Webscraping as an Investigation Tool to Identify Potential Human Trafficking Operations in Romania

Ruth McAlister
Ulster University
Shore Road
Newtownabbey
+44 (0)2890 368063
r.mcalister@ulster.ac.uk

ABSTRACT
Information communication technology has enabled criminals to remain distant from the crimes they commit with reduced risk. However, by moving this underground criminal activity online, digital evidence of communication with members of the crime group, and also victims, presents an interesting research opportunity into human trafficking and may reveal actionable information for law enforcement agencies. Specifically, this research paper investigates whether a webscraping tool could be employed to gather intelligence on organized crime groups at the recruitment stage of the trafficking operation as a means to understand their *modus operandi.* Preliminary findings presented in this paper indicate that the UK is a popular destination country for job advertisements hosted in Romania and further analysis will be undertaken to identify if there are in fact indicators of trafficking evident in these identified websites.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy - *Issues - abuse and crime involving computers, ethics, human safety, use/abuse of power.*

General Terms
Security, Human Factors, Languages.

Keywords
Organized Crime, Human Trafficking, Webscraping, Criminology.

1. INTRODUCTION
Towards the end of 2014 over 3 billion people were connected to the Internet (IWS, 2015). Not only has this connectivity increased exponentially in recent years, but also the technical capabilities of Information Communication Technologies (ICTs) with devices nowadays having increased storage and the capacity to transmit larger amounts of data at faster speeds than ever before. However, the increasing ubiquity of the Internet and our dependency on communication has left us vulnerable to crimes that take place in cyberspace. This is one reason why cybercrime is becoming an increasingly relevant component of criminological research (Williams, 2008). The theoretical, empirical and technical aspects of cybercrime have been the focus of study for researchers within a variety of disciplines. One common feature of this work regardless of the discipline is the recognition that the lack of physical boundaries and the removal of traditional jurisdictional demarcations allow perpetrators to commit crime with little fear of judicial sanctions (Britz, 2013). Not only does this facilitate criminal enterprise but it also serves to complicate its control. This ability of cybercriminals to transcend traditional geographic boundaries presents criminal justice systems across the world with their greatest challenge (Williams, 2008).

WebSci '15, June 28 - July 01, 2015, Oxford, United Kingdom
 ISBN 978-1-4503-3672-7/15/06…$15.00

A considerable advantage offered to criminals by new ICT is the relative ease of anonymity and disguise. They are able to send communications over the Internet using proxy servers, ToR, a virtual private network (VPN), multihoming, using pseudonyms and registering websites abroad to mask their identities (EUROPOL, 2014). This has enabled criminals to remain distant from the crimes they commit with reduced risk. Shelley (2011) acknowledges that these enhanced communications have facilitated the growth of human trafficking, with Dixon (2013:39) also stating that "traffickers look for newer technologies to stay a step ahead of law enforcement". It is unclear however whether the use of new technologies has *increased* the number of victims being trafficked. What is known is that the use of new technologies has made the trafficking of persons for organized crime groups (OCGs) easier to perform through all three stages of the trafficking chain; recruitment, transportation and exploitation (EUROPOL, 2014).

2. HUMAN TRAFFICKING AND THE INTERNET
Trafficking can manifest itself both within and between countries with fraud, force, threat and deception lying at the heart of trafficking. Walker & Hunt (2009) identify that victims of human trafficking often believe that they are taking legitimate jobs such as waitressing, childcare, or modeling work, only to find out when they arrive in the destination country that they have been tricked. A large part of being duped is due to the characteristics of traffickers in terms of the language they use, such as expressing love and admiration for the victim, the forwarding of fake work contracts and ultimately the promise of a better life, with people often trusting the traffickers for a variety of reasons (Walker & Hunt, 2009). Copley's (2013) work also underlines the issue that many women and young people in particular are vulnerable to such advertisements and the admiration expressed to them, however what is not known, or identified in this discussion are the motivations of their crime, nor other methods that traffickers employ in the human trafficking chain. EUROPOL (2014) have

published a comprehensive intelligence report that is considerably more expansive with regard to this. It underlines that although the Internet and smart phone technology provides traffickers with opportunities to reach a broader and more diverse pool of victims through online recruitment on social media and classified websites, it also aids transportation (buying tickets online with stolen credit cards) and the (final) exploitation phase of the process with trafficked women's pictures and contact details placed online.

3. RESEARCH APPROACH

Therefore, this research questions whether bespoke web scraping software could be usefully employed as an investigative tool as a means to combat trafficking at the recruitment stage, which, importantly is before any exploitation has occurred.

Google and Bing search engines have been employed to identify Romanian websites that may contain adverts for erotic and non-erotic jobs in the UK and beyond. Key phrases such as "jobs abroad" and "jobs abroad for women" (for example) relating to this potential employment were translated into Romanian using Google Translate. The website identification process is still ongoing with 15 currently identified.

