relation with CAM in Section 3. We describe and discuss on experiment results of analyzing the forensic artifacts in the computer of using Ares in Section 4. We conclude and discuss on future work in Section 5.

2. RELATED WORK

In literature, there are very few programs dealing with forensic artifacts in a computer running Ares. There is moreover almost no information on how to process these artifacts with other alternative programs for justifying the evidence.

On the other hand, several studies have been performed over the past decade on the relationship between CAM and active sex offenders as well as the distribution of CAM on P2P networks. However, there is still very little research on this subject when it comes to the Ares Network. In [7], they describe that the proportion of Child Pornography Possession (CP) arrestees who were identified as child molesters in cases that started out as CP investigations dropped from 16% in 2006 to 10% in 2009, while there was some surprisingly increase in the proportion of arrested CP possessors that have images depicting young children and sexual violence. They suggest that this maybe caused because some in law enforcement may be targeting those with more extreme images in the belief that these offenders are also more likely to be molesters. Unfortunately they cannot speak confidently to the issue of exactly how law enforcement should prioritize CP investigations if they want to catch active molesters. They say that this does not mean that those priorities are inappropriate for the goal of protecting children since child pornography possession has its own corrosive dynamics. But it should spur the search for additional data to evaluate the best police strategies for protecting children.

Authors in [8] concluded "Noncontact offenders anchored on lower-SAP-level indecent images of children (IIOC), with no preference in terms of the age, gender, or sexual action. In contrast, dual offenders preferred higher SAP (***Sentencing Advisory Panel***) levels and also possessed IIOC within a smaller age range, which tended to match their sexual contact victim in terms of age and gender. Moreover, the more severe the contact child sexual offense committed, the higher the proportion of penetrative IIOC possessed."

In [9], authors describe that systematically gathered and analyzed data on P2P can give an idea of the scope and characteristics of CAM on P2P networks and that such measurement can be used to combat the problem. They also mentioned that investigative tools can be used to help law enforcement with prioritizing cases. In [10], authors report that their results suggest a homology between Internet behaviors, indecent images of children (IIOC) possession and victim selection.

3. ANALYSING THE USE OF ARES NETWORK IN RELATION WITH CAM

In this section we describe first of all the worldwide use of Ares network in relation with the distribution and downloading of CAM. This was done by gathering data with a search and monitoring tool based on the original program. This information is used for the statistics by country and language. In order to analyze Ares network, we used two different Ares clients to gather information. For the most significant part of the examinations, the

original latest Ares Galaxy Sourceforge client was used. It was used to gather information about traces on the computer and traces and statistics on the Ares network. In addition an alternative client KCeasy [11] and the under-laying giFT protocol [12] was used to gather traces and statistics on the Ares network. Although the giFT protocol itself is documented [13], giFT has no documentation about the under-laying Ares protocol.

For traces on the computer the gathered data and information was checked by reading the open source and checking results with a simple hex editor and calculator. We also used AresDecrypter [14] based on the information found in the open source. AresDecrypter was validated and approved for use by the FBI's Computer Analysis Response Team. Furthermore, Encase 7 and 2 special enscripts (Ares Dat File Decryptor (V1.3.1).EnPack and Ares Registry Report (V1.1.0).enpack) [15] were used to compare the results. We also modified the original client and forced it to write information into *logfiles*. This was done for incoming and outgoing data on the network and read and written data to the operating system (registry and files). In case data was being encrypted or decrypted the data was written encrypted and unencrypted to these *logfiles*. This information was compared with network data captured with Wireshark [16] and the information that was retrieved with the mentioned tools above.

3.1 Statistics by country

Statistical information was gathered by searching for CAM related keywords and known CAM hashes. This was done worldwide with the use of the Roundup Investigative Tool over a period of 4 months, from September 2014 until January 2015. Only results for files with hashes that were already identified were gathered. The number for (known hashes) the database of known hashes, was approximately 4,1 million. This resulted in 3,568,462 unique IP-addresses worldwide. To exclude possible "accidental downloads," IP-addresses with less than 3 results were dropped, which resulted in 1,553,222 unique IP-addresses. The IP-addresses were matched to a country with the free GeoIP table from Maxmind.com[17] and with information from several sources on the internet found with Google the countries were matched with their spoken languages. The percentage of the country population and the percentage of the (IP population) total number of Internet users for the country, was calculated with information from the CIA factbook [18]. The CIA factbook information for country population was from 2014 and the IP population was from 2009. With this measurement we have to take into account that countries with high numbers of dynamic IP-addresses will score relatively higher than countries with high numbers of static IP-addresses (Figure 1 and Figure 2). We also have to take into account that the searches were only done in the English language and the known hashes used were mainly collected in investigations situated in the United States.

