Illuminating the Dark Web

Searching for geotags in dark net markets

Three years ago, Harvard student Eldo Kim emailed a false bomb threat to campus administration, seeking to avoid taking final exams. In order to evade detection, Eldo sent the email through a series of nominally anonymous services: the email originated from a self-deleting email address provided by the service Guerrilla Mail, which he accessed via the well-known anonymity software Tor. Yet within two days Eldo was apprehended by the FBI. His fatal flaw: accessing Tor through Harvard’s wireless network. When the investigators checked who on the network had been using Tor when the threat was sent, a suspect emerged, and a confession soon followed.

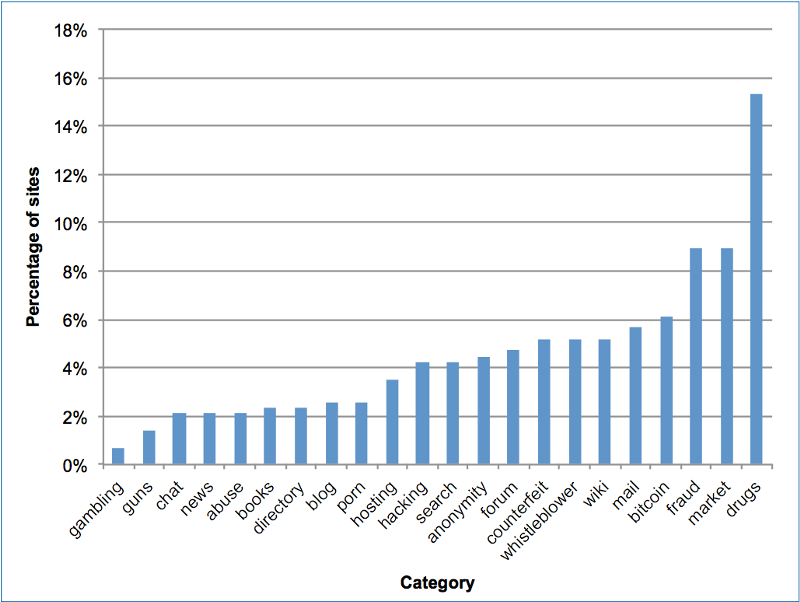
Situations like this show that even the most thorough privacy and anonymization protocols are not sufficient in ensuring anonymity; careless usage of technology is an ever-present vulnerability. Our study sought to analyze the prevalence of a similar privacy oversight in one of the most anonymity-valuing realms on the Internet: dark net markets. Is it possible that the users of such sites are unwittingly revealing their location as they post images to the markets?

Background

The dark web consists of Internet content hosted on dark nets: overlay networks that can only be accessed with special anonymity software such as Tor. Though the terms are often and erroneously used interchangeably, the dark net is distinct from the deep web, which is the collection of content on the Web that is not indexed by search engines (e.g., password-protected websites, or dynamic websites).

In the public view, the dark web — as its name may suggest — is shrouded in mystery, and often seen as an anonymous haven for criminality. This use of the dark web was noted by former British Prime Minister David Cameron at a summit against child pornography in 2014, where he stated that “the dark net is … [not] the normal parts of the Internet that we all use.”

Recent work by Dr. Gareth Owen and Dr. Nick Savage at the University of Portsmouth to quantify and categorize the usage of the dark web determined that about one third of the dark web may be categorized as dealing with Drugs, Markets, and Fraud, three activities that depend on the anonymity provided by the dark web when skirting around the law.



These activities might be best exemplified by the infamous Silk Road 1 marketplace, the “eBay of illegal drugs” that, when shut down by the FBI in 2013, comprised “more than 70% of the online drug market”.

Dark web sites such as Silk Road 1 rely on the anonymous routing provided by Tor, a specialized Internet browser and software suite which was described by the NSA in a leaked top-secret document as “the king of high-secure, low-latency anonymity.” Furthermore, Tor was acknowledged by the British Parliamentary Office of Science and Technology in March 2015 to be “by far the most popular anonymous internet communication system,” having an estimated 2.5 million daily users.

Methods

Our goal was to leverage a longitudinal archive of dark net markets (DNMs) to collect and analyze sale listing images with metadata containing location data. When a digital image is taken, there is additional information, called EXIF data, attached to the image itself. This may include date of capture, make and model of the camera, and other information. Smartphones and many modern cameras can use GPS technology to attach location information to pictures as they are taken (a “geotagged” image). In our investigation, we searched for the presence of these geotags in the images of items for sale on dark net market sites. We also aimed to determine the presence of any notable groupings and patterns in the locations found, both inter- and intra-markets.

