Ten years of attacks on companies using visual impersonation of domain names

Geoffrey Simpson
Tandy School of Computer Science
The University of Tulsa
Tulsa, Oklahoma 74104
Email: geoffrey@utulsa.edu

Tyler Moore
Tandy School of Computer Science
The University of Tulsa
Tulsa, Oklahoma 74104
Email: tyler-moore@utulsa.edu

Richard Clayton
Computer Laboratory
University of Cambridge
Cambridge, CB3 0FD, UK
Email: richard.clayton@cl.cam.ac.uk

***Abstract*—We identify over a quarter of a million domains used by medium and large companies within the .com registry. We find that for around 7% of these companies very similar domain names have been registered with character changes that are intended to be indistinguishable at a casual glance. These domains would be suitable for use in Business Email Compromise frauds. Using historical registration and name server data we identify the timing, rate, and movement of these look-alike domains over a ten year period. This allows us to identify clusters of registrations which are quite clearly malicious and show how the criminals have moved their activity over time in response to countermeasures. Although the malicious activity peaked in 2016, there is still sufficient ongoing activity to cause concern.**

I. Introduction and Background

Criminals have long been registering domain names for the purposes of fraud. A superficially similar domain name to that of a bank may make a 'phishing' email look more legitimate. A domain name that uses characters that are next to each other on the keyboard may cause poor typists to visit the wrong website. A domain name may even be registered to catch visits from people whose faulty hardware has 'flipped' a single bit of a character. In this paper we consider the registration of domain names that appear identical to an existing domain at a quick glance – which we deem visually impersonating domain names or VIDNs.

Previous work has considered the use of non-ASCII characters such as Greek or Russian glyphs that are indistinguishable from Latin letters [1], or accented characters where the accent may be too small to pick out reliably on a screen [2]. Here we consider very simple attacks where one character is substituted for another (such as G for Q – `qooqle.com`) or for a pair of characters (such as RN for M – `3rn.com`). Besides simplicity, these changes ensure that browsers or email programs never render the domain in its more detectable punycode form (`xn-...`).

We believe this type of look-alike domain is widely used in some types of Business Email Compromise (BEC) frauds. In these scams, customers are persuaded that they are corresponding with a legitimate company, when in fact a criminal has registered a visually impersonating domain name and uses it to trick victims to redirect payments, sometimes of very substantial sums, to their own bank account. BEC fraud has been tracked by the FBI's Internet Crime Complaint Center (IC3) since 2015 and losses have grown substantially over the succeeding years [3].

We start by identifying the domain names that are used by medium and large-size companies and then determine whether VIDNs have been registered during the ten year span 2009–2019. Inspection of the potentially malicious domains allows us to determine to a high degree of confidence which ones were registered by criminals and to then map the infrastructure that they used – demonstrating how this has changed over time. Unsurprisingly, we find that activity takes off in line with the growth of BEC. Over the past few years there has been a marked decline in new VIDN registrations, which we ascribe to previous hotspots of activity having deployed countermeasures and moved the criminals on.

II. Methodology

We first set out what types of visual impersonation we consider in this paper – that is, exactly what we will consider to be a 'visually impersonating domain name' or VIDN. We then explain how we used a dataset of medium and large companies to identify the domains that they use – and then what data we were able to collect about any VIDNs that might have been registered to attack those companies, or their customers.

A. Visual impersonation rules

A visual impersonation occurs when a letter, digit, or series of letters and digits are visually similar to another series of letters, digits, or series of letters and digits. In short, it is when one word visually looks like another word.

We focus on a handful of impersonations that are visually similar but are unlikely to be accidentally typed (so that we will be reasonably sure that the domain names we consider are not associated with 'fat-finger' typosquatting attacks). In Figure 1 we show how similar the two letter combination of lower case letter R and lower case letter N is to the single character lower case M, while below that we compare lower case VV to lower case W. We use the Calibri font, which is the default font of the Microsoft Outlook email application, which is very widely used in medium and large companies. Here we show a larger font size for clarity, but the default font

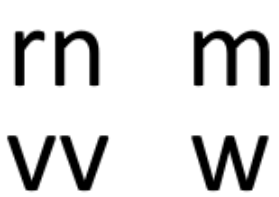

Fig. 1. rn visually compared to m, Calibri Font (top); vv visually compared to w, Calibri Font (bottom).

size in Outlook is 11 point, which makes it very hard to tell the cases apart.

There may be more than one change from the legitimate domain name. Consider the domain name `wombat.com`. A malicious actor seeking to register a VIDN would have their pick of `vvombat.com`, `wornbat.com` and `vvornbat.com`.

The full set of one- and two-character replacements that we consider is presented in Table I. Our VIDN candidate generation algorithm considers all possible replacements for each character (or bigram) in the original domain name. We exclude the top level domain from consideration, which means we do not consider attacks where `example.net` is used to attack `example.com`. Of course if ICANN were to add `.corn` to the list of top level domains then there would be even more VIDNs to consider.