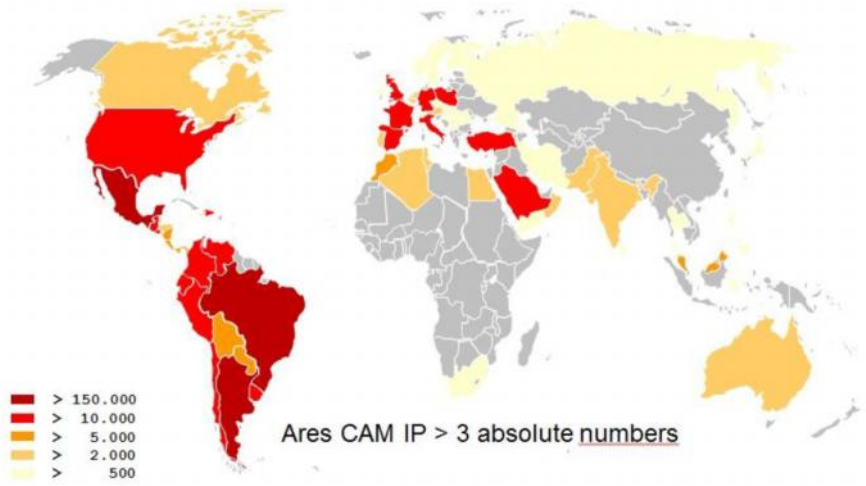


Figure 1. Absolute number of IP –address sharing at least 3 CAM files

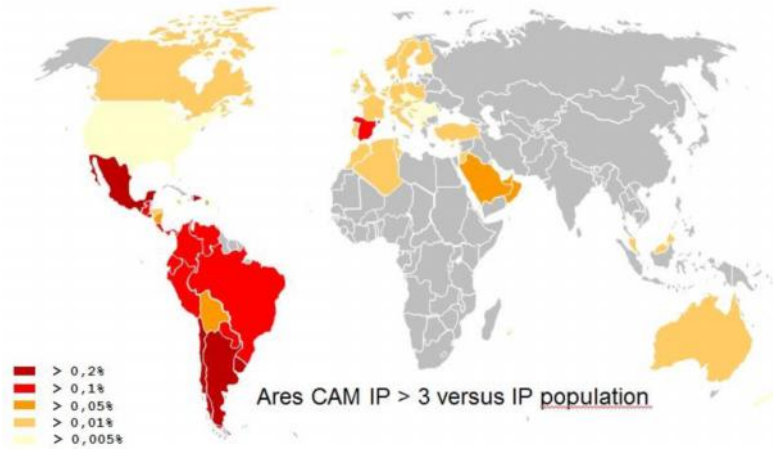


Figure 2. Percentage of IP population sharing at least 3 CAM files

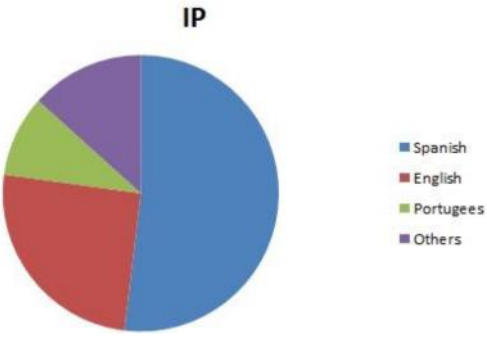


Figure 3. Match countries for IP addresses of ARES Chatrooms

There was another measurement done on the available chatrooms. This was done on three different days by starting Ares and starting the Chatroom pane and counting the advertised languages for the available chatrooms and then matching the IP-addresses of the chatrooms with the Maxmind GeoIP table. From these statistics we can see that there is definitely a close relationship between the popularity of Ares, the Spanish language and the region Latin America. It is obvious that the region Latin America is strongly represented in the absolute numbers of CAM, followed by the United States and Europe (Figure 3).

If we take a closer look and take into account the internet population for each country then we can see that after Spanish speaking countries, the Arabic speaking countries come in second place. It is likely that when internet density grows in these countries they will face the same problems with CAM on P2P.

3.2 Identifying fake files

Ares is contaminated with fake files. Searching for CAM related keywords will generate results that lead to files infected with malware. These files are normally provided by *supernodes* that probably generate theses files on the fly and therefore each search for a keyword will lead to these infected files. By generating a search request for something that really cannot exist, fake files and the related nodes can be identified and blocked. This was done by searching for a string consisting of randomly-generated hex characters.

