Investigating Encrypted Material

Niall McGrath, Pavel Gladyshev, Tahar Kechadi, and Joe Carthy

University College Dublin, Dublin, Ireland

Abstract. When encrypted material is discovered during a digital investigation and the investigator cannot decrypt the material then s/he is faced with the problem of how to determine the evidential value of the material. This research is proposing a methodology of extracting probative value from the encrypted file of a hybrid cryptosystem. The methodology also incorporates a technique for locating the original plaintext file. Since child pornography (KP) images and terrorist related information (TI) are transmitted in encrypted format the digital investigator must ask the question *Cui Bono?* – who benefits or who is the recipient? By doing this the scope of the digital investigation can be extended to reveal the intended recipient.

Keywords: Encryption, Ciphertext, OpenPGP, RSA, Public & Private Keys.

1 Introduction

Law enforcement agencies (LEA) encounter encryption in relation to the distribution of KP [1] and of TI [2] offences. For example a KP distributor encrypts the KP material with PGP and posts it into a newsgroup or interest group via anonymous re-mailer or via an instant messenger system. The accomplice who is subscribed to that group receives encrypted material and can decrypt it. The anonymity of all involved parties is preserved and the content cannot be decrypted by bystanders. The use of PGP encryption in general has been cited [3] as a major hurdle in these investigations. In addition, during digital investigations evidence is often discovered which extends the scope of the investigation. These are compelling reasons for the computer forensic investigator to be able to identify encrypted material, examine it and finally extract evidential value from it. This paper presents a methodology that was formulated from experiments and it facilitates the identification of the recipient of PGP encrypted material. As an adjunct to this a technique that identifies the plaintext file that was encrypted is presented. Subsequently a technical evaluation was carried out in a case study to validate the methodology.

2 Problem Description

The investigation of subject A is initiated and a forensic image of the hard disk drive (HDD) is taken. Analysis is carried out and it is found that there is a significant amount of ciphertext files and plaintext files containing evidence. Subject A is a suspected distributor/seller of KP and subject B whose identity is unknown is the recipient of the encrypted material. The objective of this research is to establish an evidential link

M. Sorell (Ed.): e-Forensics 2009, LNICST 8, pp. 29 – 35, 2009.

between the encryptor and the recipient of PGP encrypted material and subsequently identify the plaintext file that was encrypted. In this scenario subject A must have had subject B's public key and PGP encrypted the plaintext material to form the ciphertext. Subject B can decrypt the ciphertext with his private key when he receives it. PGP is a hybrid cryptosystem where the ciphertext created by it follows the OpenPGP message format specified in [4]. A hybrid cryptosystem is a combination of symmetric and asymmetric encryption. A symmetric key is session generated and then this is used to encrypt data. The symmetric key is then encrypted using the recipient's public key. The public key can be stored and distributed by a key server. The symmetrically encrypted data and the asymmetrically encrypted symmetric key are the major components of a PGP ciphertext data-packet. PGP also compresses data before encryption for added security because this helps remove redundancies and patterns that might facilitate cryptanalysis, compression is only applied to the symmetrically encrypted data-packet. PGP typically uses the *Deflater* (zip) algorithm for compression.

2.1 Methodology

The methodology which facilitates the investigation of PGP encryption is outlined in Fig 1 and consists of a number of steps that are described in the following sections.