Original character	replacement character	description
g	q	G for Q
q	g	Q for G
l	1	letter L for numeral 1
1	l	numeral 1 for letter L
o	0	letter o for numeral 0
0	o	numeral 0 for letter o
i	l	letter I for letter L
l	i	letter L for letter I
rn	m	RN for M
vv	w	VV for W
m	rn	M for RN
w	vv	W for VV

TABLE I
VISUAL IMPERSONATION REPLACEMENTS CONSIDERED IN THIS PAPER.

We recognize that many more visual replacements are possible than the ones considered here, including the Unicode character replacements considered in other research [4], [5], but as will be seen in the evaluation section, these visual replacements have been commonly used by criminals in the real world.

B. Identifying domain names used by companies

We elected to focus on companies as the targets of visual impersonation attacks because we know that some kinds of Business Email Compromise (BEC) attacks involve the use of look-alike email domains.

We use the Bureau van Dijk Orbis database, which holds data on over 375 million companies worldwide [6]. We selected all active US-based companies with at least 35 employees (approximately 381K firms), as well as non-US companies with at least 350 employees (approximately 184K firms). In total, this gave us 565,269 records. These records provide the Company Name, Website, NAICS Codes, and a Bureau van Dijk unique identifier, albeit not all are complete.

We picked out the hostname from the website URL, selected just the .com domains and isolated the second level. That is, from `www.example.com/index.htm` we selected `example.com`. We then excluded non-dedicated domains (such as when companies gave a Facebook page as their website). After filtering and cleaning the data we had a list of 269 759 company domain names.

C. Identifying VIDNs for company domain names

We obtained a dataset of .com zone file data from the Cambridge Cybercrime Centre, which provides a daily record of all domain name registrations and changes of name server from September 6, 2009 to June 23, 2019. Each record contains a domain name, name server name, and the start and finish dates that this entry was present in the zone file. Hence each domain can have many records, showing when it was registered (or re-registered), when it changed from one name server to another and, by deduction, when it expired altogether. The entire data set comprises 2 155 300 697 records spanning 307 765 190 unique .com domain names.

We applied our algorithm for identfying VIDNs (as outlined above) for all the 269 759 company domain names. 256 605 (95.1%) had at least one potential visual impersonation that an attacker would be able to register and many had many more than one. In total, these 256 605 domains have 249 383 735 potential VIDNs.[1] We then determined how many of these had ever been registered in the 2009–2019 time period.

Since several company domains might be attacked by the same VIDN (and company domains might be visually similar to each other – consider `gqco.com` and `qgco.com` both being attackable by `ggco.com`) we group together all the registered domains (company and VIDN) into what we call a 'canonical group'. This gave us 16 246 canonical groups containing 18 081 company names and 21 031 domain names that are VIDNs for these company names. That is to say, approximately 7% of medium or large-sized companies using a .com domain had at least one VIDN registered during 2009–2019 and therefore were potentially at risk.

We cannot of course be sure that any particular VIDN we identified was maliciously registered; it could be being used legitimately, but does not appear in the Orbis database. Further, even when a VIDN does appear to have registered for the purpose of fraud but there is more than one company in the canonical group, we cannot say which company was attacked. Naturally, we have no way of knowing what success any attack may have had.

[1]The mean number of VIDNs per domain is 971 but this is somewhat misleading – one domain has over 64 million possible VIDNs. The median value is 9 and 95% have 151 or fewer potential VIDNs.

D. The name server data

Registry zone files include the identity of the (one or more) authoritative name servers configured for each domain – typically of the form `ns1.example.com`, `ns2.example.com` etc. We extracted the second level domain name (e.g. `example.com`) from these records and deduplicated. The public suffix list was used to deal correctly with names such as `ns1.example.co.uk`.[2]

This allows us to track name server usage over time. The widespread usage of default name servers at registrars acts as a proxy for which registrar was used for initial registration. The data also indicates when domain names change hands and in particular when names are 'parked' at standard locations (usually to serve up adverts to any lingering trickle of visitors). The name servers also have the potential to help us distinguish some legitimate registrations, which choose to use name servers within the domain itself, whereas maliciously registered domains are seldom configured this way because of the extra complexity for the criminal.

E. Historical WHOIS data

We inspected historical WHOIS records provided by DomainTools. We searched for WHOIS records recorded by DomainTools within one week of the VIDN's first appearance in the .com zone file. 16 723 of 17 073 VIDNs had a matching historical WHOIS entry. We successfully parsed registrar information in all cases, and registrant email addresses for 88.2% of domains.

Additionally, as explained later in the paper, we queried DomainTools for additional domain names associated with a registrant email address. This enabled us to identify more malicious registrations beyond those satisfying our rules.

