Under the Shadow of Sunshine: Understanding and Detecting Bulletproof Hosting on Legitimate Service Provider Networks

Sumayah Alrwais[1,2], Xiaojing Liao[3], Xianghang Mi[1], Peng Wang[1], XiaoFeng Wang[1], Feng Qian[1],
Raheem Beyah[3] and Damon McCoy[4]

[1]Indiana University, Bloomington.{salrwais, xmi, pw7, xw7}@indiana.edu
[2]King Saud University, Riyadh, Saudi Arabia. salrwais@ksu.edu.sa
[3]Georgia Institute of Technology.{xliao, rbeyah}@gatech.edu
[4]New York University. mccoy@nyu.edu

***Abstract*—BulletProof Hosting (BPH) services provide criminal actors with technical infrastructure that is resilient to complaints of illicit activities, which serves as a basic building block for streamlining numerous types of attacks. Anecdotal reports have highlighted an emerging trend of these BPH services reselling infrastructure from lower end service providers (hosting ISPs, cloud hosting, and CDNs) instead of from monolithic BPH providers. This has rendered many of the prior methods of detecting BPH less effective, since instead of the infrastructure being highly concentrated within a few malicious Autonomous Systems (ASes) it is now agile and dispersed across a larger set of providers that have a mixture of benign and malicious clients.**

In this paper, we present the first systematic study on this new trend of BPH services. By collecting and analyzing a large amount of data (25 Whois snapshots of the entire IPv4 address space, 1.5 TB of passive DNS data, and longitudinal data from several blacklist feeds), we are able to identify a set of new features that uniquely characterizes BPH on sub-allocations and are costly to evade. Based upon these features, we train a classifier for detecting malicious sub-allocated network blocks, achieving a 98% recall and 1.5% false discovery rates according to our evaluation. Using a conservatively trained version of our classifier, we scan the whole IPv4 address space and detect 39K malicious network blocks. This allows us to perform a large-scale study of the BPH service ecosystem, which sheds light on this underground business strategy, including patterns of network blocks being recycled and malicious clients migrating to different network blocks, in an effort to evade IP address based blacklisting. Our study highlights the trend of agile BPH services and points to potential methods of detecting and mitigating this emerging threat.

I. Introduction

BulletProof Hosting (BPH) services rent out servers and networking infrastructure that will persist in the face of takedown attempts and complaints of illicit activities. This BPH infrastructure is a basic building block of the cyber-crime ecosystem. BPH is used by attackers as a stable base of operations from which to conduct their illicit operations that can run the whole gamut ranging from more risky activities, such as hosting botnet command and controls, launching DDoS attacks, and phishing pages to those less so, such as hosting pirated media. Originally, BPH infrastructure was solely provided by service providers[1] who catered to criminal clients and explicitly turned a blind eye to abuses emanating from their networks. However, the static nature, high concentrations of maliciousness, and reputation for not responding to abuse complaints often cause these BulletProof (BP) service providers' entire network allocations to become unilaterally blacklisted. In addition to blacklisting, these BP service providers often have difficulty finding peering points to provide stable network connectivity and in extreme cases they have been completely de-peered [1].

This increasing pressure on the monolithic BP service providers has driven many of them to transform the way they operate in order to evade these service provider (Autonomous System or AS) reputation based defenses, such as BGP Ranking [6] and ASwatch [7]. An anecdotally reported emerging trend is that the BPH services are now establishing reseller relationships with primarily lower-end hosting service providers [2]. These hosting service providers are often not complicit in supporting illicit activity, but rather either more lenient on illicit behavior or simply not investing much effort in proactively detecting or re-mediating malicious activities on their networks [2]. These lower-end service providers offer good cover for the BPH services, allowing them to leverage the better reputation of the parent providers, and as a result have a mix of both legitimate and BPH resellers, Figure 1 depicts this BPH ecosystem, which largely prevents unilateral actions against the whole service provider. This type of BPH infrastructure is not truly BP, since eventually the BPH service will likely have to move their clients to new IP addresses and network blocks. However, because the BPH service rents instead of owns the infrastructure, this strategy enables them to become more nimble and quickly move their clients when they are detected.

Quickly and accurately detecting these nimble BPH services that operate within lower-end service providers presents new technical challenges. Unilaterally blacklisting these lower-end service providers is not feasible in most cases due to

[1] We use the term "service provider" to refer to an entity that provides hosting services. Examples could be a cloud hosting provider, Content Delivery Network (CDN), Data Center (DC), Hosting Provider (HP) or an Internet Service Provider (ISP).

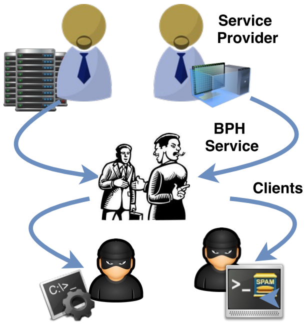


Fig. 1: BPH Ecosystem[2].

the amount of collateral damage that blocking them would cause. This forces Internet monitoring organizations to detect and point-wise blacklist IP addresses. Such protection, however, leads to a game of whack-a-mole, where blacklisted IP addresses are made ephemeral by the ability of the BPH services to move clients to new IP addresses when they are blacklisted. What is required is the ability to peer into a service provider and identify larger sets of IP addresses that have been allocated to a reseller and quickly determine the reputation of this address set. This will enable a middle ground between the overly coarse-grain AS-level blacklisting and fine-grain IP address reputation approaches, which will improve our ability to mitigate these emerging BPH service strategies.

Detecting malicious sub-allocations. To this end, we have created a set of methods to accurately detect malicious IP address network blocks. The first of these methods enabled us to identify sub-allocations of network blocks and their owners using IP Whois information provided by all five Regional Internet Registries (RIRs). Once we identified these sub-allocated network blocks, we then created a small set of labeled benign and malicious network blocks based on the manually complied blacklists of network blocks from Spamhaus [3] and lists of mostly benign network blocks based on a vetted subset of network blocks from Alexa [4] and top hosting provider lists [5]. Utilizing these labeled data-sets we were able to discover 14 key features that can assist in detecting malicious network blocks, based on three data sources: Whois, Passive DNS (PDNS), and AS reputation lists. Using these features we trained two classifiers, Support Vector Machines (SVM) and Random Forest (RF), achieving a 98% recall and 1.5% false discovery rates with a 5-fold cross-validation.

Findings. We ran a conservatively trained version of our classifier over all of the sub-allocations we found over a nine month period and detected 39K malicious network blocks. Based on these detected network blocks we were able to conduct a large-scale analysis of the BPH service ecosystem. From our analysis, we discovered previously-unknown patterns characterizing the ecosystem of this BPH infrastructure as follows:

- Parent Service Providers. Many legitimate service providers were not responsive to complaints of abuse emanating from their networks (e.g. "PEG TECH INC"). Others were conducting their business in a manner highly indicative of complicitness, such as aggressively recycling (rotating) network blocks and moving clients to new network blocks possibly in an effort to evade IP blacklisting. Additionally, we found service providers attempting to re-brand their businesses by creating many subsidiaries and ASes that delegate network blocks between them (e.g. "ColoCrossing").

- BPH Services. BPH services were registering as resellers with service providers and crossing Whois registries, countries and service providers by creating an abundance of Whois registrations (and objects). This enabled them to represent themselves as different entities with the service providers and registries. Additionally, we tracked the BPH services' movements from one service provider to another enabling the restoration of their services as a result of take-down efforts by parent service providers. Furthermore, we found these BPH services registering and dropping network blocks frequently allowing them to avoid IP blacklisting and delay take-down attempts by parent service providers.

- BPH Clients. We tracked the clients of the BPH services by analyzing the domains hosted on the network blocks and found them to host a number of malicious activities ranging from Command and Control systems to hosting pirated content. Additionally, we tracked the movement of the clients themselves through their domains and found many domains moving between at least two network blocks allowing some domains to survive for as long as 12 months before take-down.

Contributions. The contributions of the paper are organized as follows:

- New features for detecting malicious sub-allocations. We designed and implemented the first technique to detect sub-allocated malicious network blocks, achieving a 98% recall and 1.5% false discovery rate. Our approach is based upon a set of new features summarized over our analysis of a massive amount of data (e.g., Whois, Passive DNS, blacklists, etc.) as well as the ground truth collected through purchases. Of particular interest is the observation that some DNS features, such as churn, cannot be easily evaded by the BPH, due to the reliance of their services on DNS.

- New methods for validating our detection results. We performed a systematic validation of the detected sub-allocations, which is well-known to be difficult. Our validation included conventional cross-validation on the labeled set, utilization of multiple labeled sets with different qualities, identification of signals of malicious behaviors, and random sampling. This effort ensures that the findings made by our system are of the highest possible quality.

- The first large-scale study on the modern BPH ecosystem. Based upon our discoveries of malicious sub-allocations, we performed a measurement study on the BPH ecosystem, at an unprecedented scale. Our study reported 39K malicious sub-allocations, distributed across 3,200 ASes and new observations of how these services operate and evade detection, e.g., rotation of network blocks. These observations help us better understand this new BPH trend, which is critical for the ultimate elimination of this threat.

Roadmap. The rest of our paper is structured as follows. We start by presenting a background on modern BPH and how Whois IP address information is structured in Section II. Next, we describe our approach detecting malicious sub-allocated network blocks in Section III and our validation techniques

[2] The image of the two faced man is obtained from clipartfest.com

in Section IV. We then conduct a large-scale analysis of the BPH ecosystem based on the results of our classifier in Section V. In Section VI, we discuss the limitations of our detection method and potential future research directions. Finally, we present related prior research in Section VII and conclude in Section VIII.

II. Background

A. *BulletProof Hosting*

BulletProof Hosting services are a basic building block in the cyber-crime ecosystem that offer a safe haven for miscreants seeking to host all types of abusive content. Such services are resilient (aka, bulletproof) in the face of take-down efforts. This provides various protection mechanisms to their ill-willed clients, contingent upon the BPH services, money invested by their clients and the authoritative party involved. For example, abuse complaints generated by network administrators and ISPs can be simply ignored, while blacklisting by a highly influential party such as SpamHaus [3] could cause the BPH service to move the reported client(s) to a different network.

BPH services have many different structures. At one end of the spectrum are more traditional centralized BPH services with their hardware physically placed in bunkers and protected by armed guards, e.g. cyberbunker.com. These BPH services operate BulletProof Autonomous Systems (BP ASes) geared towards hosting malicious content and are highly stable. On the other end, are more fragile hosting services selling access to compromised machines for a small one-time charge. The duration of the access to these machines depends on a combination of factors, including how many complaints the illicit activity generates, when the service provider blocks access, or the maintainers of the machine detect and evict the malicious intruders. More recently, BPH services are evolving and moving away from dedicated BP ASes to legitimate ASes in an effort to blend-in and hide their clients' traffic among legitimate network traffic. This renders ASes blacklists, such as BGP Ranking [6] and ASwatch [7], impractical as blocking the whole AS would disrupt the operation of legitimate services. This type of BPH often takes advantage of the reseller programs offered by legitimate providers, as explained below.

Hosting options and price plans. BPH services offer a variety of hosting plans with stratified tiers of permitted abuse and computing resources. Computing resources can be (ordered by cost) dedicated servers, virtual private server (VPS) or shared hosting. The allowed abuse depends on the BPH service and the degree of risk they are willing to take. Generally, abuse falls into three tiers (high, medium and low). Child pornography, terrorism and anti-government content are the highest and rarely allowed by the BPH services, as observed in our study. DMCA (Digital Millennium Copyright Act) infringement and HYIPs (High Yield Investment Programs) are the lowest tier of abuse and are allowed by numerous BPH services mostly operating offshore in countries with lax laws regarding such content. Such services cost the renter an average of $10 a month for a typical VPS.

On the other hand, malware, phishing and botnet content are considered medium risk and are offered by a number of BPH services mostly found through advertising in underground forums. Their prices range from $30-$800, based upon the malicious content, computing resources and type of resilience a client needs. For example, during our research, one BPH service quoted us a dedicated server with SpamHaus protection at $650. However, VPSs for botnets would cost $250 a month on average. Figures 2 (& 14 in Appendix) show multiple hosting plans offered by the BPH services that we collected from underground forums. Some of these BPH services have operational online websites and others need to be contacted via instant messaging programs, such as ICQ. Additionally, some BPH services operate a multi-layer structure to better protect their resources, making use of proxies and fast fluxing (FF) to hide their back-end servers.

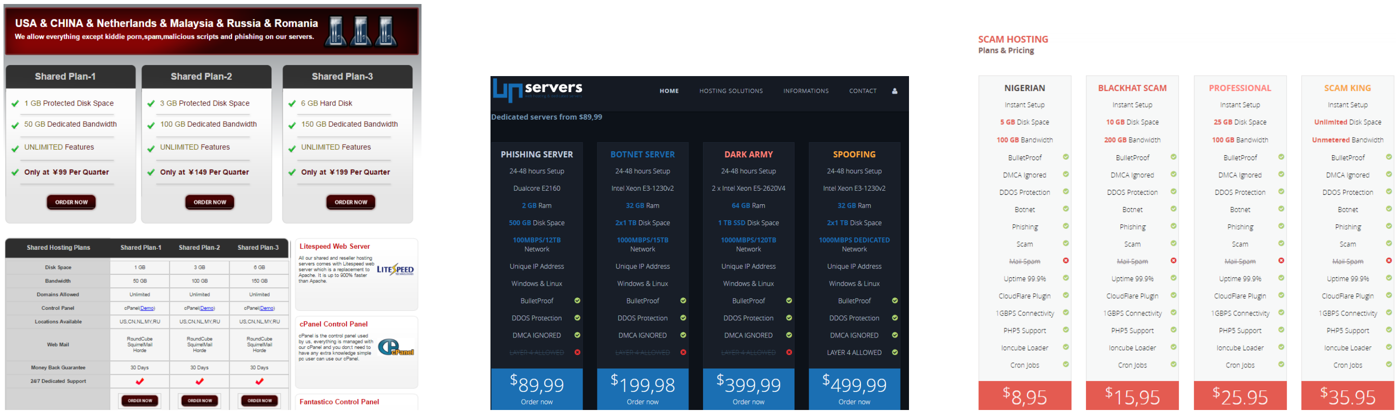
Reseller programs. Service providers such as ISPs, large hosting providers and cloud platforms offer a certain niche of clients the ability to join their *Reseller Programs*, which enables them to lease out computational resources in bulk at a discounted rate. Prominent examples of such providers include "ColoCrossing" [8], "LeaseWeb" [9], "OVH" [10], and "Voxility" [11]. These reseller programs give rise to a number of small-scale hosting providers (aka., virtual hosters) mostly run by one or a few people. Managing and operating these small hosting services is made even simpler by web hosting billing and automation platforms readily available at a small monthly charge. For instance, WHMCS [12], which costs $15/month, is a popular platform. According to their website they handle "everything from customer sign-up, to provisioning, management and support, WHMCS provides a single centralized platform for managing your web hosting business so that everyday tasks become quicker, easier, and more efficient" [12]. Figure 3 shows an example of a BPH service using WHMCS.