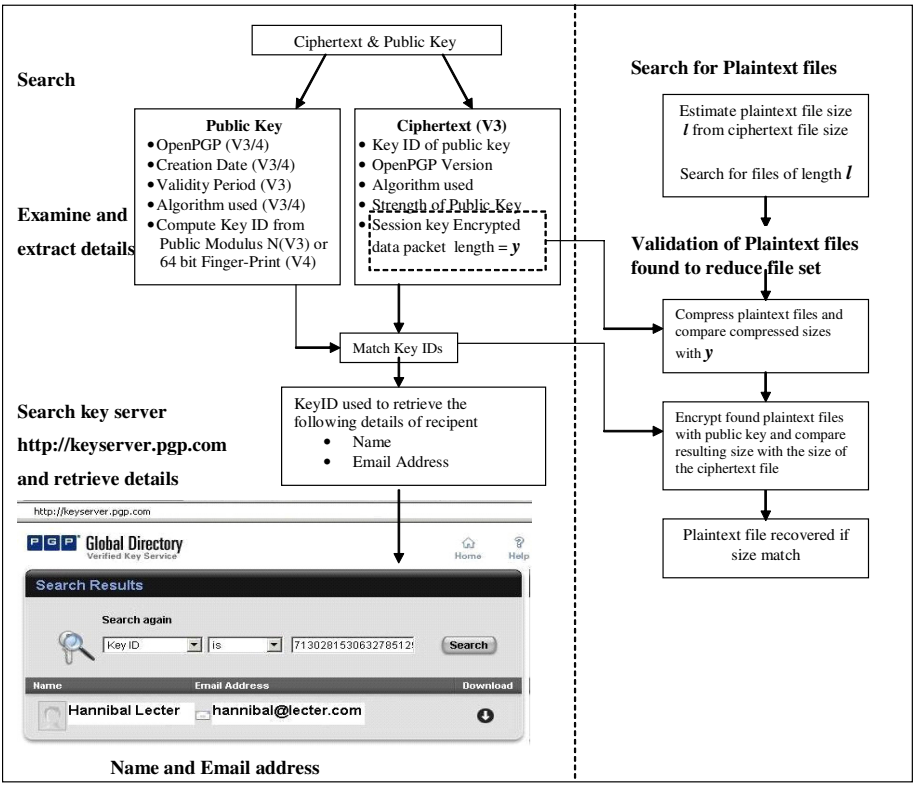

Fig. 1. Methodology for investigating PGP Encryption

In order to validate this research a framework of Java classes was created to generate OpenPGP encryption keys and data, a file parser to extract and analyse information from data, a test harness and a compression engine that examined ZIP compression.

2.1.1 Searching for Public Key and Ciphertext Artefacts

The first step of the methodology is to search for OpenPGP artefacts on the subject's HDD i.e. public keys and ciphertext files. Certain hexadecimal values can be used as signatures when searching for OpenPGP artefacts. These values have been determined experimentally and are shown in Fig 2. and Fig 3.

OpenPGP-PublicKey (V4)	Search Signature
512 Bit Key	\x98\x4d\x04
1024 Bit Key	\x98\x8d\x04
2048 Bit Key	\x99\x01\x0d\x04

Fig. 2. OpenPGP version 4 Public Key Search Criteria

OpenPGP-Ciphertext (V3)	Search Signature
512 Bit Key	\x84\x4c\x03
1024 Bit Key	\x84\x8c\x03
2048 Bit Key	\x85\x01\x0c\x03

Fig. 3. OpenPGP version 3 Ciphertext Search Criteria

When searching for public keys the linear relationship in equation 1, which has been determined experimentally, can be used to estimate the public key file size from the key strength. Key strength can be determined from section 2.1.3 below.

$$y = 3.9802x - 184 . \quad (1)$$

where x=key size in bytes and y = key strength in bits.

(1) was observed from the data of 300 generated RSA asymmetric keypairs i.e. 100 keypairs each of 512, 1024 & 2048 bit strengths.

2.1.2 Analysis of Public Key Artefacts-Examine and Extract Details

For each found public key file the next step of the methodology is to determine the following components of the OpenPGP public key: *Version* number, *Creation Date*, *Validity Period* (only for V3), *Algorithm* type and the *Public Modulus* (*N*). The low order 64 bits of *N* is the *Key ID* for a V3 key. The *Key ID* for a V4 key is the lower order 64 bits of the key finger print. This has to be computed because it is not an

element of the OpenPGP specification however the method of computation is specified in the OpenPGP standard [4]. *Creation Date* and *Validity Period* are used to search for the corresponding ciphertext file because the timestamp of the ciphertext file will obviously be greater or equal to *Creation Date* of public key and less than or equal to *Validity Period* (if V3). The location of these items in the public key is specified in [4].

