Domain Parking and the Typosquatting Feed

In an earlier post, we described the key elements of the domain parking ecosystem and discussed the risks typically stemming from a lack of appropriate regulation of this area. In the present post, we shall conduct a particular investigation revealing the connection between typosquatting, bulk domain registrations, and domain parking, by using WhoisXML API's Typosquatting Data Feed.

The Typosquatting Data Feed takes all second-level domains in all generic Top-Level Domains (TLDs) and some of the country-code TLDs that started to operate on the Internet on a given day. That is, these are newly registered or re-registered domains. It performs a lexical similarity-based clustering in search of groups of domains so that all domains in a group have similar names. Hence, the domain feed provides groups of newly registered domains that have been registered on the same day, are similarly named, and are frequently parts of bulk domain name registrations.

We have found that these sets of domains are closely related to many illicit or semi-legal activities on the Internet that deserve attention, including typosquatting, but also phishing, malware activity, etc. In addition to that, since 1 July 2020, the data are available in an "enriched" fashion, that is, part of the WHOIS information, and the IP addresses associated with the domains are also provided. We shall see below that this is very useful. So, let us see how it relates to domain parking.

1. The top 10 IPs of the Typosquatting Feed

As our starting point we use enriched data from the Typosquatting Feed from 2020-07-01 till the end of the year 2020; this 6-month period will be the starting point of our investigations. This initial sample contains 2,046,939 domains.

Even a manual observation of the IP addresses in the enriched data sets makes it clear that there are IP addresses that are very frequently associated with certain IP addresses. This leads to the idea of calculating the number of domains associated with IPs and checking the “popular” IP addresses. (Note that a domain can be related to multiple IPs, and in the case of many domains, there is no association to an IP available yet.)

Altogether there are 117,256 distinct IP addresses appearing in our sample, which means that there should be IPs which are more prone to be associated with groups of similarly named domains registered on the same day. And indeed, there are **240** IPs that are associated with at least 500 bulk-registered domains, and **1279** of them have at least 100. (Note, however, that the IPs with a few domains are also interesting in many cases, but they are not addressed in the present study.)

The top 10 IPs with bulk registrations from 2020-07-01 to 2020-12-31 are:

IP	# of connected domains	reverse DNS (nslookup)	IP owner
34.102.136[.]180	759,562	180.136.102.34.bc.googleusercontent[.]com.	Google, in use by Google Cloud customers
199.59.242[.]153	45,127	NXDOMAIN	BODIS
213.186.33[.]5	25,101	redirect.ovh[.]net.	OVH SAS, FR
94.136.40[.]51	17,380	SERVFAIL	Webfusion, Leeds, UK
51.195.76[.]235	17,310	ip235.ip-51-195-76[.]eu.	OVH, PL
184.168.131[.241	14662	p-184-168-131-241.ip.secureserver[.]net.	GoDaddy

192.161.187[.]200	13,675	192.161.187.200.static.quadranet[.]com.	QuadraNet Enterprises LLC, in a referral also “My Hosting 365”
209.141.38[.]71	13,675	parking.namesilo[.]com.	FranTech Solutions
107.161.23[.]204	13,672	parking.namesilo[.]com.	RamNode LLC, client: NameSilo
91.195.240[.]94	11,434	NXDOMAIN	Sedo Domain Parking

In the last-but-one column we check whether a domain name is directly associated with the IPs in the DNS. This can be done with a direct DNS lookup. While we expect that these are heavily shared IPs, thus not all of them will resolve, yet some do and the domain name already tends to reflect an IP, and thus infrastructure ownership. In the last column we also list the ownership information which can be found out with a direct WHOIS query or using WhoisXML API’s IP Netblocks API/lookup. A closer look reveals that these are big providers, all of whom are linked to domain parking. Next we take a closer look at the top 3.

