



# An Empirical Study Profiling Internet Pirates

Pierre Lai, Kam-Pui Chow, Xiao-Xi Fan, Vivien Chan

## ► To cite this version:

Pierre Lai, Kam-Pui Chow, Xiao-Xi Fan, Vivien Chan. An Empirical Study Profiling Internet Pirates. 9th International Conference on Digital Forensics (DF), Jan 2013, Orlando, FL, United States. pp.257-272, 10.1007/978-3-642-41148-9\_18 . hal-01460610

**HAL Id: hal-01460610**

**<https://inria.hal.science/hal-01460610v1>**

Submitted on 7 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Chapter 18

# AN EMPIRICAL STUDY PROFILING INTERNET PIRATES

Pierre Lai, Kam-Pui Chow, Xiao-Xi Fan and Vivien Chan

**Abstract** Internet piracy has become a serious problem due to the expansion of network capacity and the availability of powerful hardware. To combat this problem, industry and law enforcement need a better understanding of the behavioral characteristics of Internet pirates. This paper describes a new conceptual framework for profiling Internet pirates. Also, it presents a taxonomy based on a survey of 114 Internet pirates. The taxonomy, which includes six types of downloaders and six types of file sharers with different behavioral characteristics, provides useful insights to forensic scientists and practitioners who are focused on combating Internet piracy.

**Keywords:** Internet pirates, criminal profiling, behavioral characteristics

### 1. Introduction

Internet piracy is the act of illegally copying or distributing copyrighted digital files using the Internet [7]. Over the years, industry and law enforcement agencies have conducted major operations to stop these illegal activities. However, their efforts appear to be unable to reduce Internet piracy. It is estimated that at least 23.76% of the world's Internet bandwidth is devoted to the transfer of infringing, non-pornographic content [6].

Rogers [13] has noted that criminal profiling can benefit investigations of cyber crimes such as Internet piracy because it helps develop effective investigative and media search strategies, and helps reduce the number of possible suspects. Turvey [14] has demonstrated how behavioral evidence of cyber crimes can be analyzed and used to profile offenders. While it is promising to use profiling techniques to understand the behaviors and personal characteristics of Internet pirates, existing Internet

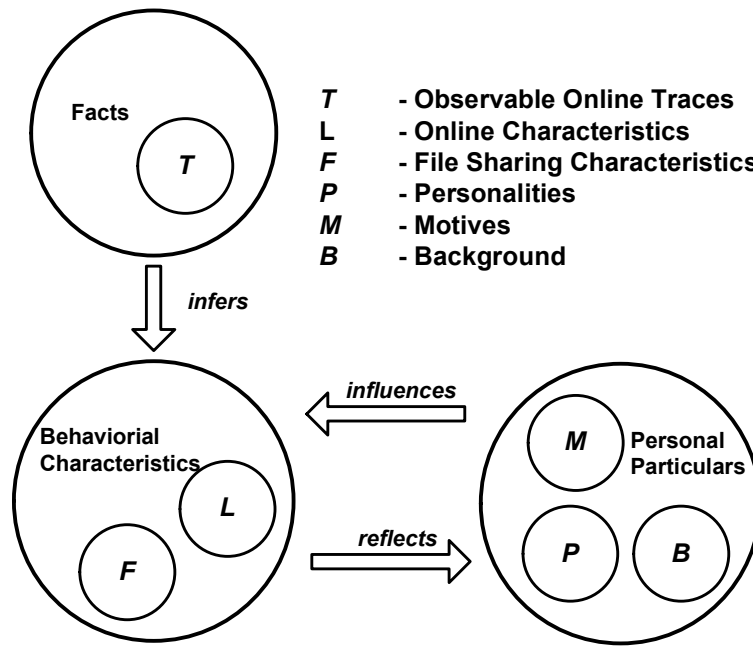


Figure 1. Conceptual framework for profiling Internet pirates.

piracy research (e.g., [7, 8]) has not attempted to relate the behaviors of Internet pirates to their personal characteristics.

This paper describes a conceptual framework for profiling Internet pirates along with a taxonomy of Internet pirates developed using an empirical study based on the framework. The taxonomy is constructed using data gleaned from an extensive survey of individuals involved in Internet piracy related activities. It includes the observed behavioral patterns and associated personal characteristics of twelve types of downloaders and file sharers.

## 2. Profiling Internet Pirates

Because cases of Internet piracy occur in the cyber world, the characteristics and inferences used in the conventional profiling process need to be redefined. In particular, crime scene evidence, which offers the principal clues in a traditional criminal investigation, is not well-defined in the context of Internet piracy.