- Becoming a Reseller. To investigate the level of authentication and verification conducted by service providers at potential resellers, we registered for a reseller package with two service providers, "ColoCrossing" [8] and "Voxility" [11]. To this end, we purchased a domain, servicehosting.org (anonymously registered) and subscribed to WHMCS to run our hosting service. We set up one fake persona with a fake name, address and a working phone number (routed through Google). Armed with our setup, we contacted the reseller programs of "ColoCrossing" [8] and "Voxility" [11] and within days we obtained access to our leased servers. We found that becoming a reseller has been streamlined on lower-end service providers with minimal verification: in our case, one service provider authenticated the validity of the phone number via a text message while another required a phone call with a representative that was mostly a sales pitch rather than authentication. Ultimately, we paid $79/month for one server with a /29 network block in the US with "ColoCrossing" while "Voxility" charged us $215/month for one server with a /28 network block in Europe (EU).

BPH services have started to take advantage of these lax policies to create virtual BPH services spanning multiple service providers and countries. Additionally, many of these lower end providers have largely offloaded the handling of abuse complaints to their resellers.

B. *IP Address Allocation*

Whois. The Internet Assigned Numbers Authority (IANA) allocates IP addresses in large chunks to one of five Regional



USA & CHINA & Netherlands & Malaysia & Russia & Romania

We allow everything except kiddie porn,spam,malicious scripts and phishing on our servers.

Shared Plan-1	Shared Plan-2	Shared Plan-3
✓ 1 GB Protected Disk Space	✓ 3 GB Protected Disk Space	✓ 6 GB Hard Disk
✓ 50 GB Dedicated Bandwidth	✓ 100 GB Dedicated Bandwidth	✓ 150 GB Dedicated Bandwidth
✓ UNLIMITED Features	✓ UNLIMITED Features	✓ UNLIMITED Features
✓ Only at ￥99 Per Quarter	✓ Only at ￥149 Per Quarter	✓ Only at ￥199 Per Quarter
ORDER NOW	ORDER NOW	ORDER NOW

Shared Hosting Plans	Shared Plan-1	Shared Plan-2	Shared Plan-3
Disk Space	1 GB	3 GB	6 GB
Bandwidth	50 GB	100 GB	150 GB
Domains Allowed	Unlimited	Unlimited	Unlimited
Control Panel	cPanel(Demo)	cPanel(Demo)	cPanel(Demo)
Locations Available	US,CN,NL,MY,RU	US,CN,NL,MY,RU	US,CN,NL,MY,RU
Web Mail	RoundCube SquirrelMail Horde	RoundCube SquirrelMail Horde	RoundCube SquirrelMail Horde
Money Back Guarantee	30 Days	30 Days	30 Days
24/7 Dedicated Support	✓	✓	✓
	ORDER NOW	ORDER NOW	ORDER NOW

Litespeed Web Server

All our shared and reseller hosting servers comes with Litespeed web server which is a replacement to Apache. It is up to 900% faster than Apache.

cPanel Control Panel

cPanel is the control panel used by us, everything is managed with our cPanel and you don;t need to have any extra knowledge simple pc user can use our cPanel.

Fantastico Control Panel

(a) 66host

servers web hosting & dedicated servers

HOME HOSTING SOLUTIONS INFORMATIONS CONTACT

Dedicated servers from $89,99

PHISHING SERVER	BOTNET SERVER	DARK ARMY	SPOOFING
24-48 hours Setup	24-48 hours Setup	24-48 hours Setup	24-48 hours Setup
Dualcore E2160	Intel Xeon E3-1230v2	2 x Intel Xeon E5-2620V4	Intel Xeon E3-1230v2
2 GB Ram	32 GB Ram	64 GB Ram	32 GB Ram
500 GB Disk Space	2x1 TB Disk Space	1 TB SSD Disk Space	2x1 TB Disk Space
100MBPS/12TB Network	1000MBPS/15TB Network	1000MBPS/120TB Network	1000MBPS DEDICATED Network
Unique IP Address	Unique IP Address	Unique IP Address	Unique IP Address
Windows & Linux	Windows & Linux	Windows & Linux	Windows & Linux
BulletProof ✓	BulletProof ✓	BulletProof ✓	BulletProof ✓
DDOS Protection ✓	DDOS Protection ✓	DDOS Protection ✓	DDOS Protection ✓
DMCA IGNORED ✓	DMCA IGNORED ✓	DMCA IGNORED ✓	DMCA IGNORED ✓
~~LAYER 4 ALLOWED~~ ✗	~~LAYER 4 ALLOWED~~ ✗	~~LAYER 4 ALLOWED~~ ✗	LAYER 4 ALLOWED ✓
$89,99 Order now	$199,98 Order now	$399,99 Order now	$499,99 Order now

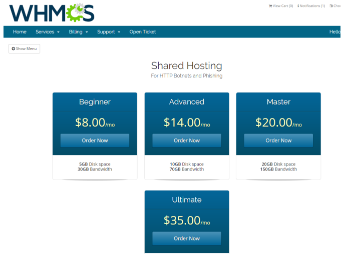
(b) bpservers

SCAM HOSTING
Plans & Pricing

NIGERIAN	BLACKHAT SCAM	PROFESSIONAL	SCAM KING
Instant Setup	Instant Setup	Instant Setup	Instant Setup
5 GB Disk Space	10 GB Disk Space	25 GB Disk Space	Unlimited Disk Space
100 GB Bandwidth	200 GB Bandwidth	100 GB Bandwidth	Unmetered Bandwidth
BulletProof ✓	BulletProof ✓	BulletProof ✓	BulletProof ✓
DMCA Ignored ✓	DMCA Ignored ✓	DMCA Ignored ✓	DMCA Ignored ✓
DDOS Protection ✓	DDOS Protection ✓	DDOS Protection ✓	DDOS Protection ✓
Botnet ✓	Botnet ✓	Botnet ✓	Botnet ✓
Phishing ✓	Phishing ✓	Phishing ✓	Phishing ✓
Scam ✓	Scam ✓	Scam ✓	Scam ✓
~~Mail Spam~~ ✗	~~Mail Spam~~ ✗	~~Mail Spam~~ ✗	~~Mail Spam~~ ✗
Uptime 99.9% ✓	Uptime 99.9% ✓	Uptime 99.9% ✓	Uptime 99.9% ✓
CloudFlare Plugin ✓	CloudFlare Plugin ✓	CloudFlare Plugin ✓	CloudFlare Plugin ✓
1GBPS Connectivity ✓	1GBPS Connectivity ✓	1GBPS Connectivity ✓	1GBPS Connectivity ✓
PHP5 Support ✓	PHP5 Support ✓	PHP5 Support ✓	PHP5 Support ✓
Ioncube Loader ✓	Ioncube Loader ✓	Ioncube Loader ✓	Ioncube Loader ✓
Cron Jobs ✓	Cron Jobs ✓	Cron Jobs ✓	Cron Jobs ✓
$8,95	$15,95	$25.95	$35.95

(c) outlawservers

Fig. 2: Examples of online BPH services with varying degrees of allowed abuse. a) Chinese BPH service offering shared hosing plans for anti DMCA and HYIP content. b) BPH service offering dedicated servers for phishing and botnet hosting. c) BPH service offering shared hosting for different types of scams. Incidentally, outlawservers and bpservers are owned and operated by the same party.



WHMCS

View Cart (0) Notifications (1) Cho

Home Services Billing Support Open Ticket Hello

Show Menu

Shared Hosting
For HTTP Botnets and Phishing

Beginner	Advanced	Master
$8.00/mo	$14.00/mo	$20.00/mo
Order Now	Order Now	Order Now
5GB Disk space 30GB Bandwidth	10GB Disk space 70GB Bandwidth	20GB Disk space 150GB Bandwidth

Ultimate
$35.00/mo
Order Now

Fig. 3: BPH service using WHMCS platform [12]

Internet Registries (RIRs). RIRs are nonprofit organizations that manage the registration of IP addresses in their regions by operating a directory service, *Whois*, to log and record all network block registrations. The Whois directory is operated in an object-oriented fashion. Figure 4 depicts the most relevant objects to a network block allocation. A network block is represented by an inetnum object which contains optional attributes identifying it, such as IP address range and/or CIDR, network name, description, allocation type (direct vs sub-allocation), organization, person(s), email and modification date. Organization and person objects also contain attributes that identify them and pointers to each other. Unfortunately, RIRs do not conform to a standard syntax, making it challenging to automatically extract certain information from them. For example, the APNIC registry does not have an organization object, while the LACNIC registry has an owner object slightly different from the organization object found in other RIRs. On the other hand, the ARIN registry has a customer object in addition to organization objects but it contains only name and address, missing additional crucial contact information. To collect information from Whois, we manually map the most relevant objects and attributes to our system to the objects in Figure 4.

Sub-allocations. RIRs allocate blocks of IP addresses to Local Internet Registries (LIR) within their region, e.g. ISPs. LIRs have the option to further split their allocated blocks and assign them to their customers. Figure 5 illustrates the cycle of a network block registration in which a direct allocation is made from a registry (RIR or NIR) to the LIR or ISP. The network block may go through further sub-allocations to customers. Also, a service provider's resellers are assigned sub-allocations and are recorded into Whois at the discretion of the service provider.

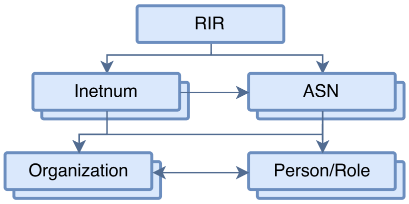


Fig. 4: Relevant Whois objects involved in a network block allocation.

Net Range:	198.46.154.88
CIDR:	198.46.154.88/29
Name:	CC-198-46-154-88-29
Net Type:	reallocated
Origin AS:	36352
Organization:	SERVI-139 [3]
Registration Date:	2016-05-02

TABLE I: Partial inetnum object in ARIN of our registered network block showing the primary key of our organization object.

- Sub-allocation example: Using our purchased reseller packages described earlier, we found that our network blocks were registered into Whois as sub-allocations. Table I illustrates a partial view of the inetnum object registered in ARIN Whois directory of our sub-allocation through "ColoCrossing". Incidentally, our other sub-allocation with "Voxility" did not use the same organization object but rather created a new one with the same attribute values.

Freshness of Whois records. Based on our observation, Whois records in all 5 RIRs are frequently updated. New objects of type inetnum, persons and organizations are created daily and even the objects currently stored in Whois are often updated. On a monthly basis, we witnessed an average of a 10% change rate of objects in each RIR. Table II illustrates RIR specific metrics evaluating the percentage of objects added and updated on a monthly basis.

[3] https://whois.arin.net/rest/org/SERVI-139/pft?s=SERVI-139

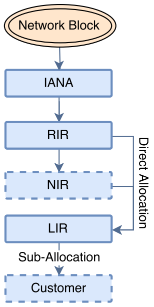


Fig. 5: Network block assignment cycle. Dashed boxes are optional. NIR is the National Internet Registry operating under a RIR but manages a smaller base of customers region, e.g. a country.

Region	RIR	Objects added	Objects updated
Europe	RIPE	8.14%	11.38%
North America	ARIN	6.58%	8.76%
Asia	APNIC	4.63%	6.71%
South America	LACNIC	10.07%	10.52%
Africa	AFRINIC	7.26%	32.21%

TABLE II: Monthly rates of Whois objects added and updated for each RIR, ordered by RIR size.

C. *Abuse of Sub-allocations*

Blacklisting. Current defenses against malicious network blocks fall into one of two categories: ASes reputation systems and IP address blacklists. Coarse-grained AS reputation systems, such as ASwatch [7], Fire[13] and BGP Ranking[6], focus on reputation metrics that can rank and detect dedicated malicious ASes (aka., BP). These AS reputation metrics are based on aggregated meta data for all IP addresses within the same AS. IP blacklists are more fine-grained, by targeting each IP address found to conduct malicious activities. As we will show below, sub-allocations in legitimate service providers are increasingly tied to malicious activities. Unfortunately, current defenses are inadequate in finding and blocking such malicious sub-allocations. To the best of our knowledge, SpamHaus's Edrop list [14] is the only blacklist that includes network blocks. However, based on our conversations with SpamHaus it is manually created and its coverage is quite limited: in the period of 9 months during our study, only 101 IP-prefixes were blacklisted.

Prevalence of sub-allocation abuse. We know based on anecdotal reports that sub-allocations are abused by BPH services in order to build a virtual hosting service running under the umbrella of legitimate service providers [2]. The scope and magnitude of this problem, however, have never been studied before. As a first step towards better understanding of this problem, we conducted an experiment with the goal of answering the following questions: are sub-allocations in legitimate service providers indeed abused, and if so, is such a threat pervasive?

Specifically, we examined the prevalence of sub-allocations in malicious activities by analyzing a 3-day snapshot (July 10-12, 2016) of 30K blacklisted IP addresses collected through our blacklist feed (BL-A, explained later in Section III-B). Looking at their corresponding network blocks reflected in Whois on July 12^{th}, 2016, we found that only 19.7% of the blacklisted IP addresses are directly allocated (i.e. managed by the service provider) while the remaining 80.3% are sub-allocations and 43% of these sub-allocations are owned and managed by a client of the legitimate service provider.