2. Detailed analysis of the top 3

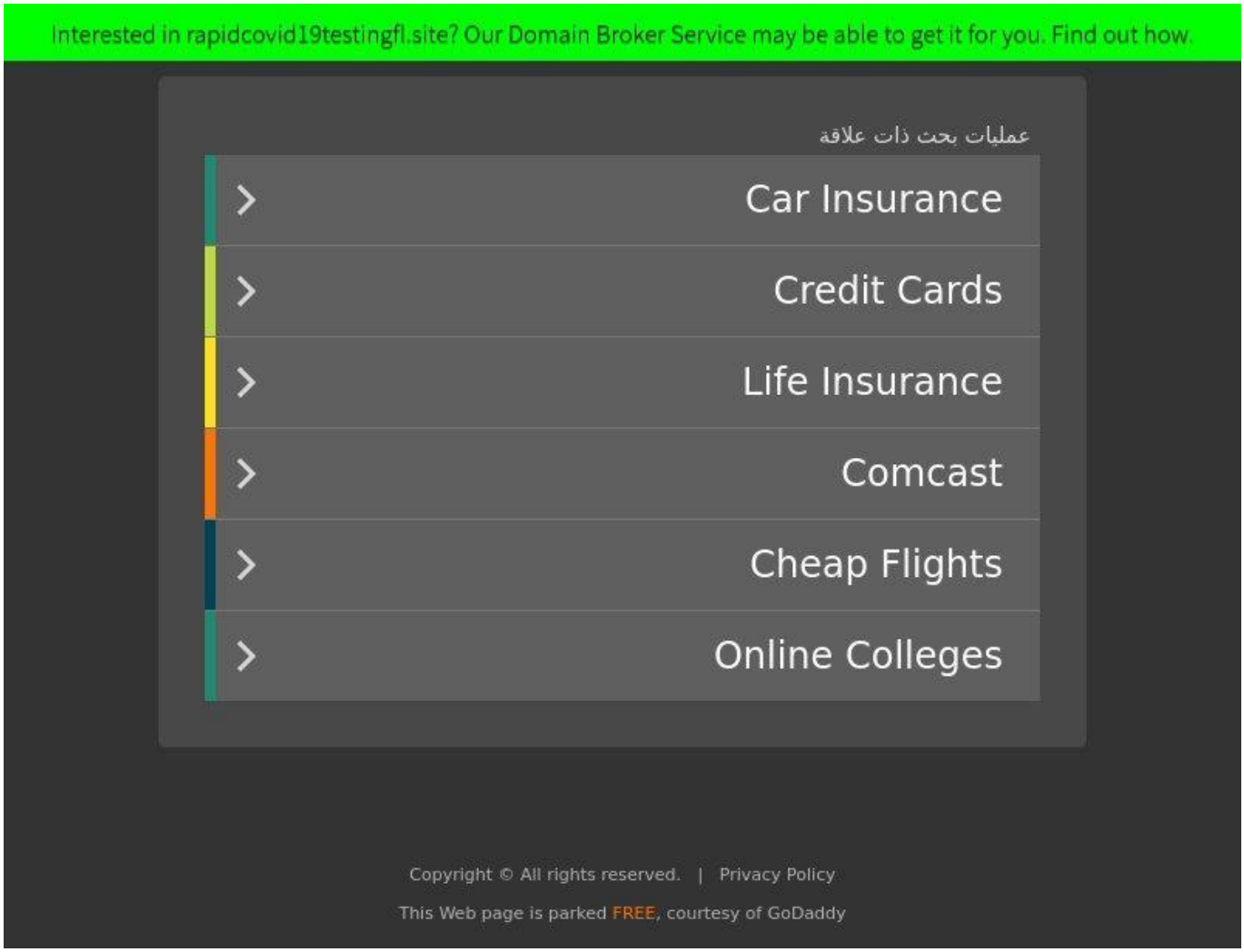
2.1 *34.102.136.180*

This IP is owned by Google and is in use by Google Cloud customers, as stated as a remark in the IP WHOIS data. A comment on IBM XForce Exchange claims that this is a landing page of GoDaddy. Taking a typical group of the many found by the Typosquatting feed, on 7 December 2020, the following 14 domains were captured:

No.	sld	tld
1	rapidcovid19testingfl	site

2	rapidcovid19testingde	site
3	rapidcovid19testingywca	site
4	rapidcovidtesting	info
5	rapidcovid19testingdc	site
6	rapidcovid19testingncaa	site
7	rapidcovid19testingusa	site
8	rapidcovid19testingla	site
9	rapidcovid19testing	info
10	rapidcovid19testingpa	site
11	rapidcovid19testingnj	site
12	rapidcovid19testing	live
13	rapidcovid19testing	site
14	rapidcovid19testing	org

The screenshot of the first one shows that it is indeed about domain parking at GoDaddy:



Consider also the group of 27 domains captured on 7 August by the typosquatting feed with the same IP, including these:

apple-billing-verificationy

iapple-billing-verification

apple-billing-verificationily

weapple-billing-verification

apple-billing-verificationable

etc., all registered in the ".com" TLD by a registrant from South Africa with GoDaddy. We cannot verify at the time of writing of this study whether they had been parked, but the likelihood of them being registered with an absolutely benign intention is low.

In addition to the 759,562 domains found on this IP by the typosquatting feed a passive DNS lookup also yields a tremendous amount of records. And, unfortunately, many of these are not quite benign; many blacklists contain this IP, and, for instance, in community-operated databases (see, for instance, https://www.abuseipdb.com/check/34.102.136.180) a tremendous amount of incident is reported ranging from ssh attacks through spamming, phishing, and blackmailing to botnet activity, etc.

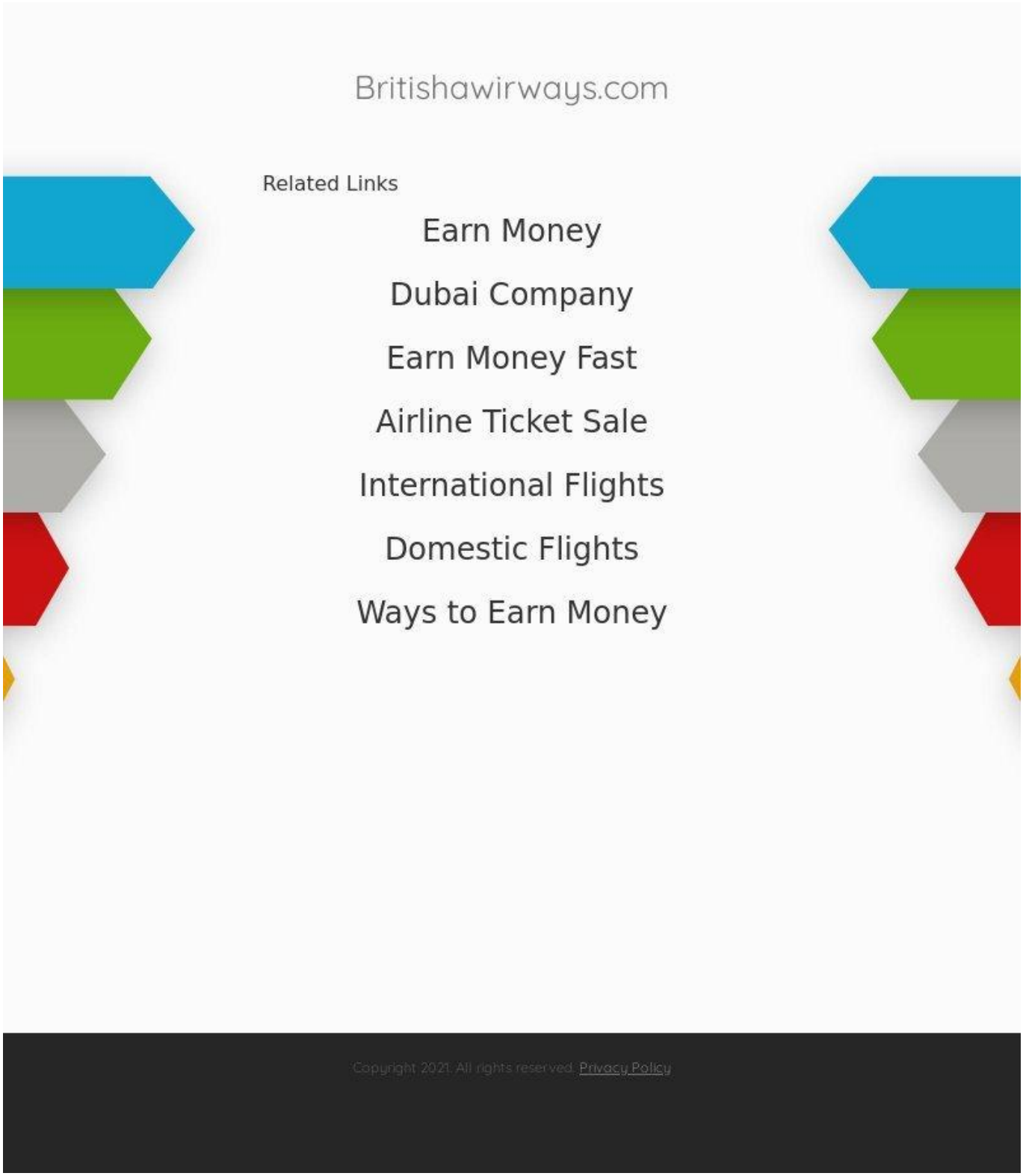
This clearly illustrates that on the one hand, big providers like Google and GoDaddy should put more effort into maintaining security measures. On the other hand, one would hardly recommend anyone to buy any of these domains that had dealings with this IP address before.