The conceptual framework for profiling Internet pirates is shown in Figure 1. The framework incorporates three categories: (i) facts (observable online traces); (ii) behavioral characteristics (online characteristics

Table 1. Category and variable description.

| Category                                | Variable                                       | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Facts                                   | Observable On-line Traces ( $\mathcal{T}$ )    | Set of traces observable on the Internet, including visible facts and data that exist online and are associated with a file sharing act. Examples are the name of the user who shared a file and the timestamp when the file was uploaded.                                                                                                                                                                                                                                                                   |
| Behavioral Characteristics              | Online Characteristics ( $\mathcal{L}$ )       | Set of attributes describing the online habits or usual practices of an individual. Examples are the amount of time a person spends online each day, the period when he or she is most active, and what he or she usually does online.                                                                                                                                                                                                                                                                       |
|                                         | File Sharing Characteristics ( $\mathcal{F}$ ) | Set of attributes associated with the file sharing activities of an individual. Examples are how often a person uploads files and the locations where the files are usually published.                                                                                                                                                                                                                                                                                                                       |
| Personal Particulars ( $\mathcal{PP}$ ) | Personality ( $\mathcal{P}$ )                  | Set of internal traits of an individual. These traits affect how a person feels, thinks or responds to external factors. Examples are if a person is positive and outgoing or quiet and introverted.                                                                                                                                                                                                                                                                                                         |
|                                         | Motive ( $\mathcal{M}$ )                       | Set of reasons why a person is engaged in a certain act. Examples are to enjoy the resources for free and to obtain the resources with minimal effort.                                                                                                                                                                                                                                                                                                                                                       |
|                                         | Background ( $\mathcal{B}$ )                   | Set of general informational items about a person. In this study, the term “background” takes on a broader meaning. It may refer to any general information about a person. Hobbies and interests and personal experience are also regarded as background information. In our empirical study, a number of general background items were selected to be included in the profiling process. Examples are if a person lives alone, gender, highest academic qualification, occupation and computer experience. |

and file sharing characteristics); and (iii) personal particulars (personality, motive and background). Table 1 summarizes the categories and variables used in the framework.

The conceptual framework incorporates three types of relationships:

- **Infer:** This relationship is a direct deduction based on an observable online trace. For example, if a post was published at 1 a.m. on December 31, 2010, it could be inferred that the person who published the post was online around 1 a.m. on that day. The observable online trace is “the timestamp of the post” and the inferred characteristic is “the person was online around the time corresponding to the timestamp.”
- **Influence:** This relationship is based on two facts: (i) an individual’s personal experience and living environment (both play an important role in behavioral development); and (ii) personality, which includes behavioral tendencies consistent over time that affect an individual’s actions [3].
- **Reflect:** Online characteristics and file sharing characteristics are behavioral characteristics that differentiate the behaviors of different people. This relationship is the inverse of influence because an individual’s personality, background and motive influence the individual’s behavior, and the associated behavior can reveal the individual’s personality characteristics.

### 3. Empirical Study of Internet Pirates

Most empirical criminal profiling studies use data gathered from convicted offenders [10, 14]. However, in Hong Kong, the one and only Internet piracy conviction was in the 2005 HKSAR v. Chan Nai Ming case [5]. Due to the single-case pool, the typical convicted offender approach cannot be used.

Our study collected data on Internet pirates by surveying individuals who have shared copyrighted materials with the aid of Chinese-speaking online public forums. This is considered an acceptable method of data collection for two reasons. The first is that in the Chan Nai Ming case, the BitTorrent (BT) file containing the copyrighted materials was published in a public online forum [5]. The second reason is that online public forums provide a convenient platform for users to exchange information, opinions and resources [4].

Online public forums, such as Hong Kong Discuss ([www.discuss.com.hk](http://www.discuss.com.hk)) and Uwants ([www.uwants.com](http://www.uwants.com)), are Internet discussion forums that allow users to have public online conversations by posting messages. According to 2010 statistics [1], Hong Kong Discuss and Uwants were the sixth and ninth most frequently visited web sites in Hong Kong. Table 2 shows the usage statistics of these two forums.

Table 2. Usage statistics for HK Discuss and Uwants.

|                                            | HK Discuss | Uwants     |
|--------------------------------------------|------------|------------|
| Registered users as of February 2011       | 2,018,728  | 2,311,809  |
| Posts as of February 2011                  | –          | 55,823,984 |
| Online users at a particular point in time | 82,513     | 67,950     |
| Maximum number of online users             | 145,753    | 141,240    |
| Posts on a weekday                         | –          | 71,642     |
| Posts on a weekend                         | –          | 113,677    |

### 3.1 Data Collection Methodology

We created a 72-question questionnaire for data collection. The questions were divided into three sections corresponding to the principal variables: (i) online characteristics ( $\mathcal{L}$ ); (ii) file sharing characteristics ( $\mathcal{F}$ ); and (iii) personal particulars ( $\mathcal{PP}$ ), which includes personalities ( $\mathcal{P}$ ), motives ( $\mathcal{M}$ ) and background ( $\mathcal{B}$ ). The category facts ( $\mathcal{T}$ ), which helps infer behavioral characteristics ( $\mathcal{L}$  and  $\mathcal{F}$ ), is omitted because in a self-reporting survey, it is deemed to be more appropriate to ask for the behavioral characteristics of an individual. However, self-reporting surveys have certain limitations, which are discussed later.