The data used for this study was the Black-Market Archives, an collection of DNM web scrapes created by an independent researcher and writer who goes by the handle Gwern. After the widely-publicized takedown of Silk Road 1 by the FBI in December 2013, Gwern began regularly scraping market sites and associated forums in order to create a more comprehensive picture of the constantly shifting DNM landscape. The freely-available archive encompasses 83 different DNMs and 40 associated forums from 2013 to 2015 (exact dates vary from site to site). This amounts to over 44 million files, taking up over 1.5TB of space.

In order to analyze the listing images inside each archive, we first searched for and compiled a list of the file path of all JPEG images to ensure that no file went untested. (Images used for listings were only in the JPEG format; any other image formats — PNG, GIF, etc. — were used for website graphics.) Then, using Python and bash scripts, we checked each image’s EXIF data for longitude or latitude data, saving the coordinates for each geotagged photo and its file path to a text file.

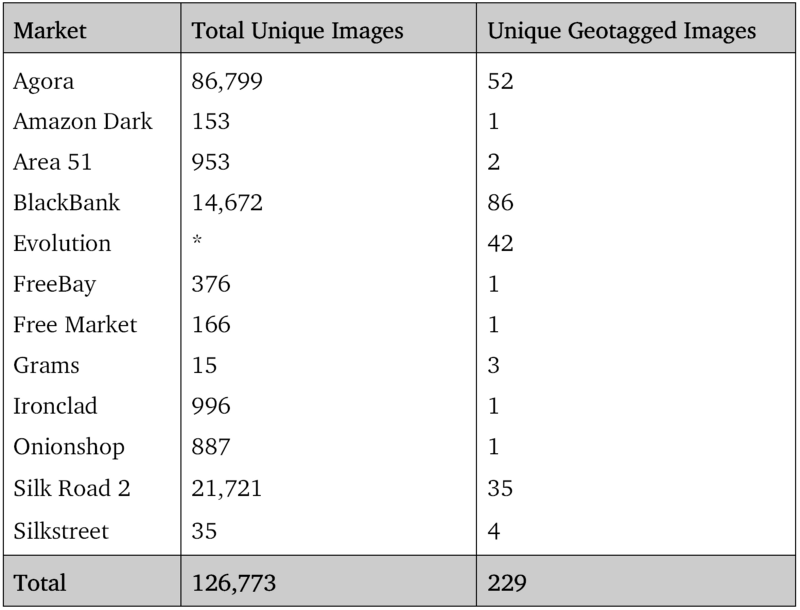
The coordinates were then extracted and analyzed to determine the number of unique coordinates — very commonly, a listing and its associated image would be hosted on a DNM for several days, resulting in the same image, and thus the same coordinates, appearing in our results multiple times.

Results

Out of these markets and forums, we located 2,276 total geotagged images, which after eliminating duplicates available over multiple days, gave 229 total unique images with associated coordinates. The coordinates—with decimals removed from the numbers to protect privacy—can be seen plotted in the map below. (The coordinates may be up to about one mile away from their true location.)

In total, we analyzed 7,522,284 images from the entire DNM archive, representing 223,471\* unique photos. Table 1 presents a summary of markets containing geotagged images:

Table 1. Markets and Geotagged Images



The total number of geotagged images found was far greater than the number of unique images. This is, in a way, a measure of the longevity of online listings — the longer a listing was hosted on the website, the greater the probability that it was re-scraped. While some images appeared only once in the archive, the particular case of the dark net market Area 51 presented several images which appeared 107 times, representing a time span of 215 days.

Agora presents an interesting case: from the archive’s beginning on January 1, 2014 through March 18 of that year, we identified 963 non-unique tagged images — the most of any market in the study. However, after March 18 we found zero images with location data through the end of the archive. The sudden cut off from what was previously a plentiful source of images suggests that the site’s administrators may have discovered this geotag vulnerability and instituted some form of metadata removal after this date.

Of course, EXIF data can not only be stripped but outright modified to point to an entirely separate location. However, patterns in our findings point to the results’ validity, rather than discounting them based on this factor. First, it was common in many cases to observe sites, typically residential, surrounded by 5–10 tagged images separated by a few meters. This suggests the behavior of sellers who are careless on a regular basis, rather than the occasional forgetfulness of not stripping data or purposeful manipulation. We also found several instances of these clusters incorporating listings on multiple sites, pointing to sellers with activities across the dark net and failing to strip their products’ location on any of the sites used.

Conclusions

The dark net markets exist to facilitate the transaction of illicit or regulated goods, such as drugs and firearms. As such, the importance of preserving anonymity of both the seller and buyer is paramount. The existence of the geotagged photos online thus represent a notable failure on two fronts: (1) a failure by the seller and (2) by the website.