With this method it is possible to discard a large amount of possibly incorrect information. In our experiment, ten unique identified fake files were downloaded and investigated. They all had nothing to do with CAM, eight of them were related to malware and two files were related to advertisements.

3.3 Gathering information as a *supernode*

Normally a *clientnode* can become a *supernode* automatically, but only if it meets minimum system requirements like the speed of the processor and connection, amount of memory, total uptime, and also a network connection that is not firewalled.

We changed the code and promoted our *clientnode* to *supernode* manually. Also the used TCP and UDP port on the NAT router is forwarded to the IP-address used by this *supernode*.

In our experiment, in the period from May 31, 2015 to June 6, 2015 during approximately 87 hours, we collect data of *clientnodes* connected and logged in to the server. This monitoring was stopped when we have 5000 *clientnodes* that share at least one file. If we have a closer look at these 5000 *clientnodes*, we notice that 24 of them share three or more files identified as CAM with the known hashes. That means 0.48 percent of the 5000 sharing users or 0.06 percent of the 38764 total connected users.

In total 426 files were identified as CAM, 103 files were images and 258 were videos and 65 files were other types. From this information it looks like the amount of CAM video files is much higher than the amount of CAM images. However we only have a table of identified hash values for CAM and could not know how many of these hashes are related to videos or images. Therefore we also compared the filenames and meta-information (serialized string, etc.) with a list of the following keywords: 'babyshivid,' 'hussyfan,' 'kingpass,' 'lsm,' 'mylola,' 'pedo,' 'pthc,' 'ptsc,' 'sdpa.' This resulted in 39 images and 180 videos that are possibly CAM. This also confirms that the amount of shared CAM videos is much higher than the amount of images.

4. ANALYSING THE FORENSIC ARTIFACTS IN THE COMPUTER

There are four important locations where we can find the forensic artifacts related to the use of Ares: the registry, the %localappdata% folder, the Ares shared folder, and download folder. Some of the data, that can be valuable for a forensic investigation, is encrypted, but the encryption routines and the variables needed to decrypt this data is available in the open source code.

4.1 Use of the program, default settings and behavior

When the program is installed, there are several default settings and users can change these settings. Almost all of the settings are stored in the registry.

To investigate a suspect, it is necessary to know which settings is default and which settings are probably changed by the user. By default, when Windows starts, Ares will load automatically connect to the Ares network. It also shows the current status on the Ares control panel. The user can also choose to show this status on the main window. If the user goes to the Control Panel pane, it also shows the users nickname and how many files there are shared. When the program starts the shown window is also the search window. Although suspects often claim that they were not aware that they were sharing something, this is most unlikely. The shared files are shown on the Library pane which is just one click away, the current and also the completed uploads are visible in the Transfer panel and if a user chooses his Control Panel there is not only a subpanel to change the shared library but also the number of current shared files are visible in the caption.

Personal GUID: When Ares is started for the first time it will call the API *Cocreateguid* found in Microsoft's windows file OLE32.dll. This will result in a GUID (Globally Unique Identifier), a unique reference number used as an identifier in computer software This GUID is saved into the value Personal.Guid under the Ares key in the registry. If this API is called again it will create a different GUID. Because the value of a GUID is 128 bits, it is a 'unique' value and it can be used to trace people who are using different IP-addresses. If the GUID found on the network is the same as the stored GUID in the registry it is strong evidence that shows

this the computer was used on the network at that time. However Ares is, unlike other networks sich as Gnutella, Gnutella2 or Emule, not sending the GUID over the network during searches, uploads or downloads. Therefore the GUID is not very useful for investigations.

Nickname: By default the nickname of the user is not set by the installation program and Ares uses an automatic composed nickname like "anon_531dbe69." Such automatic composed nicknames consist of the prefix "anon_" followed by the public IP-address of the computer in hexadecimal format. In this example the 0x531dbe69 stands for IP-address 83.29.190.105.

If the user does not choose a nickname, Ares will popup 60 seconds after starting the program with the question "Would you like to choose your nickname now?" If a user uses that popup to change his nickname the program changes to the "Control Panel Chat settings" pane and the value "anon_531dbe69" for example, is shown as his nickname and can easily be changed. It will be moreover saved to the registry. For this reason a user who is using different IP-addresses and did not set his nickname will be seen with different automatic composed "anonymous" nicknames. However we realize many users on the network with nicknames looking like automatic composed nicknames as described without changing their IP-address, making it likely that there are other Ares clients, probably derived from the original source, which are saving the automatic composed nickname by default.

Port number: On the first start Ares is automatically choosing a port number within the range from 5000 to 64999 and this port number is also saved into the registry. A user can change this port number manually with the Network settings pane. However we realize most of the users do not change this port number.