The questionnaires were posted online and were completed in an anonymous manner. It was important to ensure anonymity because sensitive information was solicited about a subject's uploading and downloading of copyrighted materials. The assurance of strict anonymity and confidentiality encouraged subjects to provide honest information, which contributed to a more accurate survey.

Online public forums were used to solicit participation in the survey. Messages were posted on three public forums: Hong Kong Discuss, Uwants and FDZone ([fdzone.forum.org](http://fdzone.forum.org)). Only those individuals who confirmed that they were 18 years or older were allowed to participate in the survey.

Over a four-month period, the system received a total of 263 questionnaire submissions, of which 133 were complete. Eliminating the questionnaires completed by individuals who had never shared copyrighted materials on the Internet left 114 complete questionnaires for analysis.

**Participant Statistics** The term “case” is used to describe the collection of responses in a complete questionnaire. The following are the background statistics about the 114 cases:

- **Gender:** 14.9% (17) female; 85.1% (97) male.

- **Age:** 27.2% (31) 18 to 21; 63.2% (72) 22 to 30; 9.6% (11) over 30.
- **Residence:** 69.3% (79) live in Hong Kong; 24.6% (28) live in other Chinese-speaking areas (Macau, Taiwan, Mainland China); 6.1% (7) live in other places (United States, Malaysia, Australia).
- **Marital Status:** 80.7% (92) are single; 19.3% (22) are married or in a stable relationship.
- **Type of Sharing:** 64.0% (73) have uploaded and downloaded copyrighted materials; 36.0% (41) have downloaded but not uploaded copyrighted materials.
- **Academic Qualifications:** 2.6% (3) completed (or will complete) Form 3 or lower; 25.4% (29) completed (or will complete) Form 5 to Form 7; 71.9% (82) completed (or will complete) a college degree or higher.
- **Employment Status:** 34.2% (39) are students; 53.5% (61) have stable employment; 4.4% (5) have unstable employment; 7.9% (9) are unemployed.

It is worth noting that the participants were not selected through random sampling. The statistics drawn from the data pool are not used to estimate the population parameters. In other words, statistical inference is not used to draw a conclusion such as the percentage of the population of Internet pirates that are male. Instead, the goal is to identify correlations among different sets of characteristics of an individual.

### 3.2 Data Preparation

Internet piracy is the copying and distribution of copyrighted files using the Internet. This definition leads to three types of Internet pirates:

- **Uploader:** Not involved in downloading; only involved in uploading (or distributing) copyrighted files
- **Downloader:** Not involved in uploading; only involved in downloading (or copying) copyrighted files
- **Sharer:** Involved in uploading and downloading copyrighted files

According to the survey results, 64% (73) of the participants are sharers, 36% (41) are downloaders, and none are purely uploaders.

The nature of Internet piracy acts leads us to divide the cases into two groups: Group I (Downloaders) and Group II (Sharers). Another

Table 3. Online characteristics.

| Label    | Characteristic                                 |
|----------|------------------------------------------------|
| $L_2$    | Online for over four hours per day on weekends |
| $L_4$    | Usually visit Facebook or use MI when online   |
| $L_5$    | Usually go to forums when online               |
| $L_6$    | Visit two or more online forums                |
| $L_7$    | Use the same user name in different forums     |
| $L_8$    | Online whenever possible                       |
| $L_9$    | Online outside working hours                   |
| $L_{10}$ | Online during early hours                      |
| $L_{11}$ | Online during work hours                       |

reason for dividing the cases is that some of the variables, such as the types of the files uploaded, are not relevant to downloaders. If all the cases were to be analyzed together, the results would be more complex and more difficult to interpret.

The final questionnaires comprised 63 variables for Group I and 58 variables for Group II. Tables 3 through 5 define the variables.

### 3.3 Methodology

Multidimensional scaling (MDS) is a statistical technique for identifying underlying patterns or structures in a set of objects. It is useful for measuring the similarities and dissimilarities of objects and displaying the results in the form of geometric representations. One of the major uses of MDS [2] is to reveal the hidden psychological dimensions in data. In the area of criminal profiling, MDS has been used to analyze criminal behavior patterns and offender characteristics [9, 12].