2.1.3 Analysis of Ciphertext Artefacts-Examine and Extract Details

For each ciphertext file found the following components of the OpenPGP ciphertext are determined: *Version* number, *Key ID* and *Key strength* of the public key used to create the ciphertext and *length* of *Symmetric Key Encrypted Data Packet*. The location of these items in the ciphertext is specified in the OpenPGP standard [4].

2.1.4 Matching the Key ID from Public Key and Ciphertext

In the next step the *Key ID* retrieved from ciphertext has to be compared with the computed *Key ID* from the public key. When a match occurs then it can be concluded that the public key was used to create the ciphertext under analysis.

2.1.5 Search Key Server – http://keyserver.pgp.com

Subsequently, the *Key Id* of public key is used to search a designated keyserver for *name* and *email address* of the owner of the public key i.e. the recipient of the encrypted material. These details are easily retrieved by inputting the *Key ID* into the website.

2.1.6 Approximating the Size of Plaintext File from ciphertext File

This is the initial step to searching for the original plaintext file. From the experiments carried out it was determined that there is a linear relationship between ciphertext file size and plaintext file size of JPG files, please see equation 2 below. Once l is evaluated then plaintext files of length l bytes can be used to reduce the size of the candidate file set.

$$y = 1.0019\, l - 8249.9 \,. \tag{2}$$

where y=length of encrypted JPG file and l = length plaintext JPG file

(2) was derived from the generated data of encrypting a number of JPG files with a 100 RSA keys each of 1024 bit strength; using AES 256 bit symmetric encryption with ZIP compression.

2.1.7 Validation

The final step is to validate the candidate plaintext file(s) that are identified by the approximating process above. This is done by passing the plaintext files through the *Deflater Engine*, which compresses the files. The *number* of compressed bytes that the plaintext file deflates to closely approximates the *length* of the data packet to be session key (symmetrically) encrypted. This size is the *length* of data packet after compression has taken place. Finally each plaintext file (in the reduced file set) is encrypted. This will definitively determine the original plaintext file that was encrypted and exchanged.

2.2 Case Study

This case study is modelled on the described problem above and the methodology is applied practically to reveal the recipient of the exchanged encrypted material and to search for the original plaintext file. Techniques for investigating recently run programs, the registry and *NTUSER.dat* file and internet search history are outlined in [5]. In addition there are specialised techniques listed in [6] for investigating AIM related incidents. An investigation was carried out based on these techniques and it was established that an incident where America Online Instant Messenger (AIM) is used in conjunction with PGP encryption took place. An encrypted file that was transferred to subject B using AIM was located. AIM provides real-time one-to-one messaging between computers and attachments can be encrypted with the recipient's PGP public key.

2.2.1 Search and Analysis of Artefacts – Extraction of Significant Information

An Encase® search was executed with the criteria in Fig 2 and Fig 3. Then the parser was run to carry out the automated analysis and matching of the public key and corresponding ciphertext *Key ID*s. When a match is detected between the two *Key ID*s output is generated, please see parser output in Fig 4.

Output from Parser
Analysis of PGP Public Key Artefact
Version 4. Computed *key ID* value is *7130281530632785129.* *Key Created Sat Jun 14 13:04:46 2008.* Validity Period: N/A because key is Version 4. *Value = 1, RSA -> Encryption* key strength is 1024.
Analysis of PGP Ciphertext Artefact
Version 3. *key ID retrieved from ciphertext is 7130281530632785129* *Key Strength is 1024 bit* *Value = 1, RSA -> Encryption* *Length of compressed data packet to be session key encrypted is 3729810.*
Match of Key IDs from Public Key & Ciphertext Files
Key ID from Public Key and Key ID from Ciphertext match -> Public Key was used to encrypt the Ciphertext.