To investigate this abuse of legitimate services, we analyzed two data sets. Firstly, looking at the set of 30K blacklisted IP addresses (explained earlier) and using the AS reputation lists collected from the BGP ranking [6], we found that only 50 IP addresses (0.17%) belong to BP ASes while the rest were from legitimate service providers. Secondly, using a set of 95 blacklisted *IP-prefixes* collected through SpamHaus [14] and mapping them to their corresponding 164 Whois network blocks, we discovered that all 164 network blocks belong to legitimate service providers and 73.7% of them are owned and managed by clients of the service providers. This observation makes us believe that indeed the abuse of legitimate service providers, especially through sub-allocations managed by third parties (i.e. clients), is pervasive.

III. Finding BPH Sub-allocations

A. *Overview*

In this section, we elaborate on the design and implementation of a new technique for detecting malicious sub-allocations under legitimate provider networks.

In our research, we capitalize on the use of sub-allocations in Whois records by building a classifier tuned to finding malicious network blocks within larger blocks of legitimate provider networks and ultimately finding BPH resellers. This endeavor is by no means trivial: BP ASes were detected in the prior research by using the malicious activities observed throughout the entire AS, whereas a sub-allocation typically only has a small chunk of IP addresses and is unlikely to generate the same magnitude of bad signals for detecting a malicious sub-allocation. In our research, we leverage a unique observation that the BPH service, even on a legitimate provider network, needs to intensively utilize DNS to support its missions, which allows us to build our detection technique on top of a massive amount of DNS data that characterizes the BPH's activities. For this purpose, we collect numerous data feeds that illuminate different aspects of the sub-allocations in question. More specifically, we first run daily Whois scans of the IPv4 address space and then harvest the DNS records corresponding to the collected Whois sub-allocations from Passive DNS [15]. Additionally, we collect a variety of reputation based lists for ASes, IP addresses, domains and IP prefixes for feature extraction and to support the training of our detection model. This effort also requires a ground truth set of *legitimate* and malicious *network blocks*. This is hard to acquire, since few of these lists exist and those that do are often noisy or severely limited in coverage. In order to overcome this obstacle, we explore a few options for collecting a representative labeled set, one of which is directly purchasing hosting packages from BPH services.

Once data is collected, it goes through our processing pipeline which entails building network block hierarchies to find sub-allocations and their owners. After that, significant features are extracted to train a classifier. The trained model is then used to scan the larger set of sub-allocations and detect the ones exhibiting similar malicious signals. Finally, we perform an in-depth analysis of the BPH ecosystem. Figure 6 depicts our processing work-flow.

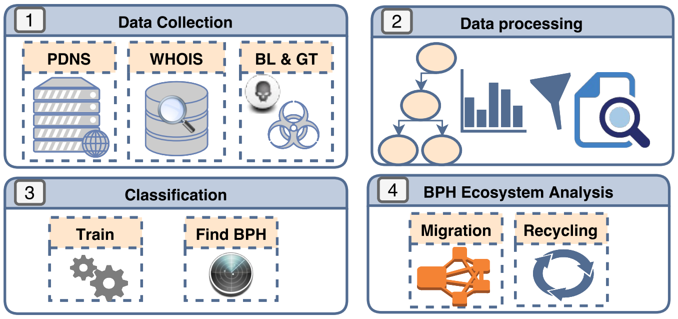


Fig. 6: Our processing work flow. BL and GT refer to blacklist and ground truth respectively.

Source	Duration(yyyyMMdd)	Size(∼)	Usage
Whois	20151225 - 20161109	9M inetnums/daily 3.4M organizations/daily 3.5M Persons/daily	Detection & Analysis
PDNS[15]	20150101 - 20160821	1.7TB	
AS Ranking[6]	20150101 - 20161007	42K ASes	
SpamHaus[14]	20160222 - 20161031	101 IP-Prefixes	Validation & Analysis
CleanMX[16]	20150801-20160714	1.5 TLD+3 700K IPs	
BL-A	20150825-20161011	1.5 TLD+3 4.4M IPs	

TABLE III: Data sets collected. BL-A is a commercial reputation based system. PDNS refers to Passive DNS.

Threat model. Given the perceived level of abuse, in this study we focus on building a classifier that can detect malicious sub-allocated network blocks. Our threat model assumes that the service providers are honest actors who will correctly update Whois records for network blocks that are delegated to resellers. If a service provider does not correctly update the Whois information we will not be able to detect BPH resellers operating within their networks. However, our initial exploration shows that many of the currently abused service providers are correctly updating sub-allocation Whois. If a service provider ceased correctly updating sub-allocation Whois, this could indicate a degree of complicity with the BPH services that might justify unilateral action against the service provider. In addition, we assume that other legitimate resellers will not enable BPH by renting them stable hosting unless, again, they are complicit with the BPH reseller. We also consider that the adversarial BPH services will attempt to evade our detection. Thus, we will identify features that likely increase their operational costs to evade detection.

B. Data Collection

As previously mentioned, our system relies on two key data-sets: Whois that is used to find sub-allocations, and Passive DNS (PDNS), which is used to extract signals indicative of malicious behavior. Additionally, complementary data sets are collected to help with validation and further analysis of the data. Table III lists the meta data of the collected data sets, a description of each data set follows.

Whois. Querying Whois over port 43 is not feasible due to the amount of traffic load it would generate when scanning the entire IP address space especially on a daily basis. Our solution is a two step process to collect daily Whois records of the full IPv4 address space. First, we download the full Whois database of each of the 5 RIRs [17], [18], [19], [20], [21]. This is accomplished by registering with each RIR and requesting bulk access to Whois. Once approved, access is granted to an `FTP` server to download the most recent copy of the Whois database. In order to collect historical records of the Whois database, we download a copy of the database once a day starting from late Dec 2015, and utilize 25 snapshots (10 days apart) in our study.

Unfortunately, depending on the RIR, the bulk snapshots of the Whois are sometimes anonymized by dropping contact information such as email and name. ARIN and APNIC are the only registries that provide the full Whois database in bulk. The other 3 RIRs omit some of the attributes and/or objects. For example, some RIRs do not include the primary keys of organizations and persons (in RIPE) and inetnum owners (in LACNIC). Additionally, AFRINIC does not include the user-names from contact emails. To supplement the missing information, we generate daily queries through an RDAP API supported by the RIR [22], [23], [24] for each missing or recently updated (according to the bulk database) object.

PDNS. To find domains hosted on sub-allocations and their corresponding IP addresses, we obtain access to a database of DNS look-ups collected by the Security Information Exchange (SIE) [25]. [4] The data set contains aggregated records of DNS look-ups over a two year period, and each record contains the number of look-ups and two time-stamps indicating the first and the last time the record has been observed to have the same value (i.e. the Rdata field in a DNS packet). We download the PDNS records that are in our scope of interest by submitting reverse queries on the sub-allocations IP prefixes through their API [15]. In total, we submitted 82K queries and collected 1.8 TB of PDNS records for our study.

AS Reputations. To compute the reputation of ASes, we leverage BGP Ranking [6], which is a public service operated by CIRCL. This service computes the reputation score of an AS based on data acquired from IP address based blacklists. We use BGP ranking for two purposes: determining ASes' reputations and bulletproof AS score. ASes' reputations are collected by downloading historical scores of each AS from 2015 to 2016 and computing the average to reflect the score of a given AS. Additionally, BGP ranking lists the top 100 ASes with the highest scores (indicating a poor reputation), which we also download on a daily basis from Jan - Oct 2016. We compute a BP AS score threshold in much the same way as ASwatch [7] by calculating the average score of the top 100 ASes for each day. We refer to this as the BP AS threshold where any AS with a score that is equal to or higher than the score is considered a BP AS.

Blacklists. We collect three types of blacklists: CleanMX, SpamHaus Edrop and BL-A. CleanMX virus watch [16] is maintained by the security community and contains historical malicious and suspicious URLs and IP addresses. We parse CleanMX lists and extract IP addresses and domains along with their listing time-stamp. The SpamHaus project [3] is a nonprofit organization that tracks cyber threats such as spam and malware. It maintains a variety of blacklists, some of which are available to the public. We use two of these lists:

[4] SIE collects data from a global sensor array that observes DNS cache miss messages and collects 200,000 observations per second and processes over 2 TB of data per day.

Edrop [14] and ROKSO[26]. The Edrop list is a manually maintained list of IP address prefixes controlled by cyber criminals which we download daily for nine months. We explain the ROKSO list below.

Additionally, we utilize a commercial reputation system, referred to throughout this paper as "BL-A", which provides real-time threat intelligence on IP addresses and domains involved in a variety of malicious behaviors such as malware, command and control, DoS attacks, botnets, exploits and vulnerabilities. BL-A also labels detected IP addresses and domains with one of 40 labels indicating the type of activities observed. Compared to CleanMX, this feed has a low false positive rate and reasonable coverage while CleanMX has a high rate of false positives, but improved domain coverage. We collected this real time feed for 14 months.

C. Ground Truth

Unlike domain name based systems where there are numerous ways to collect a labeled set, such as domain blacklists, commercial products, and active scanning of domains, finding labeled sets for network based detection systems is challenging. Previous research on BP ASes detection focuses on a handful of labeled BP ASes: e.g. 15 labeled BP ASes used by ASwatch [7]. Other AS reputation systems, such as FIRE [13], do not utilize a labeled set, but rather evaluate their results by comparing their top 10 ASes with other reputation based systems. In addition to the challenging problem of finding labeled malicious sub-allocations, finding labeled *clean* sub-allocations is also problematic due to the noise introduced by the temporary abuse of legitimate sub-allocations.

To address these challenges, we generate a labeled set with varying degrees of noise and experiment using different combinations in the training phase of the system, which is explained later in Section IV-A. Table IV provides an overview of the generated labeled sets.

Finding clean sub-allocations. Clean sub-allocations are collected from two sources: Alexa [4] and top hosting providers [5]. For the Alexa set, we collected the top 50K domains (according to Alexa [4]) that have been continuously present in the list from 2013 to 2015 (collected through the Archives [27]). For the selected list of domain names, we perform real-time DNS look-ups to obtain their hosting IP addresses and subsequently their corresponding sub-allocations. Unfortunately, this list contains many questionable sub-allocations, since it contains many sub-allocations hosting adult and copyright infringing websites. Next, we use a list of the top 500 hosting providers and search Whois to find their corresponding sub-allocations. We further split the list into the top 100 and top 500 hosting providers to reduce false positives. Additionally, we search the SpamHaus ISP [28] reputation data base for these top 500 hosting providers and if a hit is found, we label it as "Clean - Noisy".

Finding malicious sub-allocations. Next, we tackle the issue of finding malicious sub-allocations using a blacklist and through directly contacting and purchasing hosting packages from BPH services. We use two of SpamHaus's lists: Edrop [14], explained earlier, is manually maintained and thus its coverage is quite limited as it includes only 101 IP prefixes in a 9-month collection period. As such, we complement it with another slightly noisier list, ROKSO. SpamHaus maintains a Register Of Known Spam Operations (ROKSO) [26] and their actors by collecting an excessive amount of information identifying them, evidence such as websites of their operation and even some of their contact information, e.g. ICQ numbers. Incidentally, some of the actors listed are running BPH services. Within ROKSO, we find 10 BPH actors (out of 110) and are able to contact two of them. More specifically, we contacted the actor named "MailTrain" [5] through the ICQ number listed by ROKSO and purchased a server with 1 IP (to be used for botnets and spam) for $186. Additionally, ROKSO keeps track of the IP address prefixes used by the actors with two types of listings: current and archived. A current listing indicates that the IP address prefix has not been cleaned and is still used by the actor while an archived listing indicates that the IP address prefix is not used by the actor anymore. On July 6^{th} 2016, we scanned the ROKSO list and collected both current and archived listings for each actor. Using the current listings, on July 12^{th} 2016 we searched Whois for the sub-allocations matching all the listed IP address prefixes while partial matches were not considered. For the archived listings, we searched Whois but selected only the sub-allocations that were created or modified at most 10 days before the archiving date. These 10 days are used to account for the gap between ROKSO de-listing and a Whois record update. However, the resulting sub-allocations from archived listings are noisier than current listings, since it is dependent on an update of Whois records.

Purchasing from BPH services. Lastly, we scan underground forums for posts advertising BPH services and contact them for purchasing. Some of the BPH services have operational websites (although short-lived) to sell their services, as shown in Figures 2 (& 14 in Appendix). Other BPH services can only be contacted via ICQ [29] or Jabber [30] and most only accept digital currency, such as Bitcoin. Not all of our purchasing attempts were successful, some failed due to them suspecting that we were white hats because of language barriers. More specifically, in one of our purchase attempts with a Russian BPH service, which we conducted through ICQ, we were asked a few investigatory questions such as how we heard about them and why we cannot speak Russian. We proceeded as far as sending them a Bitcoin payment but within a few minutes, we received a message refusing to sell to us with a reimbursement of our payment because they believed we were white hats. Additionally, while attempting to purchase, we always asked for test IP addresses first. If they provided us with test IP addresses we did not proceed with the purchase. We bought a few of each of three hosting packages: shared hosting, virtual private servers (VPSs) and dedicated servers. Additionally, we purchased Fast Fluxing (FF) services from a few providers. The details of our purchases are described in Table V. Many of our collected IP addresses were linked to direct allocations and/or BP ASes. In total, we spent $1,155 and collected 37 IP addresses corresponding to 21 sub-allocations. Unfortunately, we cannot label many of the full sub-allocations found through the purchases as malicious since we only acquired one IP address from a BPH service. For example, one of our purchases resulted in a sub-allocation owned by Amazon which cannot be labeled as malicious based upon a single instance they gave

[5] https://www.spamhaus.org/rokso/spammer/spm806/mailtrain

Source	Label	Size	Set-A?	Set-B?
Alexa[4]	Clean - Noisy	2,238 sub-allocations	-	-
Top 100 Hosters[5]	Clean	67 sub-allocations	✓	✓
	Clean - Noisy	751 sub-allocations	-	✓
Top 500 Hosters[5]	Clean	112 sub-allocations	✓	-
	Clean - Noisy	43 sub-allocations	-	-
Edrop[14]	Malicious	15 sub-allocations	✓	✓
ROKSO[26]	Malicious - Current	876 sub-allocations	✓	✓
	Malicious - Archived	862 sub-allocations	-	-
Purchased	Malicious - Noisy	21 sub-allocations	-	-

TABLE IV: Collected ground truth data set. Sets A and B are the experimental sets we used to train the classifier. Note, the sizes of each set reflect the size after going through the filtering process explained later in Section III-D.

to us. As such, we did not use these resulting sub-allocations to train our classifier but rather for testing and validation.