MDS essentially arranges a set of objects (survey variables) on a map where the distance between two objects represents the observed correlation of the corresponding variables [2]. If an object  $A$  is in close proximity to an object  $B$  but far away from an object  $C$ , this suggests that object  $A$  and object  $B$  hold a strong relationship while a weak relationship or no relationship exists with the remote object  $C$ . Interested readers are referred to [2] for details about the mathematical and computational aspects of the MDS algorithm.

This study uses the SPSS statistical tool [11] for MDS analysis. The conceptual framework separates the variables into three sets: (i) online characteristics  $\mathcal{L}$  (Table 3); (ii) file sharing characteristics  $\mathcal{F}$  (Table 4); and (iii) personal particulars  $\mathcal{PP}$  (Table 5).

We focus first on the behavioral characteristics. The sets  $\mathcal{L}$  and  $\mathcal{F}$  were analyzed using MDS. From the two-dimensional results, clusters of variables in close proximity were identified as the observed patterns of

Table 4. File sharing characteristics.

| Label         | Characteristic                                                           |
|---------------|--------------------------------------------------------------------------|
| $F_2$         | Mostly upload songs                                                      |
| $F_3$         | Mostly upload movies                                                     |
| $F_4$         | Mostly upload television dramas                                          |
| $F_5$         | Mostly upload anime                                                      |
| $F_7$         | Mostly upload software or games                                          |
| $F_8$         | Mostly upload pornography                                                |
| $F_9$         | Mostly download songs                                                    |
| $F_{10}$      | Mostly download movies                                                   |
| $F_{11}$      | Mostly download television dramas                                        |
| $F_{12}$      | Mostly download anime                                                    |
| $F_{13}$      | Mostly download e-books                                                  |
| $F_{14}$      | Mostly download software or games                                        |
| $F_{15}$      | Mostly download pornography                                              |
| $F_{16}$      | Use two or more file sharing systems                                     |
| $F_{17}$      | Use four or more file sharing systems                                    |
| $F_{18}$      | Buy CDs/DVDs and convert them to other digital formats for sharing       |
| $F_{19}$      | Download files from other sources and make changes before uploading them |
| $F_{21}$      | Try to do something technical to avoid being caught                      |
| $G2 : F_{22}$ | Try to do something to avoid being caught                                |
| $G1 : F_{22}$ | Try to do something simple to avoid being caught                         |
| $F_{23}$      | Never did anything to avoid being caught                                 |
| $F_{24}$      | Post or get files from Hong Kong forums or websites                      |
| $F_{25}$      | Post or get files from Mainland Chinese forums or websites               |
| $F_{26}$      | Post or get files from foreign forums or websites                        |
| $F_{27}$      | Short seeding period                                                     |
| $F_{28}$      | Arbitrary seeding period                                                 |
| $F_{29}$      | Mostly upload movies                                                     |
| $F_{30}$      | Started using BT as a downloader for three years or more                 |
| $F_{34}$      | Download using BT regularly                                              |
| $F_{35}$      | Download using BT irregularly                                            |
| $F_{39}$      | Have a regular uploading time when using BT                              |
| $F_{41}$      | Use uploading tools to upload to a few locations at a time               |
| $F_{42}$      | Use a single file hosting service every time                             |
| $F_{43}$      | Use different file hosting services                                      |
| $F_{44}$      | Pay for file hosting services                                            |
| $F_{45}$      | Only share files with a small group of people                            |
| $F_{46}$      | Upload many files to file hosting services                               |
| $F_{47}$      | Download many files from file hosting services                           |
| $F_{48}$      | Upload files to file hosting services irregularly                        |
| $F_{50}$      | Download files from file hosting services irregularly                    |
| $F_{51}$      | Download files from file hosting services regularly                      |
| $F_{52}$      | Pick uploading time arbitrarily                                          |
| $F_{53}$      | Have a regular uploading time using file hosting services                |
| $F_{54}$      | Download many files using eDonkey                                        |
| $F_{55}$      | Download many files using Foxy                                           |

Table 5. Personal particulars.

