Profiling Underground Merchants Based on Network Behavior

Srikanth Sundaresan* Damon McCoy*‡ Sadia Afroz* Vern Paxson*†
* International Computer Science Institute ‡ New York University † University of California Berkeley

***Abstract*—Online underground forums serve a key role in facilitating information exchange and commerce between gray market or even cybercriminal actors. In order to streamline bilateral communication to complete sales, merchants often publicly post their IM contact details, such as their Skype handle. Merchants that publicly post their Skype handle potentially leak information, since Skype has a known protocol flaw that reveals the IP address(es) of a user when they are online. In this paper, we collect Skype handles of merchants from three underground forums—AntiChat, BlackHat World and Hack Forums—and longitudinally monitor their network behavior. Our analysis of their network behavior provides a rich profile of their likely locations, network behavior, work habits, and other dynamics. In particular, we show that these merchants do not frequently use VPN services, and even when they do, they often leak their likely geolocation by also directly using residential and cellular IP addresses.**

I. Introduction

The Internet has spawned a vibrant and active "underground" economy that hosts a range of merchants from across the globe offering semi-legal, or outright illegal services, such as botnets, DDoS, or spamming services. Central to this economy are underground forums that serve a key role in facilitating information exchange and commerce between cybercriminal (or at best gray market) merchants. Little is known about the merchants on these forums: where are they from? What sort of IP access infrastructure do they use? What are their work habits? How careful are they about their privacy?

While these forums are essentially anonymous, merchants publicly post sales threads that include details about the services or goods they sell. Merchants also often include a mechanism for customers to privately contact them to complete the sale; typically IM and VoIP handles. There is normally little risk in terms of privacy and inconvenience when publicly posting handles, since most of these services require a user to approve other users before allowing direct communication and potentially revealing their IP address. An exception is Skype, which, till January 2016, by default had a known flaw that revealed the IP of a user to other Skype users—even those that have not been approved—without them being aware of it.

In this paper, we take advantage of the fact that Skype is a popular service used by underground merchants for their business and use Skype's protocol design to characterize underground forum merchants. We continuously monitor the IP addresses associated with a randomly selected subset of 742 underground merchant Skype handles that were publicly posted to three underground forums: AntiChat, BlackHat World, and Hack Forums. Merchants on these three forums offer a variety of services that range from benign to illegal. We study the infrastructure used by merchants, how much they care about privacy, and how it relates to their geolocation and the type of service they provide. We also use longitudinal monitoring to explore different aspects of their behavior such as their work habits and travel habits. Our findings show that of the 478 active merchant accounts that we profile, only 4.8% (23) consistently use VPNs that protect their actual IP address. Even amongst merchants that engage in outright illegal activity such as botnets and malware, the percentage that consistently protect their IP address by always using a VPN only increases to 12.9% (8) out of 62. This demonstrates that many underground merchants do not consistently employ good operational security, such as using VPNs, to maintain their privacy. This enables us to measure the probable country of residence, working habits, and travel patterns for many of the active merchants that we profiled in our study.

We believe our work is the first of its kind and provides valuable insight into what can be learned about criminal merchants by monitoring their Skype handles and understanding how well they protect their privacy. Our work complements past measurement studies exploring how to monitor postings on underground markets [1], analysis of activities on these market places [2]–[8], how to link accounts based on stylometric techniques [9] and how trust is established within underground forums [10].

We make the following contributions in this paper: We collect 742 merchant handles spanning three forums and a variety of services ranging from benign to illegal, and first characterize the access infrastructure that merchants use, such as residential networks, or hosting services, and how the riskiness of the services they provide influences their use of privacy-preserving techniques such as VPNs; we then explore how well we can infer merchants' likely geolocation even when they use hosting services; we primarily rely on the fact that users typically log in to Skype from multiple devices and not all of them may be behind a VPN. We then develop techniques for detecting when accounts are linked, shared, or there are multiple actors operating different parts of the service. Our techniques demonstrate how to leverage network behavior to construct profiles of underground merchants. The profiles we build expose previously little known

aspects of the underground economy and could be useful for researchers that might want to use these measurements or combine them with other information to better understand underground merchants, or for policy makers to create more effective policies for mitigating these activities.

II. Data Collection

In order to understand the working habits and other network dynamics of underground forum merchants we collect Skype handles from three publicly accessible underground forums: BlackHat World, Hack Forums and AntiChat. We focus on Skype handles, since the Skype protocol enables us to obtain the IP address(es) associated with a Skype handle at a given time. These forums are not a complete set of underground forums; nevertheless, they are major forums for such activities and are very popular, and therefore provide an insightful cross-section of underground forums and merchants operating on these forums. We briefly explain the theme of each underground forum.