III. Empirical Analysis

We first document the prevalence of VIDN registration overall and by the number of impersonations per company. We then examine how attacker behavior has evolved over time, both in the number of attacks and the infrastructure hosts that are abused. We then distinguish between infrastructure used upon first registration and later on as domains are abandoned, resold and repurposed.

A. How many VIDNs impersonate each company?

One surprising finding is that most companies have not (yet) been targeted by a VIDN. 95% of the 269 759 companies had at least one potential visual impersonation using the simple rules outlined in Table I, yet only 18 081 of the 256 605 companies (7%) had at least one registered misspelling during the ten-year period we studied. The potential attack surface is much larger than has been actively exploited.

Thus, even though there is potential for a bias towards companies that match more of the visual impersonation rules we have set out, in practice this is not a big deal. For example, `williams.com`, with our rules, has the potential for substitution with four of the characters, 'vv' for 'w', 'l' for the 'l's and 'rn' for 'm'. This creates an opportunity for 24 different impersonations. In fact, only three had been registered.

[2]https://publicsuffix.org

Group Size	Count
1	15 698
2	1 783
3	466
4	65
5	13
6	4
7	9

TABLE II
Distribution of VIDNs per company.

Table II shows the breakdown overall in terms of impersonated domains. The vast majority of companies have only one registered visual impersonation. Around 10% of attacked companies face two impersonations, and larger numbers trail off quickly.

B. The evolution of attack-infrastructure targets

Figure 2 (left) plots the number of visually impersonating domain names based on the year the domain is first registered. We exclude from further consideration 4 083 domains that were already registered at the time our data began. While many are doubtless visual impersonations, we conservatively exclude them since they could be long-standing domains that are only similar to company websites by coincidence. The remaining 17 073 domains were first registered at some point after September 6, 2009.

We focus on those domains when first registered, since many domains are speculatively re-registered and put to use for other purposes (e.g., hosting ads or being offered for sale). While it is certainly possible that the domains are being used for impersonation years after initial registration, in practice we expect most abuse will occur shortly after the first registration, particularly since the vast majority of potential visually impersonating domains are never registered.

We can see from the plot that VIDNs were registered regularly between 2009–2012, before the rate accelerated in 2013 and 2014, peaking in 2015–2016. By 2019 the phenomenon had declined to roughly 1 000 registered domains per year.

But how many companies are targeted per year? Figure 2 (right) examines this, and the trends are broadly similar. This is as we would expect because in almost all cases there is only one associated VIDN registered for each company. See Section III-A above for more details.

What explains the huge rise, peak, and subsequent fall in activity? We can find some clues by inspecting the name servers used by the VIDNs. Figure 3 plots the number of visually impersonating domains split by the 21 most frequently-observed name servers.

A clear pattern readily appears. In the early years (2009–2012), there is no concentration in abuse at particular name servers. Beginning in 2013, though, one name server shot to

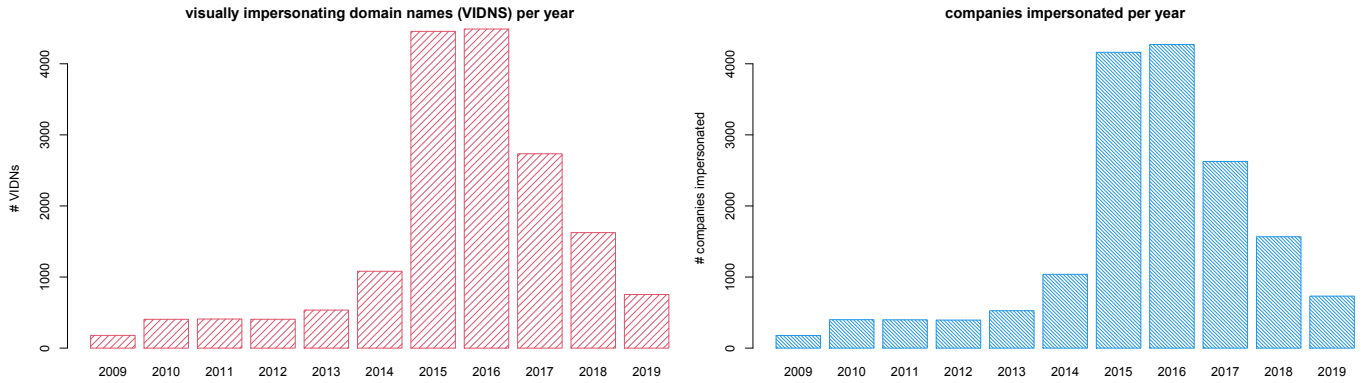

Fig. 2. Visually impersonating domain names created over time (left), along with companies affected (right).