| Label          | Particular                                                       |
|----------------|------------------------------------------------------------------|
| $PP_1$         | Live in Hong Kong                                                |
| $PP_2$         | Live in another Chinese speaking region                          |
| $PP_4$         | Female                                                           |
| $PP_5$         | Male                                                             |
| $PP_6$         | Age 18 to 21                                                     |
| $PP_7$         | Age 22 to 30                                                     |
| $PP_8$         | Age over 30                                                      |
| $PP_{10}$      | Married or in a relationship                                     |
| $PP_{12}$      | Highest academic qualification is Form 5 to Form 7               |
| $PP_{13}$      | Highest academic qualification is a college degree or higher     |
| $PP_{14}$      | Good at school                                                   |
| $PP_{15}$      | Not good at school                                               |
| $PP_{16}$      | Enjoy studying                                                   |
| $PP_{17}$      | Do not enjoy studying                                            |
| $PP_{18}$      | Student                                                          |
| $PP_{19}$      | Stable employment                                                |
| $PP_{21}$      | Unemployed                                                       |
| $PP_{23}$      | Used computers before age 16                                     |
| $PP_{24}$      | Confident about computer knowledge                               |
| $PP_{25}$      | Employed in the computer sector                                  |
| $PP_{26}$      | Positive personality                                             |
| $PP_{27}$      | Introverted or silent personality                                |
| $PP_{28}$      | Outgoing personality                                             |
| $PP_{29}$      | Easy-going personality                                           |
| $PP_{30}$      | Live alone                                                       |
| $PP_{31}$      | Have a white-collar job                                          |
| $PP_{32}$      | Have working parent(s)                                           |
| $PP_{33}$      | Have a happy family                                              |
| $G2 : PP_{34}$ | Try to hide file sharing activities from family and friends      |
| $G1 : PP_{34}$ | Try to hide file sharing activities from family                  |
| $G1 : PP_{35}$ | Try to hide file sharing activities from friends                 |
| $PP_{36}$      | Reason: Want to make contributions to others                     |
| $PP_{37}$      | Reason: Want peer recognition (earn points in discussion forums) |
| $PP_{40}$      | Intend to hide identity and avoid being traced                   |

behavioral characteristics. To examine how these behavioral characteristics are related to an individual's personal particulars, an MDS analysis was performed again on the variables of each cluster with the variables in  $\mathcal{PP}$  for each group.

### 3.4 Results

This section describes the results of MDS profiling. The profiling revealed six types of downloaders and six types of sharers.

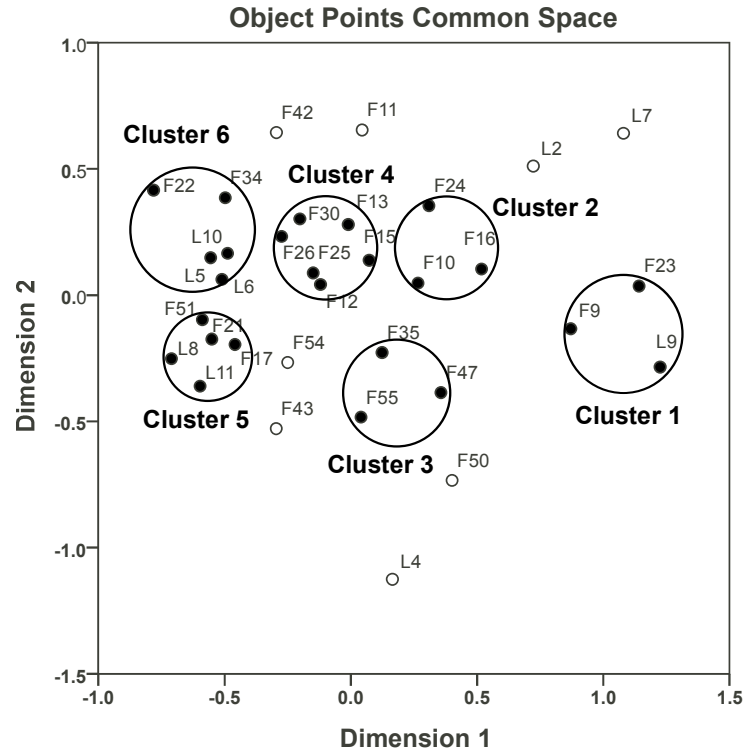


Figure 2. Group I (Downloaders): MDS of  $L_i$  and  $F_i$ .

**Group I (Downloaders).** For Group I, nine variables are related to online characteristics ( $\mathcal{L}$ ), 24 variables are related to file sharing characteristics ( $\mathcal{F}$ ), and 30 variables are related to personal particulars ( $\mathcal{PP}$ ). Figure 2 shows the two-dimensional results of MDS analysis on the sets  $\mathcal{L}$  and  $\mathcal{F}$ .  $K$ -means cluster analysis of the variables identified six clusters that grouped between three to six characteristics.

Running an MDS analysis on each cluster with set  $\mathcal{PP}$  individually yielded six two-dimensional representations. The identified behavioral characteristics and the associated personal particulars in each cluster led to the identification of the following six types of downloaders:

- **Type 1 (Music Downloaders):** These individuals are online outside of work hours ( $L_9$ ). They mostly download songs ( $F_9$ ) and do not do anything to avoid being caught ( $F_{23}$ ).