D. Data Processing

In this subsection, we elaborate on how we process the Whois records to find sub-allocations, their owners and select targeted sub-allocations, and most importantly, select the right features to find the most relevant signals of malicious sub-allocations. These features are then used to train a classifier for detecting such BPHs (Section IV).

Finding sub-allocations. The downloaded Whois bulk databases of inetnum objects (i.e. network blocks) do not include the parent network blocks of an inetnum object, except for the ARIN registry. Also, the "status" attribute in the inetnum object should typically indicate the allocation type but setting the value of the attribute correctly depends solely on the service provider who registers sub-allocations with a registry. Furthermore, each registry has different values for the status, each indicating a certain type of allocation. As such, we determine sub-allocations by building a hierarchy network tree for each network block, which enables us to capture all the parent levels of a given network block (inetnum). Once a hierarchy tree is generated, a network block is considered to be a sub-allocation if it has a parent. Table VI lists the sub-allocations generated through this process.

Identifying sub-allocations owners. We consider a sub-allocation owner to be an owner object containing the contact information specific to the sub-allocation but not its parent. To generate an owner object for a given sub-allocation, we collect identity-specific attributes such as names, emails, the email's FQDNs, street addresses, phone numbers of the inetnum and its connected organizations and person(s). Then, we cross-match it with the owner objects of all of its parents and drop the matched values from the object. If the remaining attributes of the owner object contain at least an email or an organization's key, name or person, then the owner object is accepted. Otherwise the sub-allocation is considered not to have an owner (i.e. it is managed by its parent network) because its contact information cannot be found in Whois.

To capture the resellers switching or spanning across multiple service providers and registries by creating numerous person and organization objects, we further merge the created owner objects using strong attribute values that cannot be shared by 2 different entities such as emails (and their FQDNs) and organization names and keys. The resulting daily sub-allocations with/out owners are shown in Table VI.

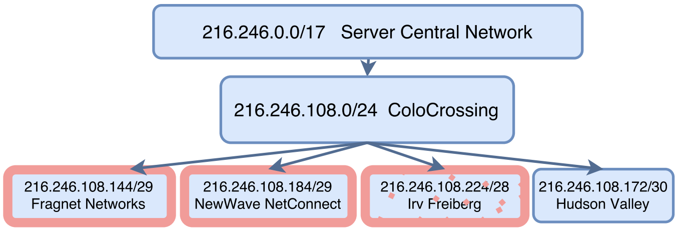


Fig. 7: Example of a partial hierarchy network tree showing 4 sub-allocations. Each node contains a network block prefix and an owner name. The dashed out sub-allocation indicates de-listing from the most recent snapshot of Whois. Sub-allocations with bold pink borders are the ones selected in the filtering process.

Further, emails and their FQDNs are treated carefully. FQDNs of popular portals such as GMAIL are not considered (since they are clearly too general), in which case the full email is used instead of its FQDN. Additionally, many service providers allocating sub-allocations would include an email (or FQDN) representing a network administrator or an account manager and as such cannot be considered related to owners. To avoid this, we drop the top 5% of all emails showing up in Whois objects. It is worth noting here that this owner generation process is used for analysis purposes later on in Section V and is not considered in the feature selection process of the classifier and as such has no impact on the correctness or accuracy of the classifier.

Filtering sub-allocations. Once sub-allocations are created, we select sub-allocations exhibiting hosting behaviors. More specifically, we select sub-allocations that are hosting more than 10 TLD+3[6] and are utilizing more than 25% of the network block according to data collected from PDNS. These two thresholds are set using the average values found in the labeled Top 500 hosting providers [5], as described earlier in Section III-B. Additionally, only leaf sub-allocations are selected because otherwise, the features of sub-allocations' children will incorrectly propagate up to them. For example, looking for the churn of TLD+3 will incorrectly consider TLD+3 hosted on its child sub-allocation.

Example. Figure 7 shows a partial view of the network block hierarchy of 4 sub-allocations owned by 4 different owners in "ColoCrossing", 3 of which are confirmed resellers. The sub-allocations of "Server Central Network" and "ColoCrossing" are filtered out because they have children and the sub-allocation of Hudson Valley is also dropped due to the lack of any corresponding DNS records. Incidentally, the sub-allocation of Irv Freiberg was detected by our system but had been de-listed from the current Whois records. Upon further examination of this network, we found it to be running a spam campaign by hosting 13 TLD+3, as shown in Table XIX in Appendix, for more than 19 months.

Feature selection and extraction. Unlike previous research on detecting BP ASes where the routing behavior of an AS or the volume of malicious behaviors observed on the *full* AS are strong indicators of a malicious (aka, BP) AS [13], [6], [7], finding indicators of malice at a finer granularity (that is, the network block) is more challenging. This is mainly due to the relatively small size of the network block and its age, e.g. sub-

[6] Throughout this paper, we use the term TLD+3 to represent the 3^{rd} Top Level Domain (TLD) of any host name while an FQDN is the 2^{nd} TLD.

BPH Service	Contact	Price	Date(s)	Acq Type	Content	Package	#IPs	RIR
66host	Website	-	20151203	Test	HYIP DMCA	-	3	RIPE
abusehosting	Website	$25	20151226	Both	HYIP DMCA	VPS	3	RIPE
bpw	Website & Online chat	$126	20160101	Purchase	HYIP DMCA Pharma	VPS	2	RIPE
bulletproof-web	Website & Online chat	$140	20151126	Purchase	Botnets Malware	VPS	4	RIPE
cyber-plane	Website	$55	20160105	Purchase	HYIP DMCA	Shared-Hosting & VPS	2	RIPE
grandhost	Email	-	20151201	Test	HYIP DMCA Port-Scan Gambling	-	1	RIPE
simplusx	Email	$150	20160105	Purchase	Botnet Spam	Shared-Hosting & VPS	2	ARIN
althost	ICQ	$350	20160115	Purchase	Botnet	VPS & FF	1	RIPE
protonhost	Website	$30	20160428	Purchase	Malware	Shared-Hosting & FF	4	RIPE
shadowhosting	Website	$25	20160428	Purchase	Malware	Shared-Hosting	2	RIPE
outlawservers	Website	$16	20160713	Purchase	Scam	Shared-Hosting	1	RIPE
mailtrain	ICQ	$186	20160711	Both	Botnet	VPS	9	APNIC & ARIN
bpservers	Website	$12	20160916	Purchase	DMCA	Shared-Hosting	1	RIPE
hostmy	Website	$40	20160918	Purchase	Botnet	VPS & FF	2	RIPE

TABLE V: BPH purchasing details. "Acq Type" refers to acquisition type which can be a test or a purchased IP address.

Region	RIR	Sub-allocations Per Day			#Owners
		#All	Owners?	#Selected	
Europe	RIPE	4M	2.2M	100K	1.3M
North America	ARIN	2.9M	2.8M	72K	2M
Asia	APNIC	928K	462K	34K	12K
South America	LACNIC	364K	357K	7K	167K
Africa	AFRINIC	86K	21K	1K	7K

TABLE VI: Daily processed sub-allocations in IPv4, ordered by RIR size. "All" represents the number of all sub-allocations while "Owners?" is the number of sub-allocations that have owners (i.e. managed by parties other than the parent service provider). "Owners" is the number of all merged owner objects we created for all sub-allocations over 25 Whois snapshots.

#	Type	Name	Normalized by
1	Whois	Sub-allocation size	-
2		Sub-allocation age	-
3	BL	AS Reputation	-
4	PDNS	Average Daily Traffic	
5		DNS Age*	Sub-allocation age
6		Average Daily TLD+3 churn*	
7		Average Daily TLD+3	
8		TLD+3 Age	DNS Age
9		Average Daily IP churn*	Sub-allocation size
10		Daily IPs	Sub-allocation size
11		IP Up-time*	Sub-allocation Age
12		IP Age	Sub-allocation Age
13-14		Net Utilization*	Sub-allocation size

TABLE VII: Selected system features. Starred Features are new PDNS features not used in previous research. Whois specific features have been used in previous research, but only for domain names.

allocations in RIPE have an average size of 130 IP addresses (∼/25), rendering domain name and IP address based blacklists largely ineffective (most blacklists limit their coverage to avoid false positives).

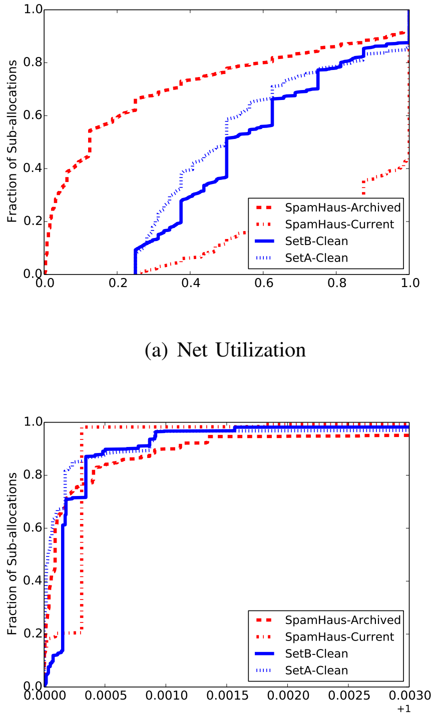
To address this challenge, we select features tailored to sub-allocations, particularly those considered to be robust, in a sense that evading these features would likely incur a significant cost (either monetary or increased blacklisting) to the malicious actors. To this end, we leverage three groups of features: Whois, PDNS and AS, totaling 14 features. Six of them have never been used in previous research. A description of each of these features is provided in Table VII.

• PDNS: BPH services tend to have a multi-layered infrastructure to better protect their back-end servers. This is usually deployed through the use of doorways (front-end websites), domain name and IP address fast fluxing, proxies and redirection servers which unavoidably will entail the use of DNS and thus presence in PDNS. We capture this behavior using 11 features, detailed in Table VII. Many previous systems detecting malicious domains have used features extracted from DNS records, e.g. Exposure [31], which we utilize and adapt to the context of sub-allocations, such as DNS look-ups (i.e. traffic), daily number and age of TLD+3 and IP addresses.

Additionally, we use 6 new features geared towards sub-allocations to find the signals of the front-end proxy layers employed by malicious actors. More specifically, we calculate the daily *churn* of TLD+3 and IP addresses. Intuitively, the clients hosted on BPH services are much less stable than those on legitimate services, who tend to come and go quickly. This observation is captured by the "Daily TLD+3 churn" feature. Furthermore, one expects to see rotation of a sub-allocation's IP addresses over time when used by BPH services to avoid blacklisting. We find this by calculating the length of a continuous duration of an IP address up-time in PDNS. BPH services would undoubtedly purchase resources, i.e. sub-allocations, as their demand increase, i.e. one expects them to monetize all resources paid for, which we measure by calculating the total usage of a sub-allocation's IP addresses. We find evidences for this intuition in the current listings by ROKSO [26], where 60% of BPH services utilize all of the sub-allocations' IP addresses as shown in Figure 8(a). We also measure the monetization of a sub-allocation by considering its DNS Age Vs Whois Age, i.e. the number of days any of the sub-allocation's IP addresses appearing in PDNS since the sub-allocation was created.

• Whois: We use Whois to extract two features: sub-allocation size and its age. Sub-allocations used by a BPH service or fully controlled by malicious actors tend to last for a few months before their owners move on to another sub-allocation. For example, in the labeled set A, described earlier in Table IV, malicious sub-allocations have an average age of ∼1K days compared to an average of ∼3k for clean sub-allocations.

• AS: Abuse of legitimate service providers will often show up in blacklists, although with a weaker signal than BP AS. We leverage this signal using the AS reputation of the sub-allocation's parent service provider (aka, AS) as shown in Figure 8(b).



(a) Net Utilization

(b) AS Reputation, a higher value indicating a worst reputation.

Fig. 8: CDF charts showing distribution of two selected features on the labeled sets.

IV. Evaluation

In this section, we present our approach for training the classifiers to detect malicious sub-allocations, and evaluation of the classifiers by testing it on two types of label sets: a highly conservative set and a noisier one. Then, we run the trained classifier on the larger unlabeled set of the filtered sub-allocations from all 5 RIRs for one Whois snapshot. We validate the results further by quantifying and showing indicators of badness, such as ties to malicious activities and de-listing from the most recent Whois records.

Training a classifier. Due to the challenges in finding representative sets of both clean and malicious sub-allocations (discussed earlier in Section III-C), we had to resort to labeled sets with different levels of noise, as presented in Table IV, to form two training sets, Set-A and Set-B, each of which contains both labeled malicious and clean samples. To select the labeled malicious sub-allocations, we used the purest lists for both Set-A and Set-B, namely Edrop [14] and current listings by ROKSO [26], providing a combined total of 891 sub-allocations. The relatively small size of the confirmed clean sub-allocations requires us to take a strategy that utilizes two sets, Set-A and Set-B, with differently labeled clean data (of different qualities and sizes) to compare the effectiveness of the models trained on them. Specifically, using the Top Hosters source [5], for Set-A, we pick only 179 sub-allocations that do not have a bad reputation according to the SpamHaus ISP [28] reputation database. As a result, Set-A is characterized by a small clean set and a larger malicious set and therefore biased towards malicious sub-allocations. For Set-B, we include all 818 sub-allocations, including the noisy ones, from the top 100 hosting providers to form its clean set, and as a result, it is balanced (between the malicious and clean sets) but has a lower quality (due to the noise). All remaining labeled sub-allocations are left out and considered for testing purposes only. Next, we experiment with two classifiers: Support Vector Machines (SVM) [32] and Random Forest (RF) [33] using 100 trees [7].