Shared folders: By default only files in the Ares 'My shared folder' are shared. This folder is set to the desktop of the user by default. The setting is stored in the registry value *Sys.Desktop* with the value *'C:\Documents and Settings\<user>\Desktop\My Shared Folder.*' Files downloaded with Ares are placed in the download folder, which is by default the same as the *'My Shared Folder.'* If a user changes the default location for the download folder it is stored in the registry value *Download.Folder*. In that case the location of the *Download.Folder* becomes the 'My shared folder.' The value *Sys.Desktop* is still in the registry with the old value but that folder is not shared anymore, unless the user shares it afterwards. The default shared folder cannot be unshared. Although users can unclick the share option, this is not changing the setting and on a restart, the old value (shared) is checked again.

Searching: The default search option is 'all' which means 'Search for generic media.' With this feature there are no further search options available. Besides the 'Search for generic media' a user can choose between 'Audio,' 'Video,' 'Image,' 'Document,' 'Software' and 'Other.' Ares uses Sha1-hashing to identify files and this makes it possible to combine downloads for one single file from different users, also known as 'file swarming' or 'multisource downloading.' The hashes are not visible in the search window but a user can make them visible by using the right mouse button option 'export HashLink.' We did several searches for well know keywords related to CAM. From the information on the Results window it is very obvious what the contents of the results could be. Result information often contains gender, ages and explicit information about the sexual aspect.

Downloading and sharing: If a user attempts to download a file but the sources are no longer available or the relevant clientnodes

are too busy, Ares will automatically search for other sources by using the sha1 hashvalue and add new sources to the download. For each started download a separate file is created in the download folder with a prefix'___ARESTRA___' (3 underscores ARESTRA 3 underscores) followed by the filename. This file is used to store the separated chunks for a download and that will lead to a complete file when all bytes of the file are downloaded. At the end of this 'arestra' file there is an appendix used to store the information about the available sources for this download. This appendix is removed on completion of the download. Downloaded files are shared after the download is completed unless the file is marked as corrupt.

Preview, watch and listening: While downloading the user can preview the downloaded part of multimedia files with a built-in player. If case of a video file Ares will create a folder *%localappdata%\ares\data\temp*. The name of that folder is a random number. Then a copy of the part that is already downloaded will be copied to that folder and played from there. After the preview, it will be deleted when the user is previewing another file. With the built in player it is also possible to play multimedia files stored on the system and that includes the completed downloads. We could not find any traces in case Ares was running and the program was used to play completed downloads or play local stored multimedia files by drag and drop these files on the Ares player window.

The player also supports Shoutcast and it is also possible that watched movies with shoutcast are placed in the data*temp* folder.

4.2 Ares and the registry

Some derived Ares clients, like "Limewire Pro" and "Limewire Plus" (despite their names, these clients have nothing to do with Limewire or Gnutella, but are instead real Ares clients) store the registry data in the same way, but with a different key and with additional values. A user can remove his complete search history or remove a single search with the Ares program. In that case the *subkeys* and *stringvalues* will be deleted. Sometimes these deleted keys can be recovered with the program YARU (Yet Another Registry Utility) [19]. Unallocated clusters and restore points can be used to carve old *ntuser.dat* files. An uninstall procedure will remove the registry keys while updating Ares will keep the old settings and therefore updating will probably only have a minor effect like adding new keys. Besides the keys that are added and changed by the Ares program, the operating system will also add and change keys to the registry during installation, use and uninstall of Ares. A search for "Ares" within the registry is a very effective way to discover Ares artifacts that are caused by the operating system. This paper however does not cover that aspect.

4.2.1 Overview of registry values

Let's examine most important registry keys and values. For each entry there is a row with 4 columns followed by a row with explanation. Sometimes the valuable information is in the name of the key or in the name of the value (searches). The values are coded in Hex or e.g. in *UnixTimeStamp* and most of the time in *BigEndian*. If needed, the decode and Endian information is available in the row with explanation.

4.2.2 Encryption methods used by Ares for the registry

A lot of the Ares encryption procedures are combinations of simple XOR and bit shifting operations. It uses a separate XOR operation for each byte in a given string. This method is known as stream cipher encryption [25]. The initial key for the XOR operation is a hardcoded word value as input parameter and

different for each procedure. Only the most significant byte of this word value is selected with a shift right 8 operation and that byte will be used for the XOR operation.

During the encryption of a given string this XOR-byte (the key) is recalculated by taking the current processed byte and adding the value of the current XOR-byte (the key) followed by a multiply and add operation with hardcoded values. The differences for the encryption procedures are the input parameters (string to encrypt and the key) and the hardcoded values for the multiply and add operation for changing the key.