The webscraping tool has been written to scrape specific information, such as, phone numbers, email addresses, job advert, job location, salary and the date the advert was posted. Scrapy is the application framework that has been used to extract the structured data. Scrapy creates logical spiders that will crawl to the identified websites; the logic of that spider is defined in python 2.7.

anuntul.py

```
# -*- coding: utf-8 -*-
import scrapy

from proj1.items import AnuntulItem

class AnuntulSpider(scrapy.Spider):
    name = "anuntul"
    allowed_domains = ["anuntul.ro"]
    start_urls = (
        'http://www.anuntul.ro/anunturi-locuri-de-munca/in-strainatate/',
        'http://www.anuntul.ro/anunturi-locuri-de-munca/in-strainatate/pag-2/',
        'http://www.anuntul.ro/anunturi-locuri-de-munca/in-strainatate/pag-3/'
    )

    def parse(self, response):
        for sel in response.xpath('//tr/td[@class="text"]'):
            item = AnuntulItem()
            item['desc'] = sel.xpath('a/text()').extract()
            item['link'] = sel.xpath('a/@href').extract()
            item['loc'] = sel.xpath('a[@class="atext"]/strong/text()').extract()
            item['phone'] = sel.xpath('span[@class="contact"]/text()').extract()
            item['email'] = sel.xpath('a[@class="amail"]/text()').extract()
            item['price'] = sel.xpath('tr[@class="raster"]/td[@class="pret"]/span[@class="pret"]
            item['timestamp'] = sel.xpath('span[@class="data"]/text()').extract()
            yield item
```

Figure 1 Spider illustrating detail to be scraped

One website (anuntul.ro) has been successfully scraped so far (as a pilot) see Figure 1. Some preliminary analysis regarding 'destination country' has taken place which is discussed below.

4. PRELIMINARY FINDINGS

In terms of some preliminary analysis, there were 103 job advertisements identified and scraped on anuntul.ro - the destination country reveals that England/UK (Anglia/UK) was the most common, followed by Switzerland (Elvetia) and Germany (Germania) as documented in Figure 2. Further analysis of this website, examining job adverts (including the language, email addresses and phone numbers) will take place later in the research process, when translated, to identify indicators of human trafficking, such as common email or phone number details and also the language used in the adverts.

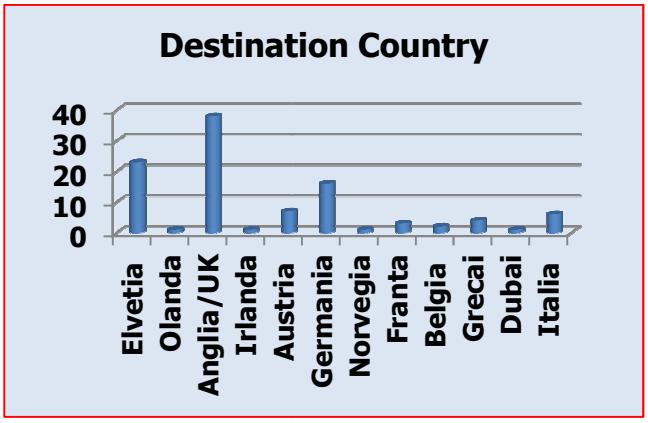

Figure 2 Job destination countries from anuntul.ro.

5. CONCLUSION

This paper has highlighted that criminals are utilizing online technologies for criminal gains. Indeed, the preliminary finding from the scraped website illustrated in this paper reveals that the UK was the most popular destination country for job advertisements. It is therefore suggested that further analysis of the scraped websites could identify indicators of trafficking and offer the potential of filling a void in relation to the *modus operandi* of traffickers, which will not only aid law enforcement investigations and policymaking, but it will also contribute to an under researched area within criminology.

6. REFERENCES

[1] Britz, M. (2013) *Computer Forensics and Cybercrime*. New Jersey: Pearson.

[2] Copley, L. (2013) Neutralising Their Invovlement: Sex Traffickers' Discourse Techniques. *Feminist Criminology*, Vol 9, No 1, pp45-58.

[3] Dixon, J.H.B. (2013) *Human Trafficking and the Internet* (*and other technologies, too)*. Available online at: http://www.americanbar.org/content/dam/aba/publications/judges_journal/2013_win_jj_tech.authcheckdam.pdf [accessed 5/10/14].

[4] EUROPOL (2014) Intelligence Notification. Trafficking in Human Beings and the Internet. Available online at: https://www.europol.europa.eu/sites/default/files/publications/intelligence_notification_thb_internet_15_2014_public.pdf

[5] Internet World Stats (IWS) (2015) *World Internet Users*. Available online http://www.internetworldstats.com/stats.htm [accessed 12/03/15].

[6] Shelley, L. (2011) International trafficking: an important component of transnational crime, in S. Okubo & L. Shelley (eds.) *Human Security, Transnational Crime and Human Trafficking*. Oxon: Routledge

[7] Walker, T.K.L. & Hunt, G. (2009) Understanding Human Trafficking in the United States. *Trauma Violence Abuse*, Vol 10, No 3 pp. 3-29.

[8] Williams, M. (2008) Cybercrime and online methodologies, in R.D. King & E. Wincup (eds.) *Doing Research on Crime and Justice*. Oxford: OUP.