Metric	SVM		RF	
	Set-A	Set-B	Set-A	Set-B
Recall (TPR)	92.2%	89.8%	96%	98%
FDR	1.2%	3.1%	2.3%	1.5%
FPR	5.5%	3.1%	11.7%	1.6%
Accuracy	92.6%	93.2%	97.8%	97.1%
AUC	93.3%	93.3%	93.1%	97.2%

TABLE VIII: Results of a 5-Fold cross-validation on two classifiers, Support Vector Machines (SVM) and Random Forest (RF) using the labeled sets of A and B.

A. *Evaluation on Labeled Datasets*

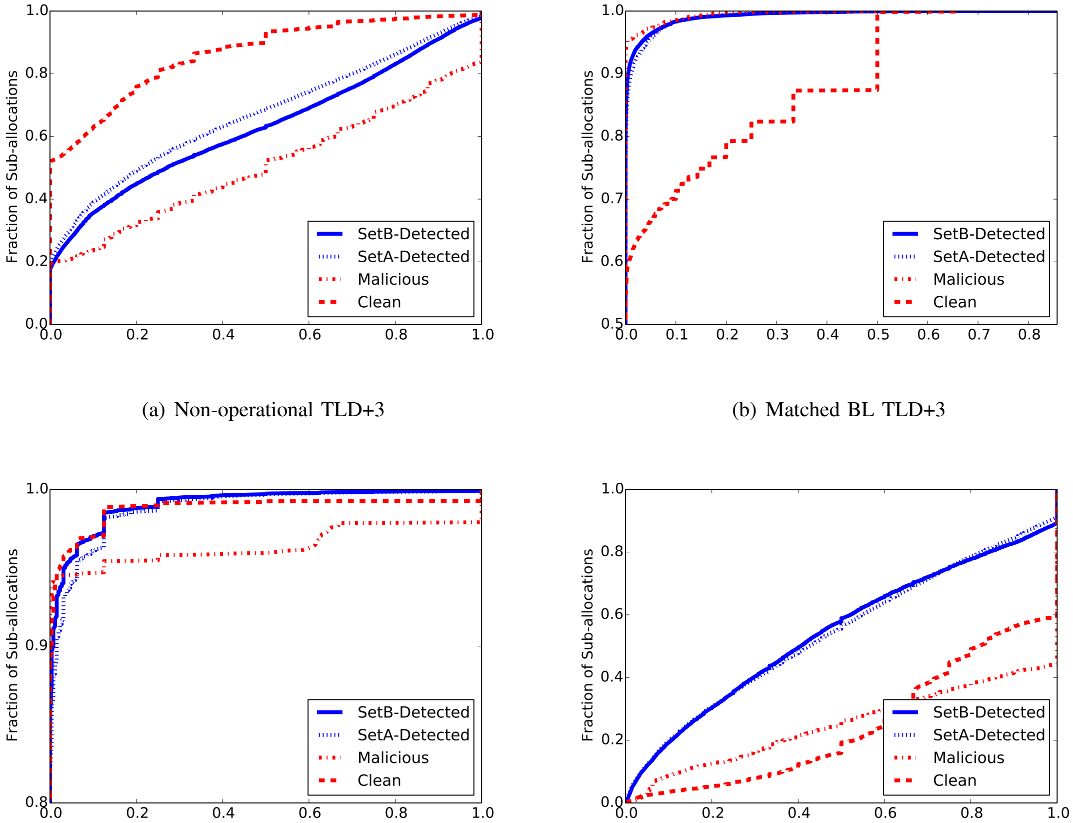
To evaluate the effectiveness of the classifiers and select the model that performs best, we employ two validation steps: 5-fold cross-validation and testing on the noisy labeled sets.

- 5-Fold Cross-Validation: We perform a 5-Fold cross-validation on each set for both classifiers, as shown in Table VIII. A description of our evaluation metrics is also provided by Table XVIII in Appendix. Clearly, RF outperforms the SVM when using the balanced Set-B but the SVM also handles the case of the unbalanced set, Set-A, much better than RF. In Set-B, the False Discovery Rate (FDR) and the False Positive Rate (FPR) are expected to be similar due to the balanced nature of the set. However, the FPR is expected to be quite high for Set-A because it represents the number of false positives out of all negatives (179) which is small compared to the number of all positives (891). We select RF as our classification algorithm for all future analysis, since it performs better on the sets, as seen by the cross-validation process.

- Using a test set: Next, we evaluate the RF models trained on both Set-A and Set-B by testing them on the much noisier labeled sets shown in Table IV, those not selected for training. For example, we test the Set-A trained model on the sub-allocations labeled "Clean-Noisy" from the Top 100 hosting providers source and the Set-B trained model on all sub-allocations from the Top 500 hosting providers source. Additionally, we test their performance on the sub-allocations containing our purchased IP addresses.

We use two evaluation metrics, True Positive rate (TPR) and True Negative rate (TNR), since their counterparts, the FNR and FPR, can be easily inferred, shown in Table XVIII in Appendix. The other metrics such as False Discovery Rate (FDR) do not apply here as each test happens on either a labeled clean set or a labeled malicious set, so the false positive rate cannot be meaningfully calculated based upon the set with a single label. The details of the test results are presented in Table IX. Overall, the model trained on Set-B works well but unfortunately is able to capture only 33% of the purchased set

[7] We limit the maximum number of trees to mitigate over-fitting on the training data.



(a) Non-operational TLD+3

(b) Matched BL TLD+3

(c) Matched BL IPs

(d) Distribution of FQDNs over TLD+3

Fig. 9: CDF charts showing 4 indicators of badness on two detected sets of sub-allocations with a frame of reference using the labeled sets. Set-A and Set-B are the set of sub-allocations detected by our classifier when trained with each of these sets.

compared to its counterpart trained on Set-A, which can detect twice as many of the purchases. The Set-B model also detects 43% of the labeled clean set (48 sub-allocations) as malicious. Taking a close look at these 48 cases, we find that they are all under the same parent service provider, LanLogic (a cloud provider), and have been registered since 1997. This indicates that they have served many clients on the same sub-allocations for a long time. Given the age of these sub-allocations, the likelihood of them at some point inadvertently hosting abusive content becomes unavoidable. Finally, we find that 4.5% of the false positives detected by our classifier run on the Set-B model are Alexa's sub-allocations that are not necessarily false positives. Looking at their indicators of badness, described later, we find that on average 1% and 3% of their hosted TLD+3 and IP addresses, respectively, are blacklisted, which justifies the noisy label assigned to it from the start (Table IV). Due to this noise, it is difficult to obtain a completely accurate evaluation of our classifier on these large testing sets and these results should be treated as estimates of performance.

Source	Label	Set-A		Set-B	
		TPR	TNR	TPR	TNR
Alexa[4]	Clean - Noisy	-	84%	-	95.5%
Top 100 Hosters[5]	Clean - Noisy	-	87.4%	-	-
Top 500 Hosters[5]	Clean	-	-	-	57.1%
	Clean - Noisy	-	76.1%	-	97.6%
ROKSO[26]	Malicious - Archived	53%	-	55%	-
Purchased	Malicious - Noisy	66.6%	-	33.3%	-

TABLE IX: Testing results of the Random Forest (RF) model trained with Set A & B on the noisy labeled sets.

B. *Evaluation on the Unlabeled Set*

Based on the evaluation of our classifier on the labeled set, we run the two trained models of Set-A and Set-B on the much larger unlabeled set of sub-allocations for one snapshot (July 12^{th} 2016) to gauge the scale and accuracy of our detector. As a result, we detected 40K (20%) and 20K(10%) sub-allocations using Set-A and Set-B respectively for training.

Indicators of badness. As previously stated, working on network blocks and more specifically sub-allocations brings in the challenge of result validation that is hampered by the difficulty in obtaining ground truth. Validation cannot be done by only correlating detected domains and IP addresses against blacklists or using malware detectors (e.g., Virus Total [34]), as most research studies have done. Validating a detected sub-allocation as a true positive entails finding either overwhelming evidence that the sub-allocation is or has hosted malicious/abusive content or has some compromised resources if not all. Such an in-depth investigation can only be performed on a case by case basis and is not quantifiable. In an effort to quantify the accuracy of our classifier, we compute a set of *indicators of badness* reflecting suspicious if not outright malicious sub-allocations. For the purposes of consistency and comparison, for each measured indicator on the detected sub-allocations, we apply the same indicator on the labeled clean and malicious sets. Following are those indicators:

• Non-operational TLD+3: All detected sub-allocations and those of the labeled sets were hosting in total over 570M TLD+3. We perform real time DNS lookup on all 570M of these domains to find the ones that have ceased to operate. More specifically, we measure the percentage of TLD+3 for which we received either an `NXDomain`, non-existent domain, or was parked with a domain parking service. This indicator proved to be quite powerful as one expects legitimate services to last a long time while malicious ones survive for a shorter duration. Figure 9(a) clearly shows that the sub-allocations detected by both trained models of Set-A, Set-B and the labeled malicious set all have a much higher rate of non-operational TLD+3 compared to the labeled clean set.

• Presence in blacklists: We show ties to malicious activities by cross-matching the TLD+3 and IP addresses of the sub-allocations against the 3 blacklists we collect, shown in Table III. Figures 9(b)&9(c) reveal a larger presence of detected sub-allocations in blacklists compared to the clean set, even though many do not have any footprints in blacklists making this indicator a slighter weaker one.

• Distribution of FQDNs over TLD+3: By manually sampling the detected sub-allocations, we noticed the behavior of them hosting many TLD+3 on one FQDN, which we quantify as `#FQDNs/#TLD+3`, as shown in Figure 9(d). Even though the labeled clean and malicious sub-allocations have fewer footprints of this pattern, we find through manual samples that the confirmed ones are clearly showing a larger rate of TLD+3 per FQDN. This is due to our selected features that implicitly capture such behavior.

• Future Whois de-listing: Lastly, we check the most recent snapshot of Whois to discover whether any of the detected sub-allocations' records were removed from Whois, essentially de-listing them. We find that 3.3% of the clean, 78% of the malicious, and 5.5% of the detected sets had been de-listed. Further analysis showed that one hosting provider (xlhost.com) in the clean set owned 77% of the de-listed sub-allocations with an average size of 8 IP addresses per sub-allocation. If we remove this one outlier from the clean set, the de-listing percentage for the clean set drops to 0.76%.

Manual sampling. Finally, we sample sub-allocations and manually investigate them case by case looking for evidence of FP. Sampling is performed in two ways: First, we randomly sample 20 sub-allocations; if we cannot tell one way or another whether a sub-allocation is a False Positive (FP), then we sample another. Out of 20 samples, 1 turned out to be a false positive. Second, we sample with biases towards false positives: specifically, we count sub-allocations that are owned by Akamai or educational institutions, old sub-allocations registered before the year 2000 and sub-allocations with one FQDN, which are 2,612 in total. If we consider all as false positives, even though some of them actually turned out to be compromised, the FDR would be 4%.

Duration	Dec 25^{th} 2015 - Aug 21^{st} 2016
Processed Whois	25 Snapshots
# Sub-allocations	Total: 39K Have Owners : 28K(71.5%)
#ASes	3,200
#Owners	19K

TABLE X: Breakdown of the detected set of sub-allocations and their meta data. "Have owners" refers to the number of sub-allocations who have an owner object.

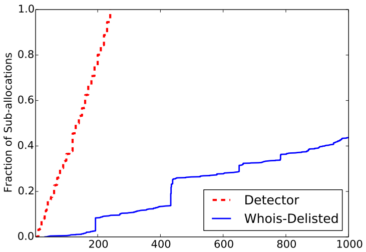


Fig. 10: CDF Comparing the performance of our classifier to Whois de-listing. We cut off the X-Axis at 1K due to the large maximum Whois de-listing age of 17K.

At this point, we performed an extensive evaluation of both trained classifiers but a model trained with Set-B has less false positives and as such we use it for the remainder of the paper.

V. Exploring the BPH Ecosystem

Landscape. As described in Table VI, the average number of selected sub-allocations for all 5 RIR is 192K per day, which we use to scan for malicious sub-allocations for each of the 25 processed Whois snapshots from December 2015 to August 2016, one snapshot processed every 10 days. In order to perform an in-depth study of the ecosystem of BPH, we choose to forgo coverage in an effort to reduce false positives. As such, we ran the detector trained on Set-B on the full set and detect 39K sub-allocations in total, averaging 20K (10%) per processed snapshot as detailed in Table X. Out of them, 17K (44%) sub-allocations were detected on the first day, out of which 738 were de-listed in Whois with an average de-listing time of 1,500 days. The remaining sub-allocations detected after the first snapshot were detected with a delay of 130 days on average. Figure 10 compares the performance of our detector to Whois de-listing. This high percentage of sub-allocations that we detected as malicious sheds light on the potential magnitude of BPH services currently operating and using this method to evade detection.

In the proceeding section, we perform a large-scale analysis of the BPH ecosystem through the lens of the 39K detected sub-allocations in an effort to better understand the extent of it[8]. Firstly, we look at the role service providers play in the ecosystem and their degree of potential complicitness by measuring a "Recycling" rate gauging the magnitude of network block turnover in an effort to clean up their IP address

[8] We point out that this analysis is likely incomplete and possibly biased, due to the inherent limitation of our detection method. Even though our starting point might not be fully representative of all BPH services, we argue that our study still captures a large number of malicious sub-allocations and amount of illicit activity.

space and avoid AS based detection methods. Next, we look into the owners of the detected sub-allocations and trace their hosting movements over multiple networks and highlight the methods they employ to evade detection by spanning their infrastructure across registries and ASes. Lastly, we explore what illicit activities the clientele of these BPH services are engaged in by analyzing the domains hosted on the detected sub-allocations. We find hosting of a variety of malicious activities and the long time they survive due to network and/or BPH service movements.