2.2 *199.59.242.153*

This is also a landing page of many parked domains, and the provider is “BODIS” this time. The IP itself has a similar reputation to the previous one, and passive DNS also has a very large number of related hits. Our 45,127 domains found solely by the Typosquatting Feed based on bulk registration detection contain a diverse zoology of bulk domain name registrations. As our example, let us quote a few from a 36-tuple of domain names, all in the TLD “.com”, captured on 23 July 2020:

baritishairways.com britishawirways.com britishairwauys.com britithairways.com britisgairways.com britishairyways.com britishairvays.com

This is a classic instance of typosquatting. A screenshot of one of these looks like that:



Again, a typical design of an ad-money collecting parking page. The WHOIS information provided in the enriched Typosquatting Feed reveals that the registrant of these 36 domains is an undisclosed entity from China, and the registration was made with Alibaba Cloud Computing (Beijing) Co., Ltd. as a registrar. We do not know if British Airways is aware of this, however, the outcome of a possible WIPO Domain Name Dispute can be easily predicted.

2.3 *213.186.33.5*

Here our provider is OVH, again a huge company who even has its own TLD. In this case, the parked pages are more correct and informative in that they clearly advertise OVH's services and the domain's availability for purchase. Yet they could not completely avoid issues: there are a few instances of typosquatting, like this one:

amazon-market	store
amazon-market	place
amazon-market	site
amazonmarket	place
amazon-market	shop

Meanwhile, e.g., from this group of 8 detected on 2020-11-03:

civolonstione	cloud
civolonstione	online
civolonstione	store
civolonstione	space
civolonstione	xyz
civolonstione	tech
civolonstione	shop
civolonstione	ovh

the marked two were phishing in November and December 2020. The IP itself is also known for abuse, though less markedly than the previous two.

3. Conclusions

Extending this analysis to the other IPs leads to essentially similar results:

- the IP address is of bad reputation and is associated with malicious activity;

- the associated domains are often parked and collect advertisement money;
- the names of the associated domains are often apparent typosquatting instances that could immediately be debated at WIPO;
- a small portion of these domains become part of phishing campaigns later on, whereas a large share of them can be considered as high risk because of their naming, WHOIS data, and other features.

From this, we can conclude that the domain parking ecosystem cries for a more rigorous regulation. Meanwhile, the use of the Typosquatting Feed can be a good starting point also in mitigating the risks associated with parked pages.