BlackHat World (BHW) is predominantly an English language forum and is one of the largest public underground forums. The primary activity of BHW is focused around blackhat Search Engine Optimization (SEO), which is the practice of manipulating search engines to increase the ranking of a site. This is normally achieved by abusive activities, such as spamming blogs and forums with links to the target site. Currently, this forum has a large number of merchants offering primarily abusive products and services to perform or enable blackhat SEO. Hack Forums (HF) is the largest public English language underground forum, with a vibrant commerce section. The merchants in the marketplace section of this forum sell illicit products and services, such as access to botnets, distributed denial of service (DDoS) and malware. AntiChat (AC) is one of the largest public Russian language underground forums; its merchants sell products and services similar in nature to HF.

Limiting our study to merchants that post their Skype handle introduces a potential bias. To understand this bias, we counted how many times the words "skype", "icq", "jabber" and "email" appear in the forums. In BlackHat World and Hack Forums, "skype" is more popular than other communication methods, making up 40% and 77% of all of those words' occurrences, respectively. "icq" predominates in AntiChat (53%), compared to 15% for "skype".

A. Skype handle extraction

We scraped subforums on each of these sites that focus on selling illicit products and services. To extract Skype handles, we stripped off all punctuation from the posts, converted all words into lower case and retrieved the non-dictionary word followed by the word "skype". In addition, we manually extracted handles from a random subset of image based sales postings. We make no claim as to the completeness or representativeness of this set of underground merchant Skype handles; however we believe that it provides a cross-section of merchants from three major forums. Table I shows a breakdown of these handles by forum and the random subset of these handles that we actively monitored due to resource constraints. We monitored 742 handles across the three forums, out of which 478 handles were "active" — meaning we were able to record a "clean" IP address. We explain how we clean our data in Section II-C. Our merchant set spans 62 countries, 450 ISPs, and 11504 IP addresses. The number of IP addresses associated per user is high because of multiple reasons. We use the term "merchant" to refer to a single handle. However, we note that a single ID might actually be used by multiple individuals (*e.g.*, in the case of a franchise operation), or vice-versa; a single individual might operate multiple handles. We also study both these possibilities in this paper. Merchants tend to log in from multiple locations like their home ISP, mobile phone, or hosting services. The number of IP address per merchant is also particularly high in countries such as India and Pakistan. This is likely due to IPv4 shortages in those countries, which leads to very short DHCP leases and frequent reallocations.

We then had a domain expert look at the sale post and manually categorize active merchants[1] based on the type of services or products that they were selling. We defined three categories based on risk: *illegal*—goods and services that might result in criminal charges, such as selling malware[2], renting botnets and offering to perform attacks (examples include Denial of Service (DoS) and hacking into accounts). *Risky*—for activities that might result in criminal or civil charges depending on the methods used to create the product or service, such as blackhat SEO, spamming, selling hosting for illicit content, proxies (these might be compromised computers that have been rented as proxies). Finally, *benign* merchants that sell services and products that have little risk associated with them, such as creating content (videos, articles, images), exchanging one form of virtual currency for another (i.e. PayPal to BitCoin), whitehat SEO, or other miscellaneous services and products (examples include selling video game leveling services and used hardware).

B. Mapping Skype handles to IP addresses

LeBlond *et al.* published a technique to obtain the IP address of Skype users using only their handles without the users being aware of it [14]. We sketch the technique briefly here and refer the reader to the paper for further details.

The technique works by attempting to set up a voice call to the user from a Skype client on a machine that has a public

[1]We limit this manual analysis to only active merchants, since this is a time intensive process.

[2]Merchants have been charged with "creation and distribution of malicious code" in Russia [12] and "facilitate criminal activity" in the USA [13] in connection with selling malware.

Forum	Total handles	Active handles	Countries	ISPs*	IPs
AC	266	192	25	163	3428
BHW	205	156	41	126	6540
HF	271	130	36	209	1536
Total	742	478	62	450	11504

Table I: *Dataset by forum. We collect measurements from March 3, 2015 to May 3, 2015. We added merchant handles in batches during the measurement period: we probed every merchant at least 484 times (20 days); and on average 917 times (38 days). *We mapped IP addresses to ISPs using the MaxMind insights service [11].*

Category	Products and Services	Active handles
Illegal	Botnet, DDoS, Malware	62
Risky	Accounts, Hosting, Proxies, Spam	270
Benign	Content, Exchange, Other	146

Table II: *Breakdown of merchants by category.*

IP address. Typically, the Skype protocol exchanges both TCP and UDP traffic between the two hosts during a call setup. The technique blocks TCP traffic, which prevents the call set up completion; however, the UDP connection is still active, and this allows us to obtain the IP address of the user. The specific signatures we look for are: 1) a UDP packet of size between 60 and 70 bytes from the client, followed immediately by a response from our server of size 58 or 59 bytes, followed by a packet (need not be immediate) of size 28 from both sides, followed by a packet of size 3 bytes from server. This signature identifies a user logged in from a public IP address. 2) a UDP packet train starting with a size 28 packet from the server, followed immediately by a packet of size 28 from client, followed by a packet of size 3 bytes from client. This signature denotes that the user is behind a NAT.