A. Service Providers

Overview. In total, we detected sub-allocations hosted on 3,200 ASes which we use to determine the actual service provider who delegated the sub-allocations to clients or resellers. Looking at the service providers, we find the majority of them fall into one of three categories; ISP, Cloud platforms and large hosting providers with an overall majority of cloud hosting services. Overall, we find that 50 service providers (ASes) account for more than half of the detected sub-allocations. The AS with the most detected sub-allocations was “PEG TECH INC” hosting 7% of the detected sub-allocations. Incidentally, we detected 50% of all of this AS’s sub-allocations and found many online reports indicating a pattern of ignoring abuse complaints and hosting malicious clients, specifically spammers. Another highly ranked AS in our detected set is “VPSQuan LLC” which we detect 37% of its sub-allocations hosting many abusive short-lived TLD+3 running brute force tools, vulnerability scanning, scams and a long list of other abusive activities, a sample of them is shown in Table XIX in Appendix. Interestingly, we detected 717 sub-allocations delegated by “Psychz Networks” accounting for 20% of all their owners and sub-allocations and again found excessive amounts of ignored abuse complaints. Additionally, we detect sub-allocations from ASes reported in prior work from OpenDNS [35] such as “King Servers”. A detailed list of the top 15 ASes in our detected set is provided in Table XI.

Recycling. By manually investigating a few service providers we found a pattern of some likely more complicit service providers actively rotating sub-allocated network blocks. This was possibly done to avoid blacklisting and to clean up the IP addresses’ reputation by delegating the network blocks to legitimate services. In an effort to quantify this behavior, we define a “Recycling” rate to capture the frequency of network block delegation while accounting for the cases of legitimate business expansion. To this end, we track and count the number of network blocks added and dropped for each processed snapshot starting from December 2015 to August 2016 and compute a recycling rate per snapshot for each AS as follows,

$$\frac{\#\,of\ Network\ Blocks\ Added + \#\,of\ Network\ Blocks\ Dropped}{Total\ Number\ of\ Blocks}$$

An overall average (for all 25 snapshots) is then computed per AS. Table XII details ASes with the highest recycling rate in our data. Additionally, we compute an average life of dropped networks blocks per AS and find that over all ASes, dropped network blocks survived for an average of 120 days. Interestingly, the ASes described above such as “VPSQuan LLC” and “Psychz Networks” had a recycling rate of Zero where they did not make any efforts in hiding or cleaning up the IP addresses reputation explaining the large number of abuse complaints. On the other hand, ASes with a high recycling rate, mostly data centers (cloud and hosting providers), were actively moving around network block delegations between a few owners. Furthermore, sub-allocations delegated and recycled by data centers had a much shorter life than those of ISPs, 121 days versus 630 days respectively.Unfortunately, this type of behavior can reduce our ability to detect malicious sub-allocations, since on average it takes our classifier 130 days before there is enough activity to reach the threshold for detection.

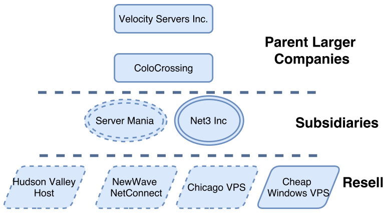


Fig. 11: Partial view of our collected organizational chart of “ColoCrossing”. Dashed boxed represent confirmed “ColoCrossing” subsidiaries.

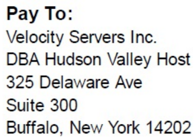

Pay To:
Velocity Servers Inc.
DBA Hudson Valley Host
325 Delaware Ave
Suite 300
Buffalo, New York 14202

Fig. 12: Bitcoin Invoice for a purchase through “Hudson Valley Host” showing vendor information.

Re-branding. Taking it a step further, we manually investigate one AS, “ColoCrossing”, that had a reputation of chronic abuse [9]. “ColoCrossing” ranked number 15 in our list of ASes with the highest recycling rates. On average, “ColoCrossing” has a 3.5% recycling rate and a network block lifespan of 200 days. From our manual investigation and active interactions that include purchases from “ColoCrossing”, we found a network of hosting providers and resellers that are owned by the same parent company. In addition to recycling network block delegations, we uncovered a pattern of re-branding and creation of many subsidiaries with different ASes. This organization would then move around network blocks between ASes that were originally assigned by IANA[36] to “ColoCrossing” and its parent company “Velocity Servers”. More specifically, Figure 11 shows “ColoCrossing” and some of the organizations with network block delegations. We found network blocks often delegated to subsidiaries and resellers (either directly or through subsidiaries). Some of these companies turned out to be subsidiaries and resellers owned and operated by “ColoCrossing” which we confirmed either through a payment chain or contact information. More specifically, we contacted “Hudson Valley Host” to purchase a reseller package and chose to pay by Bitcoin. The invoice generated through the CoinBase processing center showed the vendor to be “Velocity Servers” with the address of “ColoCrossing” as shown in Figure 12. Additionally, “ServerMania” and “Chicago VPS” had corresponding Whois records with shared contact information, an email that was the same as that used by “Hudson Valley Host”.

[9] https://twitter.com/spamhaus/status/480312720697606144

#	ASN	AS Rank	AS Name	Type	Coverage	Recycling Rate	#Detected Sub-allocations(%)	#Detected Owners	RIR	Website
1	54600	1.0001986502155999	Peg Tech Inc.	DC	7.69%	0.00%	2835(51.46%)	908	ARIN	petaexpress.com
2	11282	1.00006433491671	YUNM	DC	4.81%	0.00%	1771(49.80%)	1757	ARIN	serveryou.com
3	15003	1.0003085292176701	Ubiquity	HP	2.62%	0.38%	967(31.66%))	807	RIPE	ubiquityhosting.com
4	18779	1.0000193998423199	Energy Group	HP	2.42%	0.03%	892(8.40%)	325	ARIN	egihosting.com
5	7018	1.0000025631946801	AT&T	ISP	2.13%	0.24%	785(0.06%)	774	ARIN	att.com
6	40676	1.00020731537693	Psychz	HP	1.95%	0.36%	717(21.69%)	693	ARIN	psychz.net
7	9737	1.0000083869796701	TOT	ISP	1.86%	2.57%	687(71.04%)	2	APNIC	tot.co.th
8	6830	1.00000183689402	Liberty Global	ISP	1.58%	0.11%	581(1.03%)	103	RIPE	libertyglobal.co
9	38197	1.00009391599797	Sun Network HK	ISP	1.54%	0.20%	569(32.74%)	348	APNIC	sun.net.hk
10	5089	1.00000341462089	Virgin Media	ISP	1.42%	2.36%	525(1.06%)	7	RIPE	virginmedia.co.uk
11	62468	1.00056983197815	VPS Quan	HP	1.12%	0.00%	413(37.75%)	382	ARIN	vpsquan.com
12	53755	1.00008892325846	IOFlood	HP	1.03%	0.14%	380(26.63%)	64	ARIN	ioflood.com
13	16637	1.00000709622196	MNT Network	ISP	0.98%	0.01%	361(3.68%)	1	AFRINIC	mtn.com
14	33387	1.00132969982845	NOCIX	HP	0.85%	0.02%	313(7.24%)	139	ARIN	datashack.net
15	16276	1.0008670717274699	OVH	DC	0.78%	1.83%	286(0.42%)	157	RIPE	ovh.net

TABLE XI: Top 15 ASes covering 33.6% of all detected sub-allocations. The AS Rank represents the AS reputation collected through BGP Ranking [6], the higher the number the more malicious activities are observed. The “coverage” column represents the ratio of detected sub-allocations found in the AS out of all detected sub-allocations while the ratio in parenthesis indicates the ratio out of the ASes’s overall size. “DC” & “HP” stand for Data Center and Hosting Provider respectively.

ASN	AS Name	Type	Recycling Rate	Average block life	#Detected Sub-allocations(%)	#Detected Owners(%)	RIR	Website
61272	IST-AS	DC	19.7%	45 days	10(12%)	5(25%)	RIPE	bacloud.com
32421	Black Lotus	ISP	10%	500 days	93(94%)	59(4%)	ARIN, RIPE	blacklotus.net
62240	Clouvider Limited	DC	8%	111 days	6 (3%)	2(5%)	RIPE	clouvider.co.uk
60404	Liteserver	DC	7.4%	0	1(4%)	1 (50%)	RIPE	liteserver.nl
46475	Limestone Networks	DC	6.8%	161 days	4(0.1%)	3 (0.1%)	ARIN	limestonenetworks.com

TABLE XII: Top ASes ranked by their Recycling rate. Ratios in parentheses are computed out of the ASes’s overall size and not our detected set. “DC” stands for Data Center.

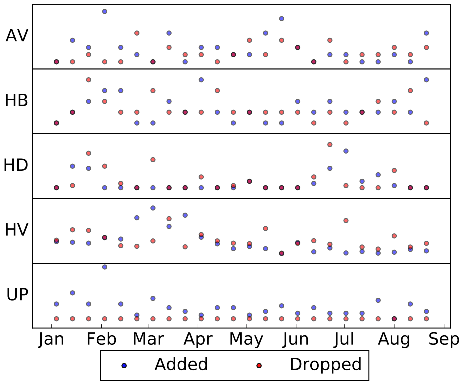


Fig. 13: Recycling behavior of the top 5 Owners showing the daily number of network blocks registered and dropped. An acronym represents a name given to the owner object, as in Table XIV. For example,“HV” stands for Hudson Valley Host.

This behavior causes “ColoCrossing” and its parent company “Velocity Servers” to not be detected by our classifier and other AS based reputation metrics [10]. We found similar re-branding and AS creation patterns by “Ecatel” and “Quasi Networks” but did not specifically interact with them to gain a deeper understanding of their structure.

B. Sub-allocation Owners

Overview. In order to capture the malicious actors and resellers managing the detected sub-allocations, we analyze the owner objects generated by the process described in Section III-D. Overall, 12K sub-allocations did not have corresponding owner objects indicating they are managed by the same party managing the parent network block. In other words, these sub-allocations were not managed by a 3^{rd} party. 28K of the detected sub-allocations had 19K corresponding owner objects. Table XIII lists the top 15 owners objects managing the detected sub-allocations. Not all owner objects are resellers, some are businesses and private customers. Almost 50% (10K) of our collected owner objects are private customers without any contact information other than a customer name. Automatically labeling resellers is difficult as they exhibit the same features as other customers and business. For the purposes of the data provided in Table XIII, we manually visited the websites of the owners to determine if they are likely to be resellers.

Reincarnation & recycling. The process of owner object generation enables us to capture owners reincarnations where many resellers span multiple service providers and even registries. Furthermore, once certain owners are blocked by a service provider, they simply register with another using similar contact information enabling us to link them. As shown in Table XIII, we found a number of owners spanning ASes (i.e. service providers) and even registries. We also find an overwhelming number of Whois objects (e.g. organizations and persons) per owner, with an average of 8 objects, making it more difficult to blacklist their Whois objects.

Additionally, we find owners repeatedly registering and dropping sub-allocations in much the same way as the service providers explained in Section V-A. Calculating the recycling rate per owner object, we find a different distribution of objects with quite a high rate of recycling, some are as high as 7%, shown in Table XIV. Figure 13 illustrates the daily churn

[10] It is unclear if this activity is intentionally evading reputation metrics or if it is benign.

#	Web Site	ASN	#ASes	#Detected Sub-allocations(%)	Object Size	Created Date (yyyyMMdd)	RIR	IsReseller?
1	mtnbusiness.co.za	16637	1	361(0.9%)	4	-	AFRINIC	N
2	NA	33387	2	143(0.36%)	1881	20100908	ARIN	-
3	yhsrv.com	54600	5	130(0.32%)	101	20120423	ARIN	-
4	NA	54600	1	110(0.27%)	225	20120501	ARIN	-
5	NA	54600	2	102(0.25%)	250	20120501	ARIN	-
6	xhostserver.com	53755	2	78(0.19%)	258	20110728	ARIN	Y
7	real.kamchatka.ru	34974	1	78(0.19%)	6	-	RIPE	N
8	NA	54600	1	76(0.19%)	93	20120601	ARIN	-
9	vultr.com	20473	1	64(0.16%)	21	20150305	ARIN	N
10	NA	54600	1	64(0.16%)	177	20120601	ARIN	-
11	NA	6147	1	62(0.15%)	2	20030901	LACNIC	-
12	NA	18779	1	56(0.1%)	86	20120424	ARIN	-
13	serverhub.com	62904	1	53(0.13%)	19	20140121	ARIN	Y
14	gddc.com.cn	-	-	42(0.1%)	3	20080328	APNIC	-
15	advancedhosters.com	3491	1	40(0.1%)	666	20040913	ARIN & LACNIC	Y

TABLE XIII: Top 15 Owner objects managing the detected sub-allocations. For readability purposes, one ASN is shown when an owner spans multiple ASes. Object size refers to the number of corresponding Whois objects collected per owner. "NA" indicates that a domain name was not found according to Whois records.

#	ID	Website	ASN	#Ases	Recycling Rate	#Detected Sub-allocations(%)	Object Size	Created Date (yyyyMMdd)	RIR
1	AV	alphavps.bg	62240	2	9.09%	15(0.04%)	38	20110121	LACNIC
2	HB	hostingbug.net	24940	1	7.69%	1(0.00%)	4	20090324	RIPE
3	HD	heberdomaine.com	24940	1	6.25%	1(0.00%)	3	20130322	RIPE
4	HV	hudsonvalleyhost.com	36352	1	4.33%	10(0.03%)	76	20121025	ARIN
5	UP	upc.ro	6830	1	3.63%	1(0.00%)	16	20011025	RIPE

TABLE XIV: Top 5 Owner objects managing detected sub-allocations with highest recycling rate. For readability purposes, one ASN is shown when an owner spans multiple ASes. Object size refers to the number of Whois objects collected per owner. ID column refers to an acronym we assigned to each owner.

of sub-allocations; of note is the recycling behavior of a subsidiary of "ColoCrossing", "Hudson Valley Host" with a 4% recycling rate.

Owner domain migration. From our interactions with the BPH ecosystem we found the following; Service providers and their subsidiaries, e.g. "ColoCrossing", tend to occasionally group multiple resellers in the same sub-allocation. Additionally, in 3 of our purchases from BPH services, we, their clients, were moved between sub-allocations when their IP addresses were blacklisted. In an effort to identify resellers and using the observation of inevitable client migration, we leverage the collected list of TLD+3 hosted on the detected sub-allocations by following their movements across sub-allocations at different time intervals.