All encryption procedures are written in the function *"helper_ crypt.pas,"* which is part of the open source code. The programmers are also using nested encryption functions like d67(d64(s,24884),7193). Here follows an example for the decryption of an important key with the name *Stats.CFRTime*. This key is filled with an encrypted *UnixTimeStamp* when Ares was started for the first time. For this experiment, the value stored in the register is 0x54 5F C0 B2 1C 92 41 CD C3 83. This value is decrypted with d67(d64(s,24884),7193). After decrypting, the first character of the decrypted value is removed because it was a random added value during encryption. The first 4 bytes that remain are a *UnixTimeStamp* stored in little endian. The following 5 bytes are used for an integrity check. Byte 5 must be 0x00 and byte 7 and byte 8 are checked against the value of the first 4 bytes. For this check these 4 bytes are hashed with a function called "whaaa," in the code this function is commented as "gnutella query routing word hashing." After this hashing, a value of 12 is added and then the value of this result has to be equal to the value stored in byte 7 and 8. This integrity check is very important because the suspect can claim that the value in the registry could have been changed by a virus or whatever. Therefore we have to do the decryption and also check the integrity of that decryption result.

The whole procedure for the value stored in the key *Stats.CFRTime* is as follows. First we have to run the d64(s,24884) operation. To make it more readable there is a small description for the decryption of the first byte and the calculation of the new XOR key. After this, the procedure is repeated in the same way until the encryption is done. The word value to start the first XOR operation is 24884 or 2 bytes with the values 11000010 00110100. The shift right 8 operation results in the first byte 11000010 and is used as the (key) XOR-byte in the first XOR operation.

For the second byte the key has to be recalculated. The formula used is b := (byte(S[I]) + b) * 12559 + 14926; so we have to take the value of the current byte of the input string add the value of the current key, then multiply by 12559 and then add 14926. The current byte in the input string is 0x54=84 and the current key is 24884. So the new key is (84+24884) * 12559 + 14926 = 313588038 or 00010010101100001111100101000110b. Because the variable is a word, only the 2 least significant bytes are used. Again followed by a shift right operation, it results in 11111001b for the new key. And this procedure is repeated until all bytes are processed in the same way.

4.2.3 Hashchecking "gnutella query routing word hashing" (Ares procedure whaaa)

In this example the values of byte 7 and 8 of the previous decrypted *Stats.CRFTime* are checked with the value of the first 4 remaining bytes. Remember we already dropped the first random byte and we also checked the 5th byte to be a Nullbyte. The value 0xD072 has to be hashed and after this a value of 12 has to be added. The formula used is word (whaaa(copy(s,1,4))+12). The result has to be

equal to the value of the first 4 bytes that are still in Little Endian. The first step in the hashing is to first lowercases the input, this will only effect hexvalues in the range from 0x41 until 0x5a (ascii A-Z). Step 2 is to convert the result into big endian and then step 3 is to multiply the result with a hardcoded hex value of 0x4F1bBCDC. In step 4, the result of this in a 32 bit program like Ares, and is only the least significant bits. Step 5 is a shift left32 followed by a shift right48. Step 7 is adding a value of 12 and then step 8, convert it to Little Endian. Result of the manual decrypted and decoded data were compared with Encase Ares Registry Report (V1.1.0), EnPack and AresDecrypter 1.3 and there were no differences worth mentioning.

4.3 Ares and the folder %localappdata%

Within this folder there are several files that contain valuable data for a forensic examination. During installation Ares will create the files *Snodes.dat, ShareH.dat, ShareL.dat, FailedSnodes.dat*. After running Ares will also create the file *DHTnodes.dat* and depending on which parts of the Ares program are used and which settings are used, Ares changes also the files *PHashIdx.dat, ChatroomIPs.dat default.m3u, 'Shared Folders.txt,' avatar.bmp, Avatar.jpg and MiniAvatar.jpg*. Due to the limitation of the number of words of this paper we only present the forensic analysis of *ShareH.dat*.

4.3.1 ShareH.dat

The file ShareH.dat is a permanent file, created during installation with a header containing the ascii string '__ARESDB1.02H_' hexadecimal value 0x5F5F415245534442312E3032485F. This file contains valuable information about shared files, in the source named as "trusted metas." A record for a shared file is added if it contains Ares data like meta-information or parameters. Therefore for each unique file that is downloaded with Ares, a record is added to this file. In that case Ares will also add a record in the file *shareL.dat*. For files that are not downloaded with Ares but added to the Ares shared files in another way, e.g. copying it to the shared folder with the operating system, there are no records added to this file. However if a user changes meta-information with Ares, for such an added file, this will also result in a record in ShareH.dat for that file with changed meta- information.