Fig. 4. Output from Parser

2.2.2 Approximating the Size of Plaintext File from ciphertext File

Since the size of the ciphertext file identified by the parser is 3,729,956 bytes, then using equation 2, / evaluates to 3,731,116.8. Then a search for plaintext files of length 3,731,116.8 bytes is carried out. This search yielded 7 candidate files. In order to further reduce this file set a validation process was carried out, which is explained below.

2.2.3 Validation

The *Deflator* engine was then used to determine what size the 7 plaintext files, short listed in the previous steps, will compress to. Using the ciphertext file from the parser output; the length of data packet to be session key encrypted is 3729810. This is the size that the original plaintext data compresses to before encryption takes place. Therefore the number of compressed bytes, that the plaintext files *deflates* to, will closely approximate to a size of 3729810 bytes. This process reduces the file set down to 2 candidate files.

Finally these 2 plaintext files were encrypted giving 2 new ciphertext files of definitive sizes. These sizes were compared with the original ciphertext file size. Hence the original plaintext file that was encrypted and exchanged with AIM was conclusively determined. This was the designated file that contained evidence. Incidentally, 3729810 is the size of the file after compression takes place and then after encryption this becomes 3729824. This is due to the fact that OpenPGP uses CFB mode encryption and this mode operates on blocks of fixed length i.e. 16 bytes. This will give rise to the "stair casing effect" in terms of size of the encrypted file. The encrypted block size will be padded out to fit the block size.

3 Conclusion

The proposed methodology encompasses the searching for PGP public keys and encrypted material, the analysis of these and subsequently the extraction of the *Key Id* of the key used to encrypt the plaintext file. This enables the identification of the intended recipient (owner of public key) of the encrypted material by searching a global directory service like PGP key server. The integrated search technique facilitates the identification of the original plaintext file. Then by viewing the contents of the plaintext it can be determined if it is evidence or not. (We are assuming that the participation of subjects A and B in the exchange of encrypted material was wilful and knowing.) If the contents hold evidence the LEA can now serve a search warrant on subject B to seize his HDD. The methodology adheres to computer forensic standards of evidence *Search & Seizure, Acquisition and Retrieval*. In addition the methodology which was carried out in the Irish jurisdiction does not violate any civil liberties and the subjects' right to privacy is upheld. Irish law operates exclusionary rules in respect of evidence which has been gathered illegally or in breach of constitutional rights. However, *Section 8* of the *Data Protection Law* in Ireland provides an investigative clause i.e. disclosure of personal data in certain cases. There is also the *Anton Piller Order* which deals with special investigative circumstances like serious crime which provides for the right to search premises and seize evidence without prior warning.

4 Research Contribution

A methodology for the investigation of encrypted material has been formulated. This methodology facilitates the extraction of evidential value from the encrypted material to enable the identification of the recipient of the encrypted material. The incorporated

search technique correlates the ciphertext file under investigation with the original plaintext file. This reduces the investigation time by carving the data under investigation into a significantly reduced file set. This is an entirely novel approach to investigating encrypted material and it is fully automated.

References

1. Carter, H.: Paedophiles jailed for hatching plot on internet (2007)
2. Joseh, S.: Hamas Terror Chat Rooms, December 11 (2007)
3. Siegfried, J., et al.: Examining the Encryption Threat, Computer Forensic Research and Development Center. International Journal of Digital Evidence (2004)
4. Callas, J., et al.: PGP Corporation OpenPGP Message Format (November 2007)
5. Bunting, S.: The Official EnCase Certified Examiner Guide. Wiley, Chichester (2008)
6. Dickson, M.: An Examination into AOL Instant Messenger 5.5. Elsevier Ltd., Amsterdam (2006) (Digital Investigation 3)