In response to this technique, Skype initially added an option, "Allow direct connections to my contacts only", that blocked this method from obtaining the user's IP address. We cannot disambiguate inactive handles from ones that have enabled this option. However, this option was off by default since it increases the latency of call establishment. In addition, users have to enable this option for each device that they use to log in to Skype. It is important to note that merchants that enable this option are another source of potential bias in our measurements; however we also note that we obtain a reasonably high hit-rate from our measurement probes (478/742, or 64%), which suggests that the bias may not be too high. Finally, we note that in early 2016—after we ended our measurement collection effort—Skype updated their default protocol such that this method is no longer effective [15].

C. Data Collection and Sanitization

We set up 7 VM hosts with public IP addresses to track the 742 handles (106 per VM). We run our measurements for a period of 2 months from March 3 to May 3, 2015. In order to reduce call setup overhead, we only start up the Skype client once per iteration. The process takes about 30 seconds per handle, this allows us to iterate through the set approximately every hour. Since we reuse the same instance of Skype for multiple handles, we sometimes see traffic overlap between two calls. For example, the signature packets (that allow us to identify the user's IP address) of a previous call might arrive while the next call is in progress; this might cause us to falsely attribute IP addresses to users. We use multiple techniques to guard against this: first we randomize the order in which we iterate through the handle list. We group IP addresses according to their prefix (which we obtain through MaxMind), and discard a prefix if we see it less than five times for that handle. Finally, we recursively check to see if an IP address has been identified with the two previous handles in that iteration more often than for the current handle; if it has, then we discard that match for that particular handle. We find that these heuristics, though slightly arbitrary, help us reduce false positives; we discard 2.4 IP prefixes per user on average.

We also use available online skype resolver services to manually verify that our implementation works correctly, and also to double-check our results wherever we see a potentially "interesting" result (*e.g.*, a user registering from multiple countries, or multiple users registering from the same IP address). We do not use the online tools as our primary data collection source because they have poor availability and they use a different technique which does not allow us to disambiguate between an online user and the last-seen IP address of an offline user.

D. MaxMind

We use the MaxMind database [11] to map IP addresses to country, ISP, and ISP type. MaxMind classifies ISPs into types such as "residential", "cellular", "hosting", "business", "government", "library", and "college". While MaxMind is reliable for obtaining the country and ISP, the ISP type is not as highly reliable, especially outside the US. In particular, MaxMind does not identify hosting services correctly in Eastern Europe. We manually crosschecked (with whois records) cases that could have potentially been anomalous, and we found 35 ISPs that are identified as residential but are likely hosting services. In our subsequently analysis we assume that hosting and business (hereby denoted as "H+B") ISPs are primarily used as VPNs and do not directly leak the merchant's likely physical location. We group "residential",

"cellular", "government", "library", and "college" ISPs as residential (hereby denoted as "R+C"); this category directly leaks the merchant's likely geographical location.

E. Ethical Framework

We confirmed with our institution's IRB that this study does not constitute human subjects research, for multiple considerations. First, both the Skype handles and the associated IP addresses are publicly accessible, the latter due to the design of the Skype protocol. Second, given the nature of underground forums, the Skype handles are by design non-identifying. Third, our IRB takes the position that IP addresses are not treated as personally identifiable. In no cases did we attempt to tie our measurements to an actual identity. As part of our data retention protocol for this study, we deleted all copies of raw IP addresses collected during our measurements after the final version of the analysis was completed. We only retained the country level data and aggregate results of our analysis.

III. Merchant Profiling

We characterize how we construct a merchant's profile based on their IP addresses, geographic locations, the types of networks they use, and how they use them.

A. IP Geolocations

We first characterize the geographic locations of merchant's IP addresses at the country level. A single merchant can use IP addresses from multiple countries; this can happen due to merchants traveling, or using hosting facilities located in multiple countries. To provide a sense of the raw data, we first plot every country we see merchants associated with based on their observed set of IP addresses. Figure 1 plots the distribution of the top-10 countries per forum. We see that Russian-speaking countries dominate AC; this makes sense as AC is a Russian-language site. The US dominates HF, where Canada, and Great Britain, The Netherlands, and India also have a significant presence. India, Pakistan, and the US dominate BHW. We reiterate that this does not translate to users actually being resident in those countries. In later sections, we analyze the type of network associated with each IP address to identify merchants' probable location(s).

From our data, we saw that an overwhelming majority of merchants, 415(out of 478, or 87%) login from a single country (92% for AC, 89% for HF, and 78% for BHW). In the next section, we dig deeper into the type of networks that users in these three forums employ.

