Forensic Science International 268 (2016) 103–115

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

The study of doping market: How to produce intelligence from Internet forums

Thomas Pineau*, Adrien Schopfer*, Lionel Grossrieder, Julian Broséus, Pierre Esseiva, Quentin Rossy

School of Criminal Justice, University of Lausanne, 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 28 July 2016
Received in revised form 14 September 2016
Accepted 16 September 2016
Available online 27 September 2016

Keywords:
Illicit market
Doping
Scraping
Intelligence-led screening
Operational monitoring
Forums

ABSTRACT

Despite the predominant role played by Internet in the distribution of doping substances, little is currently known about the online offer of doping products. Therefore, the study focuses on the detection of doping substances and suppliers discussed in Internet forums. It aims at having a comprehensive understanding of products and sellers to lead an operational monitoring of the online doping market.

Thirteen community forums on the Internet were investigated and one million topics were extracted with source code scrappers. Then, a semantic analysis was conducted with a semi-automatic process to classify the relevant words according to doping matters. Additionally, the ranking of doping products, active substances and suppliers in regards to the number of contributors to the forums were established and analyzed over time. Finally, promotion methods of suppliers were evaluated.

The results show that anabolic androgenic steroids, used to enhance body image and performance, are the most discussed type of products. A temporal analysis illustrates the stability of the most popular products as well as the emergence of new products such as peptides (e.g. CJC-1295). 327 suppliers were detected, mostly with dedicated websites or direct sales by e-mail as selling methods.

Globally, the implemented methodology shows its ability to detect products and suppliers as well as to follow their temporal trends. The intelligence will serve the definition of online monitoring strategies (e.g. the selection of appropriate keywords). Additionally, it also allows the adjustment of customs inspection strategies and anti-doping analysis by monitoring the popular and emerging substances.

1. Introduction

The online market of doping products occupies a crucial place in the distribution and accessibility of products [1,2]. The study of illicit online markets is a problem which can be described in regards to usages of the Internet to support the market. Indeed, online services may be used for several aims. They may be classified in three related processes: (1) the promotion process which uses online services to advertise and ease the accessibility to marketplaces, (2) the selling process which takes places in multiple online environments, and (3) the discussion process which aims at sharing information and reviews about products and sellers. If the selling process is central to the study of illicit online markets, promotion and discussion processes are dependent issues that are also critical to tackle the global phenomenon. Indeed, the discovery of illicit online markets requires to set up an appropriate research strategy. This article focuses on the use of information gathered during the discussion process to guide the search for online marketplaces.

Access to Internet sale spaces can take different paths, whether on the world wide web including indexed websites (surface web) and not indexed websites (deep web) or overlay networks (darknet) such as TOR.

For instance, search engines or directories can index webpages. Forums, social networking services, sharing websites or spam may also be used to broadcast hyperlinks to webpages or directly promote products [3,4]. Thus, the research of these promotion means requires the identification of third-party services and information-sharing websites as well as the definition of relevant keywords to perform target researches. Indeed, doing queries on different data sources, whether they are search engines or the content of a website (e.g. social networking services) is based on the use of keywords related to the problem to detect. Therefore, a

* Corresponding authors.
E-mail addresses: Thomas.Pineau@unil.ch (T. Pineau), Adrien.Schopfer@unil.ch (A. Schopfer), Lionel.Grossrieder@unil.ch (L. Grossrieder), Julian.Broseus@unil.ch (J. Broséus), Pierre.Esseiva@unil.ch (P. Esseiva), Quentin.Rossy@unil.ch (Q. Rossy).

http://dx.doi.org/10.1016/j.forsciint.2016.09.017
0379-0738/

set of appropriate keywords according to the targeting content should be chosen to maximize the efficiency of the process. This raises the question of keywords selection. These keywords can be defined based on those used by already known sale websites (e.g. the keywords tag in the source code or the textual content from the webpage). This strategy is limited because it assumes that a sample of websites is already known and it is solely focused on indexed websites. Moreover, the keywords obtained are not necessarily suitable for other online communication spaces and unknown marketplaces. Other sources of information should thus be used to guide the selection of keywords. The chosen approach in this research stems from the study of online information sharing spaces, in particular on community forums. The assumption is that the most discussed terms found on community forums are the most suited to search websites or sale announcements. Deep web and darknet forums were not taking into account in this study. Indeed, doping is not considered as a very priority issue by law enforcements and can be freely discussed on the surface web unlike pedophilia for example.