For investigators the contents of *ShareH.dat* is extremely valuable because the records are not removed from the file *ShareH.dat* if the shared file itself is deleted, while records in the file *ShareL.dat* will be deleted. Therefore the *ShareH.dat* is also known as the *ShareHistory*, but be careful- it is only a history of files that contained Ares information as described before. Besides the fact that records are not deleted; they also contain a sha1 hash value for the shared file making it possible to compare records with known hashes.

There is a Boolean value in each record indicating if the file is shared. If a file is shared that does not mean that the file is downloaded. It just means that the file was shared and thus available for downloading.

Ares indicates if a file is corrupt with a Boolean value. This value False indicates that there is something wrong with the file. This could be caused by a fault during downloading. When a file is completely downloaded the sha1 hash value is used to check if the file is exactly the same as the source file. If the sha1 value differs, the file will be marked as corrupt. Another reason to mark a file as corrupt is when it is identified as 'teen content.' This is done by a check on category, title, path and artist with a small list of CAM related keywords. Looking at the huge amount of CAM distributed on Ares this filtering is not

very effective. If a file is marked as corrupt, the value shared is also marked as false and the file won't be shared.

If a file is downloaded with Ares, a record contains a value 'filedate.' This is a UnixTimeStamp when the download was completed. The file itself will have a creation datewhen the download was started.

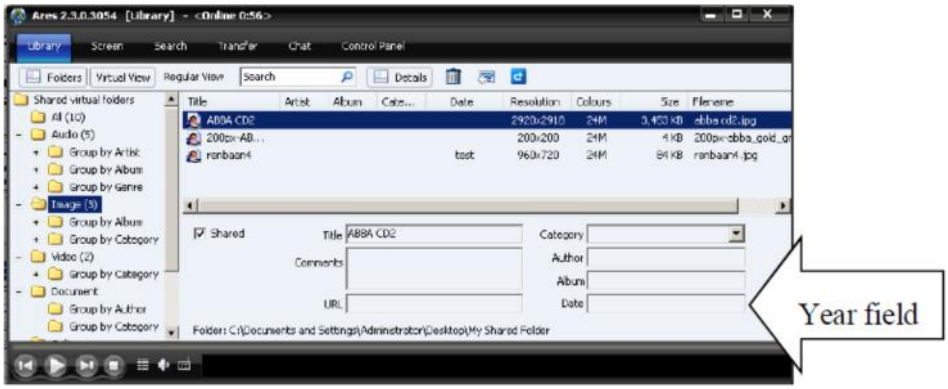


Figure 4. Changing meta-information with Ares

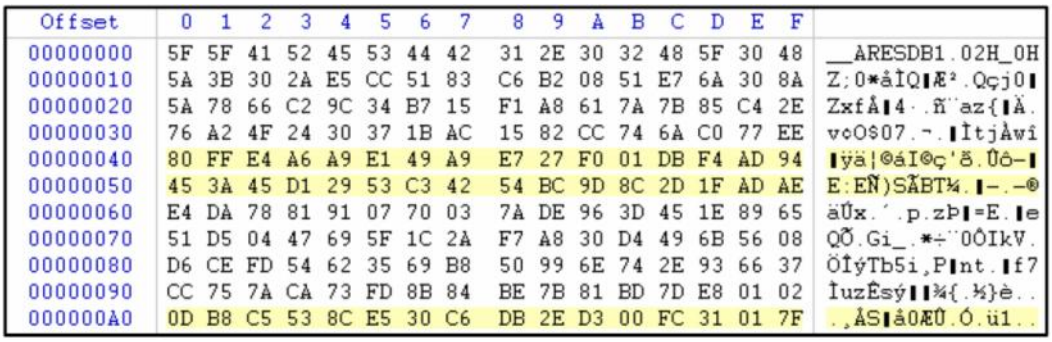
If a file was not downloaded with Ares but add to Sharel.dat in another way, e.g. changing meta -information, this field 'filedate' is empty. The field 'filedate' in the *ShareH.dat* cannot be change by the user. There is a field with the name Date in the meta-information that can be changed by the user, but this field reflects to the field Year in the *ShareH.dat* records, which is just a string value (Figure 4).