B. ISP characteristics

The type of networks merchants employ to communicate reveals the extent to which they care about availability and privacy; a merchant that solely uses VPNs likely cares about their privacy while a user that logs onto their mobile phone likely cares about being available. The merchants' choice of ISP have implications on their privacy. If a merchant is not careful enough to always log into Skype through a VPN, they will inadvertently reveal their residential or cellular ISP network which we can use to profile their likely physical location.

Table III breaks down the set of all ISPs that merchants use according to type. Note that merchants can use multiple ISPs, so the number of ISPs exceeds the number of merchants. We see that residential and cellular networks dominate in all three forums, but in differing levels. Between 16% and 22% of ISPs BHW and HF merchants used were hosting or business (H+B), larger than for AC (12%). Interestingly, H+B ISPs are most prominent amongst merchants offering Illegal services.

As we note in Section III-A, a merchant logging in from a certain country does not mean that the merchant resides or has traveled to that country. We break down the top-5 countries per forum based on the relative prevalence of the different kinds of access networks; R+C and H+B. Figure 2 shows that countries such as the US, Germany, UK, and The Netherlands have a significant fraction of hosting networks; merchants using these hosting services could be located anywhere in the world.

C. Inferring Merchant's Physical Location

We identify R+C networks as a leak, since they reveal the likely location of the merchant, at least at the country level, and sometimes even more precisely, such as the city or neighborhood. We focus on the merchants that use H+B, and whether they reveal their likely real location, knowingly or not.

In Table IV we break down the ISPs that each merchant uses according to type. The table shows the fraction of merchants that use only H+B or only R+C networks across the three forums, and the fraction of merchants that use H+B, and also R+C and potentially reveal their location. We saw that merchants do not typically use H+B networks; Table IV shows that between 14% and 22% use H+B (solely or in combination with R+B, which is captured by the "Mixed" category) services across all three forums.

We now consider those merchants that use H+B services (Table IV) to analyze leaks. Note that H+B merchants do not reveal their likely location, but Mixed do. AC and HF merchants seem to be the most careful; the percentage of AC and HF merchants that protect their actual IP address from leaking by exclusively using hosting services is around 6%, which is over four times greater then for BHW merchants. We validated this hypothesis by running the Fisher's exact test and confirmed that the difference between AC and BHW merchants, and HF and BHW merchants are statistically significant ($p < 0.05$).

This low percentage, about 1.3%, of merchants in BHW that protect their true IP address might be influenced by the

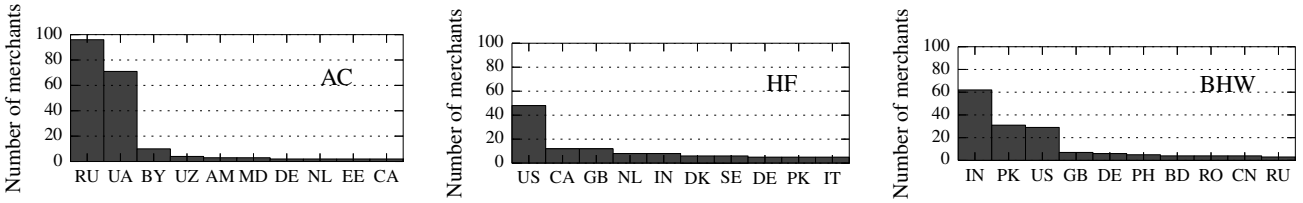

Figure 1: *Country distribution for the three forums we study. As expected, Russian-speaking countries dominate for Antichat, a Russian language forum. The US dominates for hackforums, while India dominates for Blackhatworld.*

ISP Category	AC	BH	HF	Illegal	Risky	Benign
Residential	65.9% (220)	57.4% (216)	67.7% (136)	50.4% (63)	61.5% (336)	72.1% (173)
Cellular	21.6% (72)	24.7% (93)	9.5% (19)	11.2% (14)	24.2% (132)	15.8% (38)
Hosting	8.4% (28)	14.4% (54)	19.4% (39)	34.4% (43)	11.0% (60)	7.5% (18)
Business	3.9% (13)	2.4% (9)	2.0% (4)	3.2% (4)	2.7% (15)	2.9% (7)
Other	0.3% (1)	1.1% (4)	1.5% (3)	0.8% (1)	0.5% (3)	1.7% (4)

Table III: *Type of access network used by merchant forum and type of merchant. The numbers in brackets denote the raw number of merchants.*