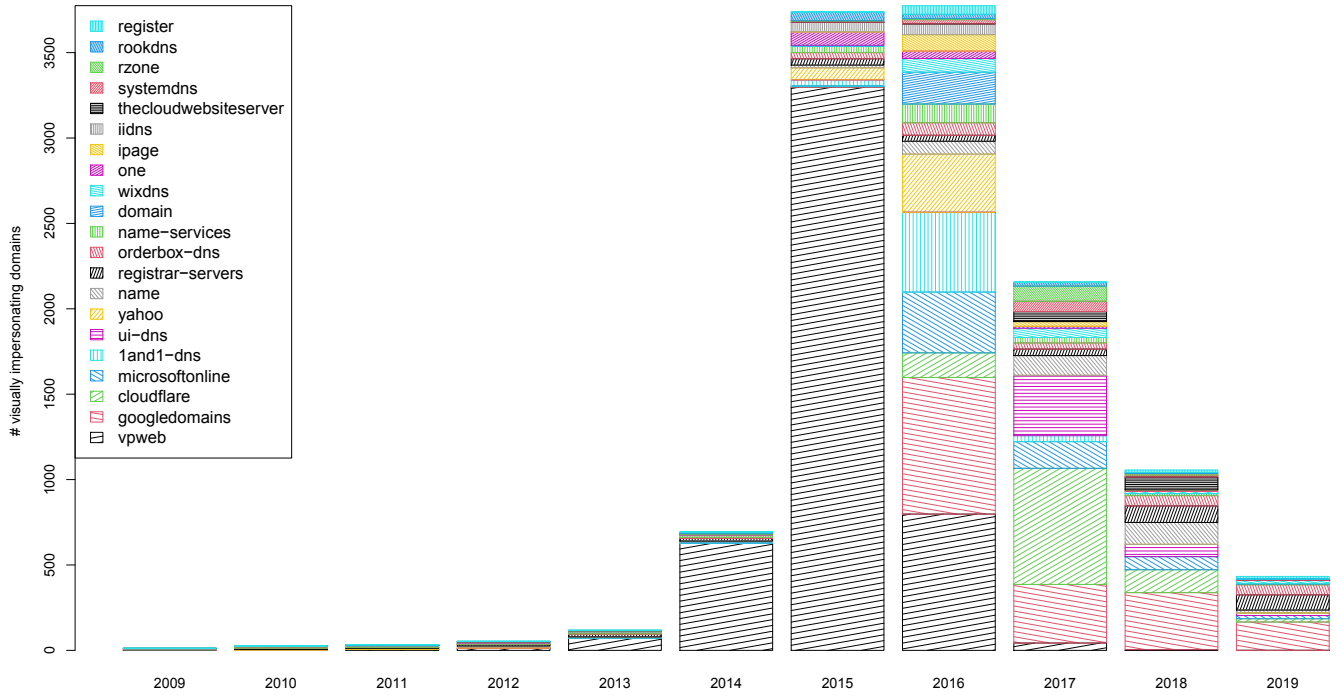

Fig. 3. Visually impersonating domain names created over time, grouped by origin name server.

prominence: `vpweb.com`, which is owned by Vistaprint. In fact, Vistaprint accounts for nearly all the growth in 2014–2015. By the time Vistaprint fell out of favor, attackers filled the gap using other name servers. In 2016, attackers could completely make up for the loss of Vistaprint, though eventually the entire phenomenon steadily declined.

Why did attackers focus on Vistaprint? They offered a free one month trial of their Web Builder product – which included some free business cards, but also a free domain name. Although Vistaprint collected a credit card number at the start of the trial they did not determine whether it was stolen (or invalid) until the month was up [7]. They certainly were not alone in attempting to entice potential customers, as Google also had similar promotions [8]. Nonetheless, Vistaprint was the most prominent. We know from prior research that once cybercriminals find a resource where they can be successful and are not quickly squashed, they often stick with it (and let others know of their discovery) [9].

What is particularly intriguing about this pattern is that it is consistent with an 'iterated weakest link' strategy [10] where attackers select a target and move onto others once the operators of one target gets a clue about the abuse taking place and start to deal with it. Once Vistaprint became less viable in 2016, miscreants moved onto a mix of other services, including `googledomains.com` (Google), `microsoftonline.com` (Microsoft), `1and1-dns.com` (1&1) and `yahoo.com` (Yahoo). When, in 2017, 1&1 and Yahoo ceased to be attractive, attackers moved on to services that used the name servers `cloudflare.com` and `ui-dns.com`. By 2018–2019, only Google's domain name business remained a primary target, and the total rates of abuse had diminished substantially from the 2015–2016 peak.