4.3.2 Decryption of ShareH.dat

The valuable data in the file, which is located after the header, is stored encrypted. This data can be decrypted with the same methodology as described in section 4.2.2. It is an XOR operation for each byte with the XOR key given as a fixed value on startup, and changing for each next XOR operation by using the value of the current XOR key added with the current byte of the input string multiplied with a fixed value of 23219 and followed by adding a fixed value of 36126. After the header we have to read the first 23 bytes and decrypt them. After decryption, the first 20 bytes are the sha1 hash value of the shared file. The next byte is a Boolean value and will tell if the file was shared. The following 2 bytes contain the length for the rest of the record. This value is in Little Endian and the maximum value is 1024. Now we know this length we can read the rest of the record which is still encrypted and decrypt it with the same code. The xor key starts again with the value 13871. This remaining part contains meta-information like artist, title and comment (Figure 5 and Figure 6). The procedure has to be repeated until all records are decrypted. The full layout of the record is available in the pascal file helper_library_db.pas in the procedure 'get_trusted_metas' which is part of the open source.

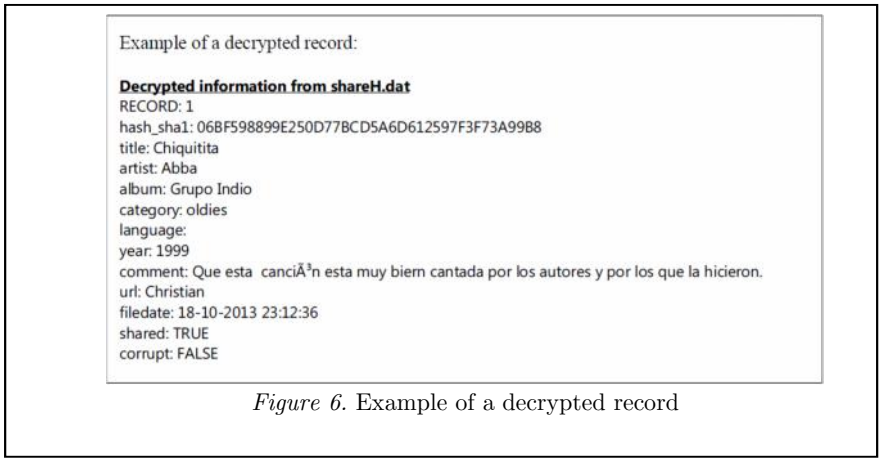
5. CONCLUSION AND FUTURE WORK

In this paper, we aim to give an overview how the Ares network is functioning. This is accompanied by an overview about the use of Ares in relation with the worldwide distribution of CAM related to countries and spoken languages.



Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
00000000	5F	5F	41	52	45	53	44	42	31	2E	30	32	48	5F	30	48	__ARESDB1.02H_0H
00000010	5A	3B	30	2A	E5	CC	51	83	C6	B2	08	51	E7	6A	30	8A	Z;0*åÌQ▮Æ².Qçj0▮
00000020	5A	78	66	C2	9C	34	B7	15	F1	A8	61	7A	7B	85	C4	2E	ZxfÅ▮4·.ñ¨az{▮Ä.
00000030	76	A2	4F	24	30	37	1B	AC	15	82	CC	74	6A	C0	77	EE	v¢O$07.¬.▮ÌtjÀwí
00000040	80	FF	E4	A6	A9	E1	49	A9	E7	27	F0	01	DB	F4	AD	94	▮ÿä¦©áI©ç'ð.Ûô-▮
00000050	45	3A	45	D1	29	53	C3	42	54	BC	9D	8C	2D	1F	AD	AE	E:EÑ)SÃBT¼.▮-.-®
00000060	E4	DA	78	81	91	07	70	03	7A	DE	96	3D	45	1E	89	65	äÚx.´.p.zÞ▮=E.▮e
00000070	51	D5	04	47	69	5F	1C	2A	F7	A8	30	D4	49	6B	56	08	QÕ.Gi_.*÷¨0ÔIkV.
00000080	D6	CE	FD	54	62	35	69	B8	50	99	6E	74	2E	93	66	37	ÖÎýTb5i¸P▮nt.▮f7
00000090	CC	75	7A	CA	73	FD	8B	84	BE	7B	81	BD	7D	E8	01	02	ÌuzÊsý▮▮¾{.½}è..
000000A0	0D	B8	C5	53	8C	E5	30	C6	DB	2E	D3	00	FC	31	01	7F	.¸ÅS▮å0ÆÛ.Ó.ü1..

Figure 5. View in hex and ASCII of ShareH.dat



Example of a decrypted record:

Decrypted information from shareH.dat
RECORD: 1
hash_sha1: 06BF598899E250D77BCD5A6D612597F3F73A99B8
title: Chiquitita
artist: Abba
album: Grupo Indio
category: oldies
language:
year: 1999
comment: Que esta cancIÃ³n esta muy biern cantada por los autores y por los que la hicieron.
url: Christian
filedate: 18-10-2013 23:12:36
shared: TRUE
corrupt: FALSE

Figure 6. Example of a decrypted record

The evidence about the use of Ares, to be found on a computer, is documented. This includes a description of the important parts of the decoding and decryption methods. At the time of writing this paper, we also worked on the forensic analysis of the communication between *clientnode* and *supernode* over Ares network protocol. For analyzing huge amount of artifact data in ARES P2P network, we are also looking at efficient data mining techniques such as [20][21][22][23].