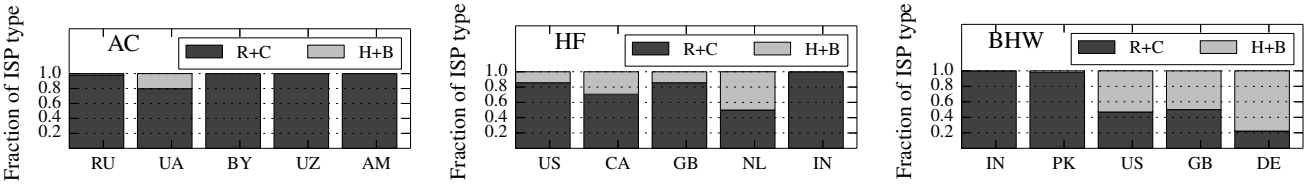

Figure 2: *A breakdown of the type of networks we observe in the most frequently seen countries per forum. Merchants prefer hosting services in North America and Europe; merchants registering from other parts of the world invariably do so from residential or cellular networks*

Forum	R+C only (%)	H+B only (%)	Mixed (%)
AC	159 (82.8%)	13 (6.8%)	20 (10.4%)
HF	111 (85.4%)	8 (6.2%)	11 (8.5%)
BHW	123 (78.8%)	2 (1.3%)	31 (19.9%)

Table IV: *Breakdown of merchants by forum and type of networks used. Mixed indicates that the merchant uses both R+C and H+B network. Only a minority of merchants use hosting services. A significant fraction of merchants that use hosting services or business networks also leak their likely real geographical location by also using residential or cellular networks.*

Category	R+C only (%)	H+B only (%)	Mixed (%)
Illegal	44 (71.0%)	8 (12.9%)	10 (16.1%)
Risky	222 (82.2%)	10 (3.7%)	38 (14.1%)
Benign	127 (87.0%)	5 (3.4%)	14 (9.6%)

Table V: *Breakdown of merchants by category and type of networks used. Mixed indicates that the merchant uses both R+C and H+B network. Merchants that offer illegal services are more likely to use hosting services. They are also less likely to reveal their likely real location by also using residential or cellular networks, though a significant fraction still do.*

fact that none of them offered goods or services that we categorized as illegal.

To validate this, we group merchants by their type and analyze the infrastructure they use. Table V again shows that most merchants predominantly use R+C networks. However, merchants that sell illegal goods or services were over three times more likely—12.9% vs. 3.7% and 3.4%—to exclusively use H+B. (We confirmed this hypothesis as well using the Fisher's exact test $p < 0.05$.) Since these merchants are probably the most concerned about their privacy, it makes sense that such merchants are more careful about revealing their true IP address and potentially their physical location.

D. IP Geolocations Revisited

We now re-visit the country distribution of merchants based on our detection of their likely real location. If a merchant uses an R+C ISP, then we classify the user as belonging to that country. Else we mark the country as "XX" — unknown. Figure 3 shows the distribution of countries we see per forum; we see that countries such as The Netherlands (NL) and Germany (DE) feature less prominently compared to their ranking in Figure 1. The US also sees fewer merchant counts, though its ranking in HF and BHW is unchanged. Interestingly, XX is in the top five for AC and HF, but not for BHW. The country distribution per merchant category was even more instructive: XX

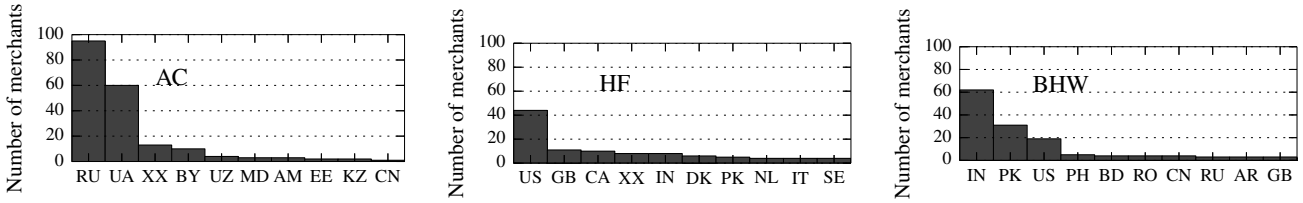

Figure 3: *Country distribution for the three forums based on our detection of the likely real location of users. Countries like The Netherlands (NL), Germany (DE) (and the US to an extent) feature less prominently than in Figure 1, as many ISPs from them are hosting services.*

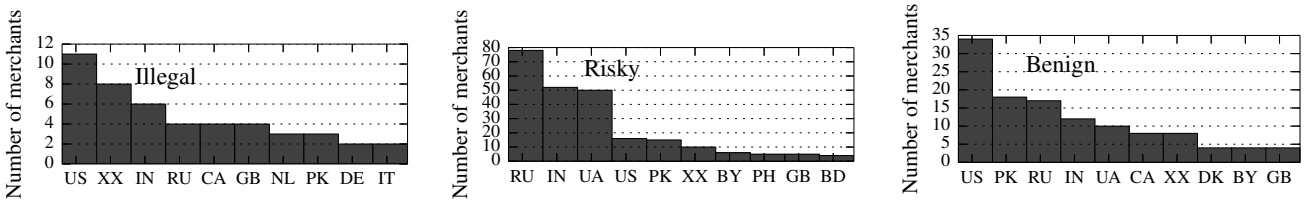

Figure 4: *Country distribution for the three classes of merchants based on our detection of the likely real location of users. Merchants who offer Illegal services are more careful about hiding their real location.*

was number two for Illegal activities. This confirms that merchants that undertake riskier activities are more careful about their privacy (Figure 4).