Name server (NS)	NS Type	1st-Use Domains	Later-Use Domains	% 1st-Use Domains	Coef. Var.
vpweb	originator	4 863	433	92	2.24
googledomains	originator	1 646	410	80	1.70
domaincontrol	dual	1 352	1 434	49	0.46
cloudflare	dual	982	526	65	2.28
microsoftonline	originator	608	149	80	2.01
1and1-dns	originator	550	162	77	2.78
ui-dns	dual	439	239	65	2.66
yahoo	originator	431	36	92	2.61
name	originator	365	145	72	1.47
registrar-servers	dual	353	356	50	1.02
hichina	dual	281	383	42	0.87
orderbox-dns	dual	273	131	68	1.16
name-services	dual	253	323	44	1.34
domain	originator	193	20	91	3.19
wixdns	originator	171	59	74	1.74
worldnic	originator	147	59	71	0.66
one	originator	140	41	77	1.99
ipage	originator	137	30	82	2.31
iidns	dual	133	135	50	1.97
thecloudwebsiteserver	dual	133	60	69	2.25

TABLE III
NAME SERVERS WITH THE MOST FIRST-USE DOMAINS.

Name server (NS)	NS Type	1st-Use Domains	Later-Use Domains	% 1st-Use Domains
renewyourname	recycling	4	5 253	0.076
systemdns	recycling	113	1 487	7.1
domaincontrol	dual	1 352	1 434	49
foundationapi	recycling	1	635	0.16
cloudflare	dual	982	526	65
vpweb	originator	4 863	433	92
googledomains	originator	1 646	410	80
hichina	dual	281	383	42.3
domainparkingserver	recycling	7	382	1.8
registrar-servers	dual	353	356	50
dnspod	recycling	93	335	22
name-services	dual	253	323	44
dnsdun	recycling	30	289	9.4
dns	recycling	36	287	11
ui-dns	dual	439	239	65
ns36	recycling	0	179	0.00
expirenotification	recycling	1	178	0.56
1and1-dns	originator	550	162	77
namebrightdns	recycling	7	159	4.2
ztomy	recycling	1	157	0.63

TABLE IV
NAME SERVERS WITH THE MOST LATER-USE DOMAINS.

C. Comparing hosting infrastructure from first to later uses

We have observed that the name servers being used when a website is first registered often differs greatly from those used later in a domain's lifetime. Put simply, some name servers appear to be preferred by miscreants for the initial registration, and these differ greatly from those used by later attempts to monetize the domain.

Let us now compare two widely used name servers: `vpweb.com` (Vistaprint) and `domaincontrol.com` (GoDaddy). Figure 4 plots the number of ingress and egress name server entries for `vpweb.com`. Here, ingress means any transition from unregistered or using a different name server to using `vpweb.com`. Egress means any change from `vpweb.com` to a different name server. Around 90% of `vpweb.com`'s egress entries are for domains that were previously unregistered. Another 5% represented a change from one `vpweb.com` name server to another, with the balance split among others. By contrast, around 80% of changes from `vpweb.com` name servers go to the `renewyourname.net` name server.

The behavior on `domaincontrol.com` is very different, as shown in Figure 5. Here we see a wide distribution of ingress and egress name servers, with concentration only in and out of `domaincontrol.com` itself.

Given the very different behavior, we sought to investigate whether the highly targeted name servers all exhibited similar behaviors, and if this is consistently different for the subsequent name servers we observed.

Table III shows the most frequently utilized name servers first used by visually impersonating domains. 92% of the domains using Vistaprint name servers in our dataset were '1st-use', viz: this name server was used when the domain was first registered. Similarly, domains hosted at `googledomains.com` (80%), `microsoftonline.com` (80%) `1and1-dns.com` (77%) and `yahoo.com` (92%) were all much more likely to there at 1st-use, rather than subsequently.

The right-most column in Table III reports the coefficient of variance (CV) for the number of visually impersonating domains hosted annually by these name servers. A coefficient less than 1 suggests that the number of domains hosted each year is relatively stable over time, whereas coefficients greater than 1 indicate high variability from one year to the next. We can see that `domaincontrol.com`, `hichina.com` and `worldnic.com` have low CV, suggesting that attacks did not concentrate there for shorter periods, whereas most others exhibit high variability from one year to the next.

Table IV indicates the most common name server used subsequent to the first name server. Top of the list is `renewyourname`, where 5 253 of 5 257 domains were served by other name servers initially (unsurprising, given its name). Similarly `systemdns`, `foundationapi`, `domainparkingserver` and `dnsdun` are almost never used to serve a newly registered visual impersonating name, but are widely used subsequently. We anticipate that the vast majority of these later uses happen after the initial impersonation has taken place, once the domains fall into the domain reseller and repurpose markets.

We build on these observations to classify name servers into the following groups:

- **Originator**: name servers in which at least 70% of domain names hosted are 1st-use (minimum 10 domains served);
- **Recycling**: name servers in which at most 30% of domain names hosted are 1st-use (minimum 10 domains served);
- **Dual**: name servers in which between 30–70% of domain names hosted are 1st-use (minimum 10 domains served);
- **Niche**: name servers hosting fewer than 10 domains in the dataset.