Type 1 downloaders are usually male ( $PP_5$ ) aged 22 to 30 ( $PP_7$ ). They live in Hong Kong ( $PP_1$ ) and have obtained (or will obtain) a college degree or higher ( $PP_{13}$ ). They have used computers since

they were young (before age 16) ( $PP_{23}$ ). They usually have stable jobs ( $PP_{19}$ ) and happy families ( $PP_{33}$ ).

- **Type 2 (Movie Downloaders):** These individuals use two or more file sharing systems ( $F_{16}$ ) to download movies ( $F_{10}$ ), and post or get files from local forums or web sites ( $F_{24}$ ).

Type 2 downloaders do not have a specific set of personal particulars.

- **Type 3 (P2P Downloaders):** These individuals do not use BT regularly ( $F_{35}$ ). Instead, they download files from file hosting services ( $F_{47}$ ) and by using Foxy ( $F_{55}$ ).

Type 3 downloaders do not have a specific set of personal particulars.

- **Type 4 (Anime/e-Book/Porn Downloaders):** These individuals mainly download anime ( $F_{12}$ ), comics, e-books ( $F_{13}$ ) and pornography ( $F_{15}$ ). They usually post or get files from the Chinese mainland ( $F_{25}$ ) and foreign ( $F_{26}$ ) forums or web sites. They are experienced BT downloaders, having used it for three years or more ( $F_{30}$ ).

Type 4 downloaders do not have a specific set of personal particulars.

- **Type 5 (Cyberlocker Downloaders):** These individuals download files from file hosting services ( $F_{51}$ ) on a regular basis (e.g., every day or every weekend). They try to go online whenever possible ( $L_8$ ), even during working hours ( $L_{11}$ ). They use four or more file sharing systems ( $F_{17}$ ) and use advanced techniques (e.g., onion routers or proxy servers) to avoid getting caught ( $F_{21}$ ).

Type 5 downloaders have outgoing personalities ( $PP_{28}$ ). They usually live alone ( $PP_{30}$ ) in a Chinese-speaking region outside Hong Kong ( $PP_2$ ). Typically, they are over 30 ( $PP_8$ ) and are unemployed ( $PP_{21}$ ). They are cautious about their online identities and actively avoid being traced by others ( $PP_{40}$ ). They tend to conceal their file sharing activities from family ( $PP_{34}$ ) and friends ( $PP_{35}$ ).

- **Type 6 (Forum Downloaders):** These individuals usually visit forums when they are online ( $L_5$ ) and visit more than two online forums ( $L_6$ ). They usually go online during the early hours ( $L_{10}$ ). They use BT to download regularly ( $F_{22}$ ) and use simple methods (e.g., disconnecting from the swarm right after a download is complete) to avoid getting caught ( $F_{34}$ ).

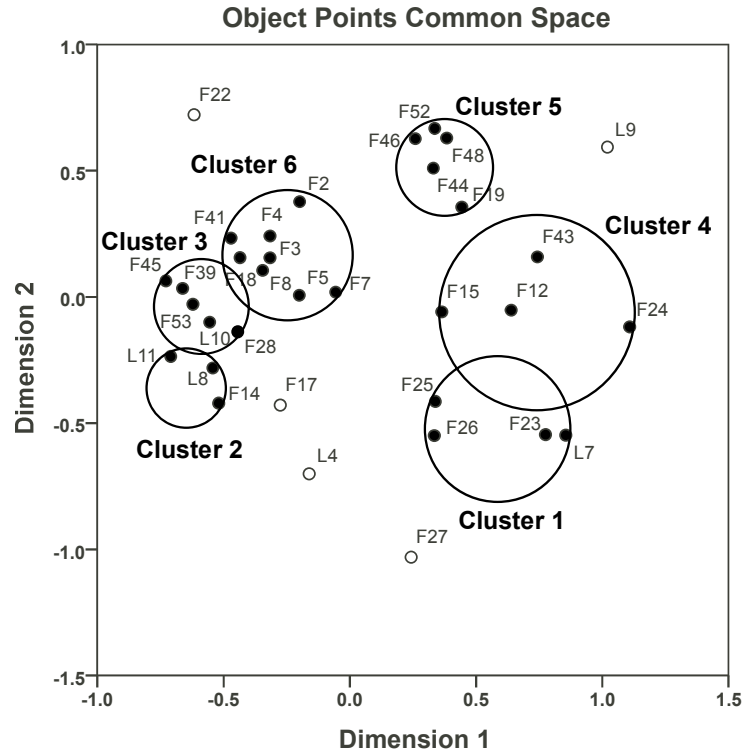


Figure 3. Group II (Sharers): MDS of  $L_i$  and  $F_i$ .

Type 6 downloaders do not have a specific set of personal particulars.