We note that for a small number of cases, we observed merchants using R+C networks from multiple countries; these are likely due to the merchant traveling, or multiple individuals in different countries using the same id, or due to a misclassification of the network type by MaxMind. We analyze some of these cases further in Section IV.

E. Diurnal Analysis

In this section we add depth to our profiling by inferring merchant behavior from their diurnal online/offline patterns. First we try to understand diurnal patterns at a high level: what are their work habits? In order to do so we devise a heuristic that approximates their diurnal patterns.

We only consider merchants that are active at least 24 times (recall that we probe approximately once an hour). We are left with 449 merchants, for whom we have, on average, 410 samples. We generate a time series for each merchant using the hour of day when they are active, and we normalize the number of hours we see a merchant online by the number of hours when our probing apparatus was active between the time we first saw that merchant and the time we last saw them. We thus obtain the average fraction of a day that a merchant is online, and we multiply it by 24 to get the average working hours per day.

We plot a CDF of the number of active hours per merchant across the three datasets in Figure 5. We see that between about 50–60% of merchants are active between 4–12 hours a day. This suggests that a majority of them are professional. About 20–25% of merchants register more than 16 hours a day, which suggests that they are continuously logged in from a mobile phone or a desktop.

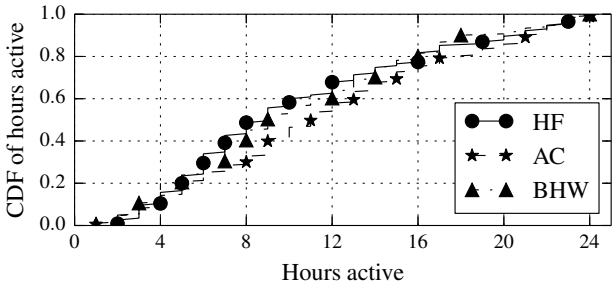

Figure 5: *CDF of hours active per merchant. We see that merchants are active for long hours; a majority are active more than 8 hours a day.*

Locating users based on diurnal pattern: A case study In general, it is difficult to use only diurnal patterns to locate a merchant. Merchants might have differing work-schedules, due to personal preferences, as well as due to business constraints. For example, a merchant in India that wants to interact with American customers probably would change their work patterns accordingly. Also, many small countries, particularly in Europe, share the same (or similar) time zones. We observed one AC merchant logging in from both Ukraine and Canada. Based on the location from where the user logs in most frequently, we would place the merchant in Canada. However, we also know by their logins from their residential network that the merchant is more likely to be in Ukraine. We studied their diurnal patterns in detail and saw that the user is active starting about 2AM Canadian Eastern time, which is roughly about 9AM Kiev time. This

validates our hunch (based on using residential IP addresses to locate the merchant) that the user is more likely to be in Ukraine. This case study confirms the promise, as well as the challenges of using diurnal patterns to geolocate users.

IV. Linking (or De-Linking) Merchants

In this section, we explore the use of IP-level and diurnal analysis to link seemingly independent merchants, track merchant travel, de-link merchants that appear to be superficially the same, or identify accounts that are operated by multiple people.

A. Linking independent accounts

Sometimes posts in the same forum having different handles might be operated by the same person. We attempt to link such handles using their IP addresses and activity signatures. We link merchants using a two-stage approach. We see whether two merchants consistently register the same IP address and we then match their activity using the timestamps in which they share the IP address. The second stage is important because of the high rate of churn that we notice in IP addresses in developing countries such as India and Pakistan (likely due to IPv4 shortages in those countries). We require that two merchants match an IP address within an hour of each other, at least five times (this helped us avoid false positives due to IP churn). We also do not consider IP addresses that are non-residential or non-cellular, because merchants could use the same hosting services, which might have the same external IP addresses.

Using this technique, we were able to match one pair of merchants in HF, and 7 in BHW. We also found a triplet of merchant IDs BHW that were linked. In many cases, these merchants have no outwardly connection to each other. For example, the HF pair consistently match on a Pakistan Telecommunication Company Limited (PTCL) IP address and also across time. They offer two different classes of services; merchant 1 offers booter DDoS services, while merchant 2 Linux DDoS botnet services. Interestingly, the first handle also consistently uses a hosting service alongside residential services to log in, while the second handle does not register through a hosting service at all.