REFERENCES

Analysis and characterization of Peer-to-Peer Filesharing Networks. Retrieved on February 2015 from http://personales.upv.es/jlloret/pdf/icosmo2004.pdf

New client – Ares Link Retrieved on February 2015 from http://www.gnutellaforums.com/ general - gnutella-gnutella-network-discussion/13828-fyinew-client-ares.html

Nathaniel, Leibowitz et al. Deconstructing the Kazaa Network, Retrieved on January 2016 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8970&rep=rep1&type=pdf

Archie, Kuo, Ethan Le, Spotlighting Decentralized P2P File Sharing, Retrieved on January 2016 from http://cs.sjsu.edu/faculty/stamp/CS158B/syllabus/papers/DecentralizedP2P.doc

239 F.3d 1004 Retrieved on February 2015 from

https://law.resource.org/pub/us/case/reporter/F3/239/239.F3d.1004.00-16403.00-16401.html

Ares Galaxy / Discussion / Open Discussion: To the developers of this project URGENT! Retrieved on February 2015 from

http://sourceforge.net/p/aresgalaxy/discussion/384787/thread/1a7f4503/

Janis Wolak, David Finkelhor & Kimberly J. Mitchell, "Trends in Arrests for Child Pornography Possession: The Third National Juvenile Online Victimization Study (NJOV-3) published in bulletin 4-13-2012 by the University of New Hampshire, Crimes Against Children Research Center

Matthew L. Long, Laurence A. Alison and Michelle A. McManus concluded in their article 'Child Pornography and Likelihood of Contact Abuse: A Comparison Between Contact Child Sexual Offenders and Noncontact Offenders'

Janis Wolak, Marc Liberatore and Brian Neil Levine wrote an article 'Measuring a year of childpornography trafficking by U.S. computers on a peer to peer network', published in Child Abuse & Neglect 38 (2014) 347–356

Michelle Ann McManus, Matthew L. Long, Laurence Alison & Louise Almond wrote an article 'Factors associated with contact child sexual abuse in a sample of indecent image offenders', published in Journal of Sexual Aggression: An international, interdisciplinary forum for research, theory and practice, DOI: 10.1080/13552600.2014.927009

KCeasy, Retrieved on February 2015 from http://sourceforge.net/projects/kceasy/

giFT: Internet File Transfer, Retrieved on February 2015 from http://gift.sourceforge.net/

giFT's Interface Protocol, Retrieved on February 2015 from http://gift.sourceforge.net/docs/0.11.x/interface.html

Ares Dat File Decryptor Retrieved on January 2016 https://www.carbonaria.nl/aresdecrypter.html

Guidancesoftware.com/appcentral Retrieved on February 2015 https://www.guidancesoftware.com/appcentral/pages/searchresults.aspx?k=ares

Wireshark · Go Deep. Retrieved on May 2015 https://www.wireshark.org/

GeoLite Legacy Downloadable Databases Maxmind Developer Site, Retrieved on February 2015

http://dev.maxmind.com/geoip/legacy/geolite/

The World Factbook, Retrieved on February 2015

https://www.cia.gov/library/publications /the-world-factbook/

Yet Another Registry Utility, Retrieved on January 2016 https://www.tzworks.net/prototype_page.php?proto_id=3

N-A Le-Khac, L. Aouad and M-Tahar Kechadi Distributed Knowledge Map for Mining Data on Grid Platforms, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.10, October 2007

N-A Le-Khac, L. Aouad and M-Tahar Kechadi, A New Approach for Distributed Density Based Clustering on Grid Platform, Data Management. Data, Data Everywhere, Volume 4587 of the series Lecture Notes in Computer Science pp 247-258

L. Aouad, N-A. Le-Khac and M-T. Kechadi, "Lightweight Clustering Technique for Distributed Data Mining Applications," Chapter in Advances in Data Mining. Theoretical Aspects and Applications, Volume 4597, Lecture Notes in Computer Science pp 120-134, 2007

L. Aouad, N-A. Le-Khac and M-T. Kechadi, Grid-Based Approaches for Distributed Data Mining Applications, Journal of Algorithms & Computational Technology, Vol. 3 (4), 2009, pp.517-534