Once migrated TLD+3 are found, we look for *groups* of TLD+3 moving together through sub-allocations by building a graph to find connected components. More specifically, we build a graph $G = (V, E)$ where V is the set of nodes representing sub-allocations, and an edge is created between two sub-allocations indicates a domain migration of more than 50% of one sub-allocations' TLD+3 set. After which, we generate connected components using a Python Package NetworkX [37], where each component represents a group of TLD+3 moving *together* between at least two sub-allocations. As a result, we identified 592 groups, a sample of 5 groups is provided in Table XV with a velocity rate of 1 day per sub-allocations.

The top moving group spanned 35 sub-allocations and was mostly serving Command and Control (CnC). Additionally, two other groups of moving TLD+3 had one common domain pattern (e.g., 713811.xyz & 939211.xyz) and lasted for over a year. We believe this is more likely a private customer of the service provider rather than a BPH service as one expects a variety of patterns when multiple clients are involved.

C. BPH Clients

We study the clients of the BPH services by analyzing the complete set of TLD+3 domains hosted on the detected sub-allocations. In total, we collected a set of 260M TLD+3, out of which 87.7% are found to have an average life of 1 day indicating the pervasiveness of fast fluxing likely to evade blacklisting. Using our commercial blacklist, BL-A, described in Section III-B, we find that 115K TLD+3 domains are present on this blacklist. Table XVI shows that 50% of the blacklisted domains are used to distribute malware or spyware, and 46% are running botnet command and control servers. We note that this blacklist focuses on these two activities and thus might be biased, but this does show what is likely a small subset of the malicious activities clients of these BPH services are engaged in.

Client domain migration. To track clients moving between sub-allocations and sometimes even BPH service(s), we follow TLD+3 movements over the detected set of sub-allocations and track the list of TLD+3 found to be hosted on sub-allocations at different time intervals indicating network movement. Out of a partially processed set, due to the large processing cost incurred while cross-matching TLD+3 across sub-allocations, we found that 1.6M TLD+3 have migrated between at least two sub-allocations. Network movement enables TLD+3 to survive for a longer period as can be seen in Table XVII which shows 5 selected TLD+3 hoping between at least 10 sub-allocations and lasting for months. On average, we found the moving TLD+3 to hop 3 detected sub-allocations (10 being the maximum) with an average life of 6.7 months.

#	#Sub-allocations	#TLD+3	#FQDN	#ASes	Life-Time(Start-End)	RIRs	Note
1	35	841	52	3	20110103-20160524	ARIN, RIPE	Spyware CnC
2	33	3713	3713	1	20150413-20160920	ARIN	`\p[4,5].\p[4,5].tld`
3	14	12352	754	2	20160116-20160824	ARIN	`\p[6].tld`
4	8	186	32	3	20140421-20160702	ARIN, RIPE	Mobile CnC
5	6	9347	6708	1	20150428-20160910	APNIC	Fast Flux

TABLE XV: Sample of 5 large groups of TLD+3 moving together between a large number of sub-allocations. Samples of TLD+3 for each group can be found in Table XIX in Appendix.

Type	Size(%)
Botnet CnC	53K(46%)
Spyware	43K(38%)
DriveBy Sources	13K(12%)
Dynamic DNS	3k (2%)
Credential Drop Sites	93 (0.08%)
Total	115K

TABLE XVI: Types of malicious activities found in 115K TLD+3.

TLD+3	Overall Age	Activity
1-factoring.ru	12 months	Spam [11]
a.biomuders.at	1.3 months	Carding [12]
apple-chasy-dlia-vas.ru	11.8 months	Pushdo botnet [13]
apilogin.ru	7.1 months	Malware [14]
iinbanks.ru	11.6 months	Spam & Phishing [15]

TABLE XVII: 5 Selected TLD+3 hoping at least 10 detected sub-allocations.

VI. Discussion

We have presented a method for accurately detecting malicious sub-allocations of network blocks and validated this method to the extent that we could. In this section, we will discuss some of the limitations of our method in terms of evaluation, scope, and robustness to evasion. We will also highlight a few of the potential follow-up studies that could be performed based on our method and findings.

A. *Limitations*

Ground Truth. One of the primary limitations of our method is the lack of ground truth information about which sub-allocations are actually controlled by BPH services, which are simply poorly managed, and how much malicious activity is truly emanating from each of these sub-allocations. To overcome this we have validated our classifier's performance using a number of data-sets and approaches. Our validation approach involved manually creating and using existing high quality labeled lists of benign and malicious sub-allocations and validating that our classifier is fairly accurate when trained and tested on these lists. Unfortunately, these lists tend to be smaller and might be biased towards the extremes since these sub-allocations are likely the easiest to investigate and correctly label. To overcome these possible biases, we also evaluated our classifier on what is likely a noisier set of labeled sub-allocations. As expected our classifier performs worse on this set, but it is difficult to measure the true precision and recall since the labeling of the sub-allocations is imperfect. We attempt to identify some features, most of which are not used by our classifier, to validate these results, but again this only gives us a sense of how well it is performing and not a precise metric. Finally, we ran the trained version of our classifier that had what we believe to have lower recall and higher precision over the entire set of sub-allocations we found in all five of the RIR Whois data-sets. We did some manual sub-sampling of detected sub-allocations. This again indicated that our classifier is fairly precise, but we could not measure the recall rate without expending more manual effort to explore undetected sub-allocations.

Scope of Detection. Our classifier is focused on detecting maliciousness within a particular type of network block that has been sub-allocated from the parent owner. This focus was chosen based on anecdotal reports of how BPH services have evolved to evade blacklisting, our own results from purchasing BPH, and our analysis of what is included in the blacklist. Again it is difficult to understand what recall of the overall BPH infrastructure we achieved. It is clear that we are detecting a large number of network blocks that are likely malicious and probably outperforming both AS reputation metrics and IP address blacklists, both of which these BPH services have adapted to evade.

Robustness of Detection. Since we are attempting to detect an adaptive adversary it is important to consider the robustness of any detection system. Again it is difficult to evaluate the robustness of our features and classifier. However, when choosing features we did consider robustness and selected some features, such as network utilization and domain churn that in order to evade would likely cause the BPH services to either increase their costs, decrease their client's rate of abuse, or become more susceptible to blacklisting.

There are other methods of avoiding detection, such as high rates of network block recycling, which exploits the fact that our approach takes a few months to detect malicious sub-allocations. Complicit service providers make it more difficult to detect malicious sub-allocations because their high sub-allocation recycling rates cause each sub-allocation to have a limited footprint in passive DNS before the BPH service is moved to another sub-allocation. An example of this is "Colocrossing", for which we were able to detect only 44 (out of 7K) sub-allocations using our Set-B trained classifier. Another evasion strategy is for BPH services to quickly abandon the sub-allocation, which again will drive up their costs of operation. These activities might also be anomalous and themselves be detectable by adding additional features and tracking domain movements more aggressively. However, our current detection approach will not be effective against actively colluding service providers.

[11] http://www.joewein.de/sw/spam-bl-1.htm

[12] https://krebsonsecurity.com/2016/05/carding-sites-turn-to-the-dark-cloud/

[13] http://www.malwareurl.com/ns_listing.php?as=AS48666

[14] http://www.malwareurl.com/ns_listing.php?ip=185.121.25.7

[15] http://www.joewein.de/sw/bl-log-2016-01-12.htm

B. Ethical Concerns

In order to study the potential illicit activities that may take place in sub-allocations, we conducted two types of purchasing; resellers packages and BPH hosting. We contacted several legitimate service providers to participate in their reseller programs in order to investigate whether and how reseller's information is propagated to Whois. During our interactions with the providers, we made sure that no harm was done to them in our experiments and investigations. Additionally, as mentioned earlier, we contacted the BPH services identified from underground forums for the purposes of understanding their malicious activities, ecosystem, infrastructure and to collect ground truth data. Our purchasing activity has been explicitly reviewed and approved by The University of Indiana. We did our best to avoid paying them whenever possible: specifically, we first requested test IP addresses, which are free and only for the extremely interesting targets that did not offer such free trials did we move forward to purchase their services. We believe that the value of our work outweighs the relatively minor ethical concerns resulting from the small financial support provided to these BPH hosting providers through our purchases.

C. Future Work

Our features, classifier, and analysis of this ecosystem is a starting point for detecting these evasive and agile BPH services. It exposes the wealth of information from Whois, passive DNS, and other sources that can be collected and transformed into useful features for detecting malicious network blocks. As future work, we plan on exploring the feasibility of leveraging our improved understanding of how these BPH services operate and the expanded ground truth provided by the results from this study to develop even more effective and timely detection approaches. These approaches could be based more on Whois data, recycling and re-branding patterns, and less or not at all on passive DNS which we have found to be a useful but slow detection feature. The goal is to create a system, similar in spirit to PREDATOR [38], that might be able to proactively, at sub-allocation registration time, predict if a sub-allocation will likely be benign or malicious.

VII. Related Work

Working with sub-allocations. To the best of our knowledge, there has been no prior academic work on leveraging IP address sub-allocation information to find blocks of malicious IP addresses at the sub-ASN granularity level. Mahjoub from OpenDNS has presented two systems that mention using sub-allocation information from Border Gateway Protocol (BGP) and Whois feeds. The first system presented in 2014 called Marauder [35] focused on identifying "leaf-ASNs" or Border Gateway Protocol (BGP) sub-allocations and using the network structure of BGP routing tables to detect anomalous and likely malicious leaf-ASN network blocks that are peering with legitimate service providers. The other system, SPRank [39], is based on what they describe as an "IP range fingerprinting" approach and focused on detecting the same threat we are exploring which is detecting BPH providers operating within legitimate service providers. Neither of these presentations provide enough details, such as methodologies, capabilities, or assessments of their systems to understand how they operate and how well they perform. What we can reconstruct from these presentations is that they reinforce the notion that large amounts of data can be potentially used to detect malicious sub-allocated network blocks. In our work, we have presented what we believe to be the first detailed methodology and evaluation of a machine learning based system for detecting malicious sub-allocated network blocks.

The only other network block based work we are aware of is the Spamhaus Edrop list [14] which is a smaller list of manually compiled malicious network blocks that we used for the training and validation of our machine learning based detection system.

Detecting Malicious AS. There have been several methods proposed for detecting malicious ASes. Chen conducted a study that analyzed longitudinal trends of malicious IP addresses and made the observation that many of them exhibit spatial correlations that can be mapped back to a small set of ASes [40]. Fire [13] was published in 2009 and was one of the first systems for methodically detecting BP ASes. The method Fire used was to aggregate information temporally and spatially from multiple blacklists in order to detect elevated concentrations of persistently malicious activity within an AS's IP address allocations. There have been multiple additional studies that have refined this blacklisting approach for detecting BP ASes [41], [42]. More recently, Shue, et al. [43] noted that BP ASes often fast-flux their BGP routing information to evade detection. ASwatch [7] leveraged this signal that fast-fluxing of BGP routing information is a strong indicator of a BP AS to build a classifier to detect BP ASes before they appear on blacklists.

We have used these prior studies as one feature in our detection system (i.e. AS ranking), but these systems alone can not produce a network block based detection model since they focus on detecting the parent AS. Thus, as part of our study we have identified a number of new features tailored to our specific goal of detecting fine-grained malicious sub-allocated network blocks that enable us to build an accurate detection system.

Legitimate Service Providers Abuse. Another source of concentrated abuse arises from poorly managed service providers that are either compromised or have lax vetting processes for their clients. Collins, et al. [44] showed that misconfiguration can be used to predict increased probability of future abusive activity from an AS. Zhang, et al. [45] also found a correlation between scanning results indicating evaluated levels of misconfiguration and abuse likely due to compromised systems. Other studies have also noted that legitimate services providers, including ISPs [46] and cloud hosting services [47], are often used for abusive hosting. When selecting the features for our classifier, we attempted to avoid detecting compromised infrastructure, however this is a challenging task because it is sometimes difficult to differentiate between malicious activity generated from none-compromised as opposed to compromised hosts. We found a few instances where our classifier detected compromised network blocks when a large portion of the hosts within the network block were compromised for a long period of time.

Fast fluxing. Corona, et al. [48], used features from passive DNS to detect fast-fluxed domain names. There have been a number of follow-up studies that have identified improved sets of features from passive DNS that can be used to detect fast-fluxed domain names [49], [50]. In our study, we have used this concept of identifying features from passive DNS that indicate higher rates of churn, but the features we have developed are tailored to our domain and integrate other signals of malicious network blocks.

VIII. Conclusion

In this paper, we have presented the first publicly described method for detecting malicious network blocks that have been sub-allocated. As part of our methods we have highlighted the ability to collect Whois information and merge this with passive DNS to create a set of features that can accurately detect malicious sub-allocated network blocks. This is useful for detecting modern BPH services that have evolved from operating purely static BP ASes to agile services that rent network blocks from large service providers. After validating our detection method we then use it to perform a large-scale study of the malicious network blocks that it detects. This study sheds new light on the techniques used by these BPH services to evade AS based reputation metrics and fine-grained IP address blacklisting. We also highlight the magnitude of this problem by detecting 39K malicious network blocks, none of which are allocated to what are considered BP ASes and many of which have little to no coverage in IP address blacklists.

The method we have presented is by no means optimal and we have pointed out some of the limitation and potential directions for future research in this area. Our hope is that the detection methods and ecosystem analysis we presented will serve as a building block to guide and spur additional open research on this topic. The end result ideally being increasingly effective and robust methods for detecting the evolving techniques used by BPH services to hide their infrastructure.

Acknowledgment

We thank Stefan Savage and Vern Paxson for their valuable feedback and our reviewers for their insightful comments and suggestions. We thank Paul Pearce for his help and support. We also thank those who provided data for our study, but wish to remain anonymous. This work was supported in part by National Science Foundation under grants 1223477, 1223495, 1527141, 1618493, 1237265, 1619620 and gifts from Google. Any opinions, findings, conclusions or recommendations expressed in this paper do not necessarily reflect the views of the NSF.

References

[1] B. Krebs, “Host of Internet Spam Groups Is Cut Off,” http://www.washingtonpost.com/wp-dyn/content/article/2008/11/12/AR2008111200658.html, 2008.