The described approach is comparable to the process named "Intelligence-led crime scene processing" formalized by Ribaux et al. [5,6]. This process aims at guiding the collection of forensic traces on a crime scene in the light of the knowledge previously accumulated about criminal activities. In particular, this knowledge is produced from the intelligence process which aims to detect and follow the evolution of crime problems [7]. The process is decomposed according to the following steps: (1) data collection; (2) integration in a structured memory; (3) detection of specific problems (through pattern recognition); (4) analysis of the detected problems; (5) intelligence production to guide the decision-making; and (6) impact assessment of the actions taken. The Fig. 1, adapted from Ribaux [7], express the steps of the intelligence process in the case of illicit markets on the Internet.

The arrow "intelligence-led screening", on the Fig. 1, describes the research process of selling spaces on the Internet. This process is guided by the knowledge acquired during the data analysis of the crime problem. In this case, the analysis of community forums is used as a source of knowledge to identify and select the keywords needed to detect the selling spaces of doping products.

Forums analysis involves setting up a process that can be described by the previous approach. Thus, these discussion spaces are integrated in the global monitoring process of illicit markets on the Internet. The acquisition phase (1) includes a research of forums containing discussions about doping products, followed by an acquisition of their content. The acquired raw data includes general information about posts, members as well as the promotional content (i.e. advertising, hyperlinks). The integration phase (2) aims at structuring the relevant information by a semantic analysis of the topics to extract information on products, active substances and suppliers. The detection phase (3) is mainly on the identification of the most discussed terms and the identification of suppliers and their contact means. The analysis step (4) focuses on the temporal analysis of detected products and sellers in order to evaluate their trends and follow their evolution. Additionally, activity patterns linked to selling contact means such as dedicated websites or e-mail are analyzed. Finally, (5) the intelligence inferred on selling and consumption practices aims at guiding decision-making. In particular, to define relevant keywords to search the Internet for illicit online marketplaces. Moreover, it allows the adjustment of custom inspections strategies and anti-doping analysis by monitoring the popular and emerging substances.

2. Previous studies

A few studies have already examined the market of doping products on the Internet. Krug et al. [8] have described the products that can be found on the market in Germany. The products were seized by customs between 2010 and 2013. Steroids were the major kind of seizures with 87.5% of the total number. In this category, esters of testosterone, mehtandienone, boldenone and trenbolone were the main substances. Donati [1] pointed that the world trafficking of doping products, in particular anabolic agents, is sharply increasing through the Internet. In 2000, a first study was conducted to highlight the easiness to buy doping products online [9]. However, the search strategy was not clearly described. Additionally, Cordaro et al. [10], who did not provide information on their strategy of keywords selection to detect websites, have investigated 30 marketplaces selling anabolic agents to evaluate their availability on the Internet. They concluded that an average of 52 products were available with esters of nandrolone, methandrostenolone and of testosterone in the highest proportion. Clement et al. [11] studied the diffusion of nonprescription steroids on the Internet. They selected 5 keywords related to the brand names of steroid products. They classified 100 hyperlinks found with Google in 2006 to show the proportion

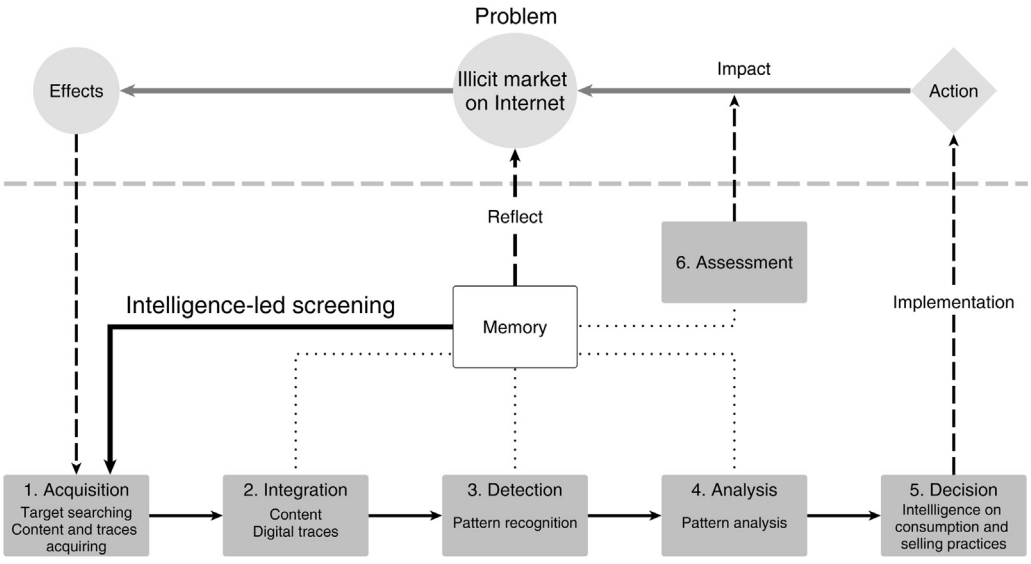

Fig. 1. Operational monitoring process in the case of illicit markets on the Internet.