B. Global accounts

We saw in Section III-A that a vast majority of merchants register from a single country. Only 13% of merchants register from multiple countries, and fewer still, 4%, register from residential or cellular ISPs from multiple countries. Such cases could mean one of many things: the merchant could be traveling, or could be using international roaming SIM cards, or the same account could be operated by multiple people in different countries. It is sometimes difficult, but not impossible to track travel, but it is not possible to differentiate between a single user using multiple SIM cards (and multiple devices) and multiple users (though we could make educated guesses based on diurnal patterns, but we do not deal with that in this paper).

We saw 17 cases that suggested that the user was traveling. In most cases, we found a clear signal that the user was traveling. For example, in one case, we saw a merchant with an IP address that pointed at a Mexican residential ISP, went offline for 23 hours, surfaced in Germany for 3 hours, and then went offline for 12 hours, and then resurfaced in Ukraine. The merchant has registered an Ukrainian IP address since then. In this case, we can surmise that the user traveled from Mexico to Ukraine with a stopover in Germany (a popular stopover point in Europe). There were other cases where teasing out travel patterns were more difficult. For example, one user was associated with multiple IP addresses in Romania. This user also logged in for 2 days from a residential IP address in the Czech Republic, alongside one Romanian residential IP address. On closer inspection, we found that the Romanian IP address was logged in for long periods, sometimes for 24 hours a day. This suggests that this device is a desktop, which is always turned on (even when the merchant is not using it). Filtering out that IP address gave us a clear, non-overlapping travel signature.

In another case, we observed that the same account registered two users in different countries, India and the UK, often at the same time. The distance between the two countries means that traveling between them in short timeframes is infeasible, so it can only mean that they are two different people, or the same person using multiple SIM cards, with at least one of the SIMs roaming internationally.

C. Profiling a Service with Multiple Handles

It is also interesting to see how many real people are behind a service. Many forum postings include multiple handles. In some cases all of the handles are operated by the same person. However, we see in some cases that they are not, and that there are multiple people behind a service. In effect, we do a similar analysis as for understanding global accounts, except that we know from the forum that two handles are linked to the same service. We studied one case which had two handles included in a HF posting advertising hosting for abusive tools. One account is for servers, while the other is for sales. We find that they are not only two people, but also in two countries. The person behind servers resides in Russia, while the person behind sales is in the UK, with evidence of travel to Russia. They also use the same hosting service in the Netherlands, which might be connected to their service.

V. Discussion and Future Work

We have presented our initial data-driven analysis of underground merchants based on profiling their IP addresses. This section will serve to provide a higher level view of our analysis to put it into context and discuss how our analysis might be used to better understand underground

merchants. In addition, we will describe future work that we are planning to undertake that will improve our methods and allow them to scale to larger analyses.

A. Operational Security

Based on our results from Table II we find that 87% (58/66) of the underground merchant engaged in illegal activities leak their likely real IP address and country of residence. We also find from Figure that 11 of the these merchants are located in the United States of America and 4 of them are located in Canada. Both of these countries have laws against computer crimes and strong enforcement of these laws. This indicates that our methods of profiling underground merchants might be useful to law enforcement in these countries.

B. Travel Detection

From our analysis it appears to be possible to detect when a merchant is traveling based on profiling their IP addresses. Using this method on a larger scale might enable us to measure common travel patterns of underground merchants and build up groups of merchants that travel to larger gatherings. This might also be useful for apprehending an underground merchant when they have traveled to a different jurisdiction.

C. VPN Identification

In our analysis, we detect VPNs based on the client IP address, using the Maxmind database. However, it is possible that if the database is incorrect, we might be incorrectly classifying VPN IPs as non-VPN. In order to verify the database, we attempted to compare application layer latency between the server and the client with the network latency. For every probe from Skype where we successfully obtained an IP address, we attempted to a) measure the application-layer latency between our server and the client which we attempted to obtain from the Skype control packets, and b) IP latency between the server and client IP addresses using ping. Our conjecture was that if the two latencies are similar then the client is not using a VPN, while if the application layer latency was significantly higher, then the client is using a VPN. Unfortunately, our samples for application-layer latency were too noisy for us to successfully validate and use this technique. Instead we randomly picked multiple IP addressed and validated them manually using the whois service. We plan to fine-tune this technique in order to strengthen our analysis.

D. Automated Methods for Linking (or De-Linking) Merchants

In Section IV, we provided case studies of single merchants using multiple accounts, merchants traveling, and similar accounts possibly manned by several individuals. In all of these cases, we use very clear signals. We find high correlation between IP addresses to link multiple accounts, or plausible travel signatures. Much of this effort was manual, and required knowledge about the network across different countries and providers. While a sharing a single IP within a small window of time might be highly suggestive in a US residential network, it may not mean much for a cellular network, or for a residential network in countries like India and Pakistan which have high churn. However, is possible to at least narrow down plausible candidates for such cases using our heuristics. For example, users registering from residential networks in multiple countries is so rare that that in itself becomes a signal for deeper analysis. These heuristics are a work in progress, and we plan to fine-tune them for automatic detection of such signatures.