**Group II: Sharers.** For Group II, six variables are related to online characteristics ( $\mathcal{L}$ ), 28 variables are related to file sharing characteristics ( $\mathcal{F}$ ), and 24 variables are related to personal particulars ( $\mathcal{PP}$ ).

Figure 3 shows the two-dimensional MDS analysis results for the sets  $\mathcal{L}$  and  $\mathcal{F}$ .  $K$ -means cluster analysis identified six clusters that grouped between three and eight characteristics.

Six two-dimensional representations were obtained by running an MDS analysis on each cluster with the set  $\mathcal{PP}$ . The six clusters led to the identification of the following six types of sharers:

- **Type 1 (Forum Sharers):** These individuals use the same user name in different online forums ( $L_7$ ) and do not take any measures to avoid getting caught ( $F_{23}$ ). They usually post files to or get files from Mainland Chinese ( $F_{25}$ ) and foreign ( $F_{26}$ ) forums or web sites. Type 1 sharers do not have a specific set of personal particulars.

- **Type 2 (P2P (Public) Sharers):** These individuals go online whenever possible ( $L_8$ ), even during working hours ( $L_{11}$ ). They mostly download software or games ( $F_{14}$ ).

Type 2 sharers usually live alone ( $PP_{30}$ ). They do not do well in school ( $PP_{15}$ ) and do not enjoy studying ( $PP_{17}$ ). They share files because they seek peer recognition ( $PP_{37}$ ) (e.g., earn points in discussion forums).

- **Type 3 (P2P (Private) Sharers):** These individuals go online during the early hours ( $L_{10}$ ). They usually share their files with a small group of people ( $F_{45}$ ). When they upload files using BT, they usually stay in the swarm for arbitrary periods of time ( $F_{28}$ ). They have regular uploading times when using BT ( $F_{39}$ ) and file hosting services ( $F_{53}$ ).

Type 3 sharers have easy-going personalities ( $PP_{29}$ ). They do not do well in school ( $PP_{15}$ ) and do not enjoy studying ( $PP_{17}$ ). They usually live alone ( $PP_{30}$ ) and share files to gain peer recognition ( $PP_{37}$ ).

- **Type 4 (Anime/e-Book/Porn Sharers):** These individuals upload anime ( $F_{12}$ ) and pornography ( $F_{15}$ ). They usually post files to or get files from local forums or web sites ( $F_{24}$ ). They use multiple file hosting services for sharing ( $F_{43}$ ).

Type 4 sharers usually live in Hong Kong ( $PP_9$ ). They do not have any other specific personal particulars.

- **Type 5 (Cyberlocker Sharers):** These individuals download files from other sources and make some changes to the files before uploading them ( $F_{19}$ ). They do not regularly ( $F_{48}$ ) upload many files to file hosting services ( $F_{46}$ ) and are willing to pay for better transfer speeds ( $F_{44}$ ). They usually do not have regular uploading times ( $F_{52}$ ).

Type 5 sharers do not have a specific set of personal particulars.

- **Type 6 (Media-Shifter Sharers):** These individuals upload most file types (songs ( $F_2$ ), movies ( $F_3$ ), TV dramas ( $F_4$ ), anime ( $F_5$ ), software and games ( $F_7$ ), and pornography ( $F_8$ )). They usually buy CDs/DVDs and convert them to other digital formats for file sharing ( $F_{18}$ ). They use uploading tools to upload to a few locations at the same time ( $F_{41}$ ).

Type 6 sharers have easy-going personalities ( $PP_{29}$ ). They usually live alone ( $PP_{30}$ ) and try to hide their file sharing activities from

family and friends ( $PP_{34}$ ). They share files with others because they seek peer recognition ( $PP_{37}$ ) and want to make contributions to others ( $PP_{36}$ ).

## 4. Discussion

This section discusses the limitations of the survey and the accompanying analysis, along with plans for future research to address the limitations.

### 4.1 Limitations

The results of this study suggest a tendency or higher probability that relationships exist between behavioral characteristics and personal particulars. The taxonomy of Internet pirates described in this paper is neither exclusive nor exhaustive. It was derived from the responses of 114 survey participant and, as such, has the following limitations:

- **Self-Reported Data:** The online survey obtained self-reported data from volunteers. There is the possibility that the reported data may be biased. Also, it is possible that some survey participants were dishonest or were incapable of providing accurate answers to certain questions.
- **Participant Pool:** Only individuals aged 18 or above were allowed to participate in the survey. We anticipate that a large proportion of Internet pirates are younger than 18 (e.g., students from secondary schools and even primary schools). Clearly, the taxonomy would not apply to these individuals. Another issue is that the survey participants were clearly limited to Internet pirates who were interested in helping the study.
- **Cultural Dependency:** In general, the results of criminal profiling studies are highly culture-dependent. This is due to the fact that the environment surrounding an individual's life influences the individual's perceptions of issues. Therefore, the results of this study may not be applicable to non-Chinese cultures.