[2] M. Goncharov, “Criminal hideouts for lease: Bulletproof hosting services,” http://www.trendmicro.fr/media/wp/wp-criminal-hideouts-for-lease-en.pdf, 2015.

[3] “The spamhaus project,” https://www.spamhaus.org/.

[4] Alexa, “Alexa top global sites,” http://www.alexa.com/topsites, May 2015.

[5] “The top 500 sites on the web,” http://www.alexa.com/topsites/category/Computers/Internet/Web_Design_and_Development/Hosting.

[6] “Bgp ranking.” http://bgpranking.circl.lu/.

[7] M. Konte, R. Perdisci, and N. Feamster, “Aswatch: An as reputation system to expose bulletproof hosting ases,” in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, ser. SIGCOMM ’15. New York, NY, USA: ACM, 2015, pp. 625–638. [Online]. Available: http://doi.acm.org/10.1145/2785956.2787494

[8] “Colocrossing resellers program,” https://www.colocrossing.com/services/resellers.

[9] “Reseller partner program,” https://www.leaseweb.com/partner-programs/reseller.

[10] “Become an ovh partner,” https://partners.ovh.com/become-a-partner.

[11] “Voxility - iaas for service providers and large websites,” https://www.voxility.com/info.

[12] “Whmcs web hosting billing and automation platform,” https://www.whmcs.com/.

[13] B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda, “Fire: Finding rogue networks,” in *Computer Security Applications Conference, 2009. ACSAC ’09. Annual*, Dec 2009, pp. 231–240.

[14] “The spamhaus don’t route or peer lists,” https://www.spamhaus.org/drop/.

[15] “Farsight security information exchange,” https://api.dnsdb.info/.

[16] “Viruswatch – viruswatch watching adress changes of malware URL’s,” http://lists.clean-mx.com/cgi-bin/mailman/listinfo/viruswatch/.

[17] “Obtaining bulk whois data from arin,” https://www.arin.net/resources/request/bulkwhois.html.

[18] “Riswhois,” https://www.ripe.net/analyse/archived-projects/ris-tools-web-interfaces/riswhois.

[19] “Apnic bulk access to whois data,” https://www.apnic.net/manage-ip/using-whois/bulk-access.

[20] “Lacni request for bulk whois,” http://www.lacnic.net/en/web/lacnic/manual-8.

[21] “Afrinic bulk whois data,” https://www.afrinic.net/library/membership-documents/207-bulk-whois-access-form-.

[22] “Ripe database documentation,” https://www.ripe.net/manage-ips-and-asns/db/support/documentation/ripe-database-documentation/.

[23] “Registration data access protocol (rdap),” http://www.lacnic.net/en/web/lacnic/registration-data-access-protocol.

[24] “Registration data access protocol (rdap),” http://rdap.afrinic.net/rdap/.

[25] I. SIE, “Security information exchange (sie) portal,” https://sie.isc.org/.

[26] “The rokso list,” https://www.spamhaus.org/rokso/.

[27] “The wayback machine,” https://archive.org/.

[28] “Spamhaus isp area,” https://www.spamhaus.org/isp/.

[29] “Icq with video calls, free messages and low-cost phone calls,” https://icq.com.

[30] “Jabber. the original xmpp instant messaging service,” https://www.jabber.org/.

[31] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, “Exposure: A passive dns analysis service to detect and report malicious domains,” *ACM Trans. Inf. Syst. Secur.*, vol. 16, no. 4, pp. 14:1–14:28, Apr. 2014. [Online]. Available: http://doi.acm.org/10.1145/2584679

[32] C. Cortes and V. Vapnik, “Support-vector networks,” in *Machine Learning*, 1995, pp. 273–297.

[33] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[34] VirusTotal, “Virustotal - free online virus, malware and URL scanner,” https://www.virustotal.com/, 2013.

[35] D. Mahjoub, “Marauder or Scanning Your DNSDB for Fun and Profit,” http://www.slideshare.net/OpenDNS/marauder-or-scanning-your-dnsdb-for-fun-and-profit-source-boston, 2014.

[36] “Internet assigned numbers authority,” www.iana.org.

[37] “NetworkX,” https://networkx.github.io/.

[38] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, “PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration,” in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, 2016, pp. 1568–1579.

True Positive Rate (Recall) =	TP/ (TP+FN)
False Negative Rate (FNR) =	FN / (TP + FN) =1 - TPR
True Negative Rate (TNR)=	TN/ (TN+FP)
False Positive Rate (FPR)=	FP / (FP + tn)= 1-TNR
False Discovery Rate (FDR)=	FP/ (TP+FP)
Accuracy=	(TP + TN) / (TP+FP+FN+TN)

TABLE XVIII: Description of the evaluation metrics used. Acronyms TP, FP, TN, FN stand for, True Positives, False Positives, True Negatives and False Negatives respectively.

[39] D. Hubbard and D. Mahjoub, "Using Large Scale Data to Provider Attacker Attribution for Unknown IOC's," https://www.rsaconference.com/writable/presentations/file_upload/air-r04-using_large_scale_data_to_provide_attacker_attribution_for_unknown_iocs-.pdf, 2016.

[40] Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of internet malicious sources," in *INFOCOM 2008. 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 13-18 April 2008, Phoenix, AZ, USA*, 2008, pp. 2306–2314.

[41] *Malicious Hubs: Detecting Abnormally Malicious Autonomous Systems*, vol. Mini-conference 14: Secu, 03/15-19/2010 2010.

[42] F. Roveta, G. Caviglia, L. Di Mario, S. Zanero, F. Maggi, and P. Ciuccarelli, "Burn: Baring unknown rogue networks," in *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, ser. VizSec '11, 2011, pp. 6:1–6:10.

[43] C. A. Shue, A. J. Kalafut, and M. Gupta, "Abnormally Malicious Autonomous Systems and Their Internet Connectivity," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 220–230, Feb. 2012.

[44] M. P. Collins, T. J. Shimeall, S. Faber, J. Janies, R. Weaver, M. De Shon, and J. Kadane, "Using uncleanliness to predict future botnet addresses," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '07, 2007, pp. 93–104.

[45] J. Zhang, Z. Durumeric, M. Bailey, M. Liu, and M. Karir, "On the mismanagement and maliciousness of networks," in *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*, 2014.

[46] Z. Li, S. Alrwais, Y. Xie, F. Yu, and X. Wang, "Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures," in *Proceedings of the 2013 IEEE Symposium on Security and Privacy*, ser. SP '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 112–126. [Online]. Available: http://dx.doi.org/10.1109/SP.2013.18

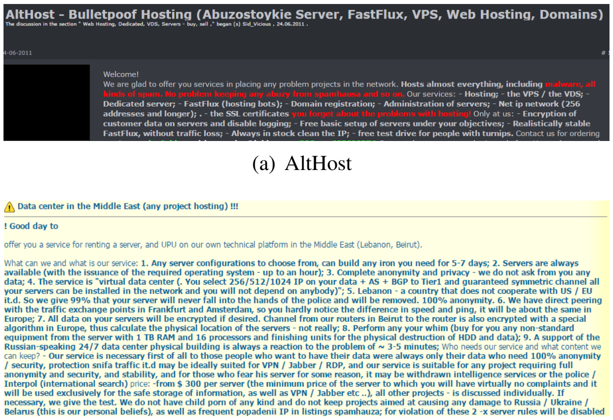
[47] X. Liao, S. Alrwais, and K. Y. et al., "Lurking malice in the cloud: Understanding and detecting cloud repository as a malicious service," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, 2016.

[48] I. Corona, R. Perdisci, and G. Giacinto, "Early detection of malicious flux networks via large-scale passive dns traffic analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. undefined, pp. 714–726, 2012.

[49] Z. B. Celik and S. Oktug, "Detection of fast-flux networks using various DNS feature sets," in *2013 IEEE Symposium on Computers and Communications, ISCC 2013, Split, Croatia, 7-10 July, 2013*, 2013, pp. 868–873.

[50] D.-T. Truong and G. Cheng, "Detecting domain-flux botnet based on dns traffic features in managed network," *Security and Communication Networks*, vol. 9, no. 14, 2016.

IX. Appendix



AltHost - Bulletpoof Hosting (Abuzostoykie Server, FastFlux, VPS, Web Hosting, Domains)

The discussion in the section " Web Hosting, Dedicated, VDS, Servers - buy, sell ," began (s) Sid_Vicious , 24.06.2011 .

4-06-2011 #

[redacted]

Welcome!

We are glad to offer you services in placing any problem projects in the network. **Hosts almost everything, including malware, all kinds of spam. No problem keeping any abuzy from spamhausa and so on.** Our services: - **Hosting; - the VPS / the VDS; - Dedicated server; - FastFlux (hosting bots); - Domain registration; - Administration of servers; - Net ip network (256 addresses and longer); . - the SSL certificates you forget about the problems with hosting!** Only at us: - **Encryption of customer data on servers and disable logging; - Free basic setup of servers under your objectives; - Realistically stable FastFlux, without traffic loss; - Always in stock clean the IP; - free test drive for people with turnips.** Contact us for ordering

(a) AltHost

Data center in the Middle East (any project hosting) !!!

! Good day to

offer you a service for renting a server, and UPU on our own technical platform in the Middle East (Lebanon, Beirut).

What can we and what is our service: **1. Any server configurations to choose from, can build any iron you need for 5-7 days; 2. Servers are always available (with the issuance of the required operating system - up to an hour); 3. Complete anonymity and privacy - we do not ask from you any data; 4. The service is "virtual data center (. You select 256/512/1024 IP on your data + AS + BGP to Tier1 and guaranteed symmetric channel all your servers can be installed in the network and you will not depend on anybody)"; 5. Lebanon - a country that does not cooperate with US / EU it.d. So we give 99% that your server will never fall into the hands of the police and will be removed. 100% anonymity. 6. We have direct peering with the traffic exchange points in Frankfurt and Amsterdam, so you hardly notice the difference in speed and ping, it will be about the same in Europe; 7. All data on your servers will be encrypted if desired. Channel from our routers in Beirut to the router is also encrypted with a special algorithm in Europe, thus calculate the physical location of the servers - not really; 8. Perform any your whim (buy for you any non-standard equipment from the server with 1 TB RAM and 16 processors and finishing units for the physical destruction of HDD and data); 9. A support of the Russian-speaking 24/7 data center physical building is always a reaction to the problem of ~ 3-5 minutes;** Who needs our service and what content we can keep? - **Our service is necessary first of all to those people who want to have their data were always only their data who need 100% anonymity / security, protection snifa traffic it.d may be ideally suited for VPN / Jabber / RDP, and our service is suitable for any project requiring full anonymity and security, and stability, and for those who fear his server for some reason, it may be withdrawn intelligence services or the police / Interpol (international search)** price: **-from $ 300 per server (the minimum price of the server to which you will have virtually no complaints and it will be used exclusively for the safe storage of information, as well as VPN / Jabber etc ..), all other projects - is discussed individually. If necessary, we give the test. We do not have child porn of any kind and do not keep projects aimed at causing any damage to Russia / Ukraine / Belarus (this is our personal beliefs), as well as frequent popadenii IP in listings spamhauza; for violation of these 2 -x server rules will be disabled**

(b) GlobalNetwork

Fig. 14: Forum posts of BPH advertisements (translated from Russian).

TLD+3	Start-EndDate (yyyyMMdd)	Note
mail236.viralmoneyraising.net	20150508-20150509	All 13 TLD+3 and 1 FQDN found on the detected network block owned by Irv Freiberg (216.246.108.224/28).
mail228.viralmoneyraising.net	20150101-20160807	
mail226.viralmoneyraising.net	20150101-20160807	
mail232.viralmoneyraising.net	20150510-20150510	
mail238.viralmoneyraising.net	20150510-20150510	
mail230.viralmoneyraising.net	20150510-20150510	
mail229.viralmoneyraising.net	20150509-20150520	
mail235.viralmoneyraising.net	20150509-20150509	
mail227.viralmoneyraising.net	20150101-20160807	
mail231.viralmoneyraising.net	20150509-20160611	
mail237.viralmoneyraising.net	20150510-20150510	
click.viralmoneyraising.net	20150101-20160807	
mail234.viralmoneyraising.net	20150509-20150509	
hg1q2s.online	20160707-20160820	6 short lived TLD+3 (out 397) hosted on (107.149.32.128/26) through VPS Quan.
fbgjz778.com	20150101-20150128	
slez778.com	20150101-20150216	
hg888u.racing	20160723-20160820	
lbl778.com	20150101-20150109	
f76ub.racing	20160804-20160813	
hgubt1.date	20160808-20160820	
hg8ry3.host	20160808-20160820	
g17tk.racing	20160810-20160820	
ritarorasco.com	20160213-20160307	Sample of TLD+3 in group 1
vizinurion.ru	20160214-20150607	
metiztransport.ru	20151125-20151215	
vvservop.at	20160528-20160820	
jufugers.ru	20150112-20150114	
lzhgt.xn3jl.xyz	20151210-20151210	Sample of TLD+3 in group 2
axbbr.fgvcb.xyz	20160116-20160116	
meqh.ergbd.xyz	20160116-20160116	
4tw6.dfhr3.xyz	20160111-20160111	
mz2c.rn1h7.xyz	20151204-20151204	
598770.top	20160814-20160820	Sample of TLD+3 in group 3
131622.xyz	20160713-20160820	
578290.top	20160814-20160820	
335235.xyz	20160619-20160820	
282720.top	20160814-20160820	
cobite.ddns.net	20160407-20160414	Sample of TLD+3 in group 4
sawa5001.no-ip.org	20150616-20150618	
justvirusahmed.zapto.org	20160406-20160411	
fuukrie.ddns.net	20160625-20160625	
xdayshk.ddns.net	20160408-20160410	
6z4ziw.csgwth25.com	20160702-20160702	Sample of TLD+3 in group 5
audel.dzqds3xt.com	20160710-20160710	
cc7qf.asgdy3xt.com	20160608-20160710	
rdkbtghbcgriztd.com	20150910-20150910	
07tl9.gfghj3xt.com	20160629-20160629	

TABLE XIX: TLD+3 hosted on various detected sub-allocations.