E. Limitations

Our measurement study has some limitations due to the nature of both the subject matter and the technique we use. The set of merchants we use are essentially random; we pick them from popular online open forums without attempting to verify them. We are not aware of a better approach to choose merchants. The technique we use also limits the number of merchants we can study. It requires that we use servers with public IP addresses to run Skype. This restriction limits us to profiling 742 merchants. The technique also does not tell us whether a negative probe is due to the merchant being inactive or the merchant having turned on the feature that blocks this technique. We might therefore be undersampling the more security-aware merchants. Finally, accuracy of our analysis depends on the Maxmind database that we use to classify IP addresses. While Maxmind provides good mapping of IP addresses to countries and ISPs, we do not know how accurate is its classification of IP type (though we manually verify some of them using whois), in particular for addresses outside of the US.

We also note that our profiling methods can be deceived by an underground merchant that always utilizes a roaming SIM which uses IP addresses in another country. Our methods could also be deceived by using R+C proxies located in the same country. However, a merchant with this sophisticated of an operational security policy could also evade our profiling by disabling direct connections on all their Skype clients.

VI. Conclusion

We presented how the Skype handles of underground merchants can be used for profiling their network behavior and characteristics, tracking their diurnal and travel habits, and linking/de-linking different accounts. We tracked 742 Skype handles from three active underground forums. While the specific method we used has been turned off by default by Skype, our study shines a light on many operational aspects of underground merchants, such as their geolocation information, which in turn can be used to infer their likely

physical location and online behavior such as how long they stay online, their travel behavior, and whether or not they use privacy mechanisms, such as VPNs, to hide their locations. Our techniques to profile these merchants can be extended to any platform that reveals merchant IP addresses.

ACKNOWLEDGMENTS

Our thanks to Mark Allman and Louis Dekoven for valuable feedback on our study. This work was supported by the US National Science Foundation under grant CNS-1237265. Opinions and findings are those of the authors.

REFERENCES

[1] H. Fallmann, G. Wondracek, and C. Platzer, "Covertly Probing Underground Economy Marketplaces," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, ser. Lecture Notes in Computer Science, vol. 6201. Springer Berlin Heidelberg, 2010, pp. 101–110.

[2] S. Afroz, V. Garg, D. McCoy, and R. Greenstadt, "Honor Among Thieves: A Common's Analysis of Cybercrime Economies," in *Proceedings of the eCrime Researcher's Summit*. IEEE, 2013.

[3] N. Christin, "Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace," in *Proc. WWW*, 2013, pp. 213–224.

[4] "An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants," in *Proc. CCS*, 2007, pp. 375–388.

[5] V. Garg, S. Afroz, R. Overdorf, and R. Greenstadt, "Computer-Supported Cooperative Crime (Short Paper)," in *19th International Conference on Financial Cryptography and Data Security*. Springer-Verlag, 2015.

[6] T. J. Holt and E. Lampke, "Exploring stolen data markets online: products and market forces," *Criminal Justice Studies*, vol. 23, no. 1, pp. 33–50, 2010.

[7] R. Thomas and J. Martin, "The underground economy: priceless," *USENIX login;*, vol. 31, no. 6, Dec. 2006.

[8] M. Yip, N. Shadbolt, and C. Webber, "Why forums?: an empirical analysis into the facilitating factors of carding forums," in *Annual ACM Web Science Conference*. ACM, 2013, pp. 453–462.

[9] S. Afroz, A. Islam, A. Stolerman, R. Greenstadt, and D. Mccoy, "Doppelganger Finder: Taking Stylometry to the Underground," in *IEEE Symposium on Security and Privacy*, May 2014, pp. 212–226.

[10] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker, "An Analysis of Underground Forums," in *Proc. ACM Internet Measurement Conference*, 2011, pp. 71–80.

[11] "MaxMind - IP Geolocation and Online Fraud Prevention," https://www.maxmind.com/.

[12] L. Constantin, "Russia arrests creator of the devastating Blackhole exploit kit, 12 others," http://tinyurl.com/qbr7zwr, 2013.

[13] "Manhattan U.S. Attorney and FBI Assistant Director in Charge Announce 24 Arrests in Eight Countries as Part of International Cyber Crime Takedown," http://tinyurl.com/7sapezo, 2012.

[14] S. Le Blond, C. Zhang, A. Legout, K. Ross, and W. Dabbous, "I Know Where You Are and What You Are Sharing: Exploiting P2P Communications to Invade Users' Privacy," in *Proc. ACM Internet Measurement Conference*, 2011, pp. 45–60.

[15] "To our gamers: IP will now be hidden by default in latest update," http://blogs.skype.com/2016/01/21/to-our-gamers-ip-will-now-be-hidden-by-default-in-latest-update/, 2016.