### 4.2 Future Directions

Despite the limitations of the study, the findings indicate that there are two typologies, downloaders and sharers. Future research in this area should perform empirical tests based on these typologies and the conceptual framework for profiling Internet pirates. For example, using

the typologies to predict actual Internet pirating behaviors could help determine the accuracy of the typologies.

Additionally, the applicability of the conceptual framework to other crimes, such as Internet auction fraud, should be examined.

From the design perspective, future studies should include participants from other online communities (e.g., social networks), convicted cyber criminals and even non-criminals as a control group. The use of larger and diverse populations would significantly enhance the resulting data analysis and provide better profiles of Internet criminals. Also, MDS analyses and *K*-means clustering comparisons could help provide quantitative evaluations of the taxonomies.

## 5. Conclusions

The conceptual framework for profiling Internet pirates presented in this paper incorporates six fundamental variables: observable online traces, online characteristics, file sharing characteristics, personality, motive and background. These variables are further divided into three conceptual categories: facts, behaviors and personal particulars. In the conceptual framework, behaviors can be influenced by personal particulars and inferred by observed facts. Similarly, an individual's behaviors reflect his or her personal particulars.

MDS and clustering analyses of the results of the survey of 114 Internet pirates yielded a taxonomy comprising six types of downloaders and six types of sharers, each with different sets of behavioral characteristics. Some of the downloader and sharer types are also characterized by their personal particulars.

Criminal profiling techniques have been shown to be immensely useful in traditional criminal investigations. Profiling frameworks and taxonomies for specific types of cyber crimes would aid forensic scientists and practitioners in devising effective and efficient investigation plans. As Internet piracy and other cyber crimes become even more rampant, it is crucial that law enforcement agents and other officials who combat these crimes have better understanding of the underlying criminal behaviors.

## References

- [1] Alexa Internet, Top sites by countries: Hong Kong, San Francisco, California ([www.alexa.com/topsites/countries/HK](http://www.alexa.com/topsites/countries/HK)), 2012.
- [2] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, New York, 2005.

- [3] A. Caspi, B. Roberts and R. Shiner, Personality development: Stability and change, *Annual Review of Psychology*, vol. 56, pp. 453–484, 2005.
- [4] K. Chow, K. Cheng, L. Man, P. Lai, L. Hui, C. Chong, K. Pun, W. Tsang, H. Chan and S. Yiu, BTM: An automated rule-based BT monitoring system for piracy detection, *Proceedings of the Second International Conference on Internet Monitoring and Protection*, 2007.
- [5] Court of Final Appeal of the Hong Kong Special Administrative Region, Final Appeal No. 3 of 2007 (Criminal), between Chan Nai Ming and HKSAR, Hong Kong, China ([legalref.judiciary.gov.hk/lrs/common/ju/ju\\_frame.jsp?DIS=57111](http://legalref.judiciary.gov.hk/lrs/common/ju/ju_frame.jsp?DIS=57111)), 2007.
- [6] Envisional, An Estimate of Infringing Use of the Internet, Technical Report, Cambridge, United Kingdom ([documents.envisional.com/docs/Envisional-Internet\\_Usage-Jan2011.pdf](http://documents.envisional.com/docs/Envisional-Internet_Usage-Jan2011.pdf)), 2011.
- [7] N. Fisk, *Understanding Online Piracy: The Truth About Illegal File Sharing*, ABC-CLIO, Santa Barbara, California, 2009.
- [8] J. Gantz and J. Rochester, *Pirates of the Digital Millennium: How the Intellectual Property Wars Damage Our Personal Freedoms, Our Jobs and the World Economy*, FT Prentice Hall, Upper Saddle River, New Jersey, 2005.
- [9] E. Hickey, *Serial Murderers and Their Victims*, Wadsworth, Belmont, California, 2006.
- [10] R. Kocsis, *Criminal Profiling: Principles and Practice*, Humana Press, Totowa, New Jersey, 2006.
- [11] R. Levesque, *SPSS Programming and Data Management: A Guide for SPSS and SAS Users*, SPSS, Chicago, Illinois, 2007.
- [12] G. Palermo and R. Kocsis, Offender profiling: An introduction to the sociopsychological analysis of violent crime, *Journal of the American Academy of Psychiatry and the Law*, vol. 33(3), pp. 421–423, 2005.
- [13] M. Rogers, The role of criminal profiling in the computer forensics process, *Computers and Security*, vol. 22(4), pp. 292–298, 2003.
- [14] B. Turvey (Ed.), *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, Academic Press, Oxford, United Kingdom, 2012.