Tables III and IV include a column indicating the grouping assigned to the top 1st-use and later-use name servers. We can

Fig. 4. Name Server ingress and egress for vpweb.com

see that for the top first-use name servers, all are originators or dual use. For the top later-use name servers, most are recycling, with a few dual and top originators.

Table V breaks down the results overall. The vast majority of name servers observed are niche (1 545), but these account for only 6.8% of 1st-use domains and 8.9% of later-use domains. The 37 originators account for 58.6% of all observed 1st-use domains, while the 84 recycling name servers account for 57.7% of later-use domains. The dual-use name servers account for a more balanced but lower share.

What can we glean from this table? First, a small number of originator name servers accounted for most of the initial usage of visually impersonating domain names. Countering this abuse could have been concentrated at these hosts as well. Moreover, the recycling and dual name servers produce quite a bit of noise that can largely be ignored if the goal is to disrupt attacks leveraging visual impersonation of companies.

Category	# NS	1st-Use		Later-Use	
		#	%	#	%
originator	37	10 083	58.6	1 776	8.7
recycling	84	502	2.9	11 716	57.7
dual	73	5 471	31.8	5 006	24.7
niche	1 545	1 165	6.8	1 804	8.9

TABLE V
NAME SERVER CLASSIFICATION.

IV. What can we learn from historical WHOIS?

Thus far, we have identified connections between VIDN registrations through the name servers used. In most cases, we would expect that more than one criminal actor utilizes the same name server. Registrant information provided to WHOIS can offer more direct evidence of a relationship between VIDN registrations.

A. Registrants who register multiple VIDNs

In some cases, registrant information was obscured, either with privacy WHOIS or by listing the service provider as registrant. Regrettably, the latter is what happened with Vistaprint. For all Vistaprint VIDNs, the registrant is listed as Vistaprint rather than the individual who signed up for the domain.

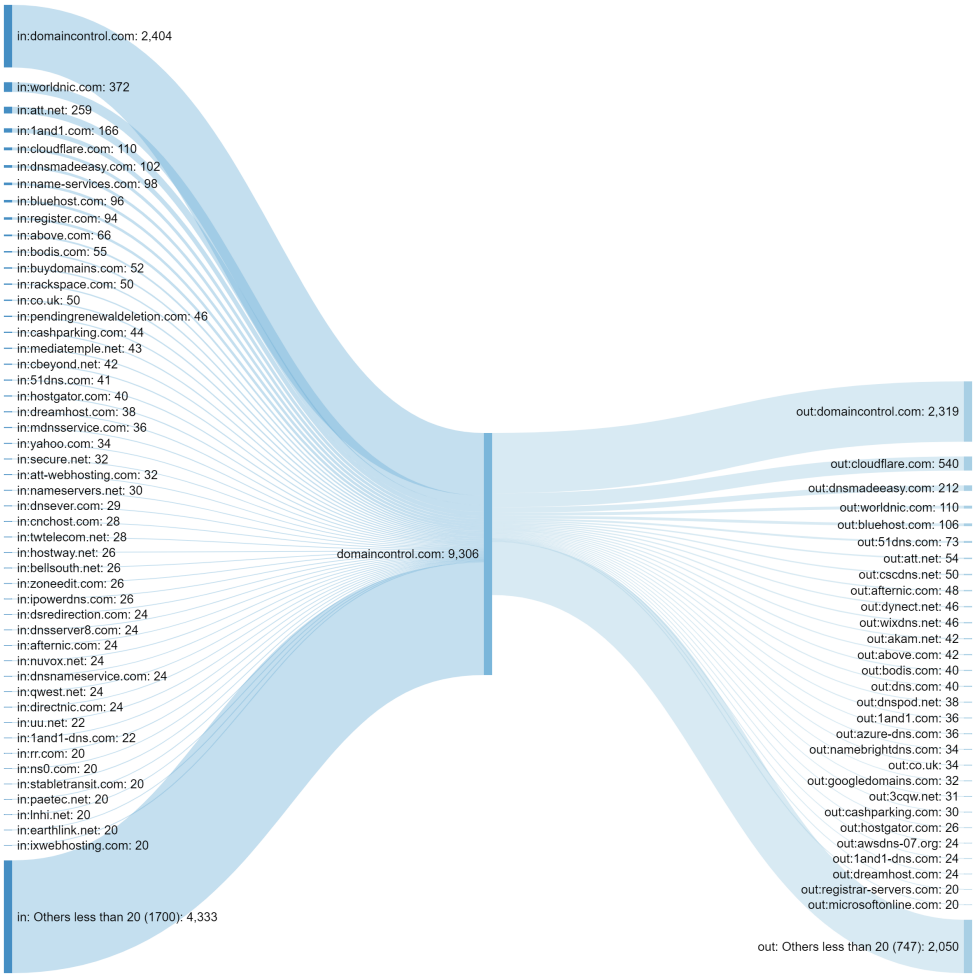

Fig. 5. Name Server ingress and egress for domaincontrol.com

Nonetheless, we can still learn quite a bit about linkages between VIDNs. In theory, a cybercriminal worth her salt would either enable privacy protections or use throwaway email addresses and have a different name and address for every registration. Doubtless, many people do. Nonetheless, there remains a substantial number who use the same contact information when registering multiple VIDNs.

Table VII indicates the number of registrant emails that sign up for different numbers of VIDNs. For example, while 6 035 VIDNs have unique registrant email addresses (3 382 of which are obscured by privacy protections), 85 email addresses have between 5 and 10 VIDNs.

We decided to drill down and focus on just those registrants who registered more than 5 domains but did not hide their contact information. After excluding emails associated with providers (e.g., Vistaprint), we are left with 92 users registering a total of 775 VIDNs.

Table VI shows information for all of these registrants, presented in the order in which the user first registered a VIDN. Consider the registrant email `bunsourr1965@gmail.com` (row is bolded in the table). Five domains were registered, all with registrar 1&1 Internet SE: `caliberpavinq.com`, `qeritommedical.com`, `lefthandbrewinq.com`, `leveraqeis.com` and `pmppropertyqroup.com`. Note that all five of the attacked domains have nothing in common sectorally – there is a brewery, roadworks contractor, medical provider and real estate company. What they do have in common is how they visually impersonate each company, all using use the g-to-q substitution. Moreover, we can see that the first domain was registered on 2016-04-05, while the last was registered two weeks later on 2016-04-19.

The very next entry is for `jmaddy421@gmail.com` which registered five domains beginning just two days later, on 2016-04-21, using the same registrar as `bunsour1965@gmail.com` did. Moreover, the impersonating domains also use the same g-to-q substitution: `barlosiqns.com`, `hohmanplatinq.com`, `inteqritypays.com`, `rainbowqraphics.com` and `reliablecontractinq.com`

While we cannot state definitively that these two registrants are in fact the same criminal, the circumstantial evidence is

Registrant	Registrar	1st Domain	Last Domain	# Days	# VIDNs
legalizationalism@legislator.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2014-01-30	2015-03-11	405	6
tai4ted@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2014-06-27	2017-05-08	1 046	5
fredchalson@gmail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2014-08-03	2015-07-07	338	6
YuMing@YinSiBaoHu.AliYun.com	HICHINA ZHICHENG TECHNOLOGY LTD.	2014-10-12	2018-05-06	1 073	97
tonitoney@gmail.com	GoDaddy.com, LLC	2015-07-08	2015-07-30	22	5
823015516@qq.com	eName Technology Co.,Ltd.	2015-09-27	2015-09-27	0	3
823015516@qq.com	Hangzhou Aiming Network Co.,Ltd	2015-09-27	2015-09-27	0	3
nino.brown3000@yahoo.com	1&1 Internet AG	2015-09-30	2015-10-09	9	6
13950170988@163.com	HICHINA ZHICHENG TECHNOLOGY LTD.	2015-10-06	2015-11-12	37	8
s.frayne@dsgschool.net	GoDaddy.com, LLC	2015-10-13	2015-10-13	0	1
s.frayne@dsgschool.net	TUCOWS, INC.	2015-10-19	2015-10-20	1	4
alexandercregier@ebooksseller.com	1&1 Internet AG	2015-10-21	2015-10-26	5	6
oneofakind414@outlook.com	1&1 Internet AG	2015-10-28	2015-10-29	1	6
69999@qq.com	Bizcn.com,Inc.	2015-11-04	2015-11-05	1	5
offlceor@offlce.org	PDR Ltd. d/b/a PublicDomainRegistry.com	2015-11-04	2015-12-23	49	6
2622580066@qq.com	WEBCC	2015-11-09	2015-11-11	2	6
jbellinato@treaddwwayscorp.com	Rebel.com	2015-11-15	2015-11-15	0	5
927822@qq.com	Hangzhou Aiming Network Co.,Ltd	2015-11-18	2015-11-20	2	6
knell.gary@yahoo.com	Domain.com, LLC	2015-11-30	2015-12-07	7	5
CENTS@DR.COM	ENOM, INC.	2015-12-16	2016-01-07	22	21
KMORNEY213@GMAIL.COM	ENOM, INC.	2016-01-11	2016-01-21	10	9
ginaemohr@yahoo.com	Ascio Technologies, Inc	2016-01-22	2016-01-26	4	6
nobiedhillbrannon@yahoo.com	Ascio Technologies, Inc	2016-01-26	2016-01-28	2	6
BEATRICE.LAFON@CLALRES.COM	FastDomain Inc.	2016-02-10	2016-02-12	2	5
bin1255218@163.com	Hangzhou Aiming Network Co.,Ltd	2016-02-18	2016-02-29	11	5
jejeabh@gmail.com	1&1 Internet SE	2016-03-24	2016-05-02	39	7
bunsourr1965@gmail.com	**1&1 Internet SE**	**2016-04-05**	**2016-04-19**	**14**	**5**
jmaddy421@gmail.com	1&1 Internet SE	2016-04-21	2016-04-27	6	5
ckogovsek@keystoneprn.com	Domain.com, LLC	2016-05-11	2016-05-11	0	5
jockoverfelt435@yahoo.com	Domain.com, LLC	2016-05-11	2016-05-17	6	10
dson00901@gmail.com	Domain.com, LLC	2016-05-12	2016-05-18	6	7
woodhouseheart@gmail.com	Register.com, Inc.	2016-05-12	2016-05-13	1	5
kirstenfebz@gmail.com	MESH DIGITAL LIMITED	2016-05-24	2016-05-26	2	16
new@kolomaster.net	MESH DIGITAL LIMITED	2016-05-30	2016-05-31	1	5
remax311@gmx.com	Domain.com, LLC	2016-06-06	2016-06-07	1	6
jamesliilio@yandex.com	MESH DIGITAL LIMITED	2016-06-13	2016-06-13	0	11
bankymoney8@gmail.com	Domain.com, LLC	2016-06-15	2016-06-21	6	5
richardclinton1111@gmail.com	Domain.com, LLC	2016-06-22	2016-06-23	1	5
dennis_chapman01@hotmail.com	1&1 Internet SE	2016-07-05	2016-07-18	13	6
steph7webb@att.net	Domain.com, LLC	2016-07-05	2016-07-13	8	6
sally.tillman@gmx.com	Domain.com, LLC	2016-07-06	2016-07-20	14	19
manager@weererr.com	MESH DIGITAL LIMITED	2016-07-14	2016-07-14	0	5
gdowns@srnwautoblok.com	Domain.com, LLC	2016-07-20	2016-07-21	1	5
moneyharuna@gmail.com	Domain.com, LLC	2016-07-21	2016-07-22	1	9
sambornetals1@gmail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2016-07-21	2016-11-29	131	5
ownerwire@gmail.com	Domain.com, LLC	2016-07-22	2016-07-24	2	8
rshine@manltobacorp.com	Domain.com, LLC	2016-07-27	2016-08-11	15	6
pinidnuj@aol.com	1&1 Internet SE	2016-08-01	2016-08-03	2	5
eilleen306@gmx.com	Domain.com, LLC	2016-08-02	2016-08-31	29	36
officialportal1@gmail.com	Domain.com, LLC	2016-08-09	2016-08-11	2	6
rroth@romanrnfg.com	Domain.com, LLC	2016-08-09	2016-08-09	0	5
adeball929@yahoo.com	Domain.com, LLC	2016-08-15	2016-08-19	4	10
via-1@mail.com	Name.com, Inc.	2016-08-18	2016-08-21	3	9
lydiamobarak@yahoo.com	1&1 Internet SE	2016-08-25	2016-09-12	18	5
caliiseluise@outlook.com	1&1 Internet SE	2016-09-04	2016-09-06	2	5
jhawkin@envinologic.com	GoDaddy.com, LLC	2016-09-13	2016-09-28	15	12
gregapel@aol.com	GoDaddy.com, LLC	2016-09-14	2016-09-23	9	6
scot.fluharty@infrasourcelnc.com	Google, Inc.	2016-09-15	2016-09-19	4	5
emaier@qwlisk.com	Google Inc.	2016-09-22	2016-09-22	0	5
jayhoekstra@outlook.com	1&1 Internet SE	2016-10-12	2016-10-14	2	10
dhoffarth@citadeldrillinq.com	1&1 Internet SE	2016-11-30	2016-12-08	8	8
vickycox1985@yahoo.com	1&1 Internet SE	2016-12-14	2016-12-22	8	8
akhzah@outlook.com	TUCOWS, INC.	2016-12-18	2016-12-18	0	1
akhzah@outlook.com	REGISTER.IT S.P.A.	2017-01-02	2017-01-04	2	6
anoldluna@gmail.com	Cronon AG	2017-02-05	2017-02-07	2	15
hvdinternationa@gmail.com	Cronon AG	2017-02-13	2017-02-14	1	5
vty9001@outlook.com	REGISTER.IT S.P.A.	2017-02-21	2017-02-22	1	6
bovona@12storage.com	Cronon AG	2017-03-06	2017-03-06	0	5
jdykes221@outlook.com	1&1 Internet SE	2017-03-07	2017-03-13	6	5
amber@selbybooks.com	1&1 Internet SE	2017-03-27	2017-03-27	0	5
anold.luna@12storage.com	Cronon AG	2017-03-27	2017-03-28	1	6
mkisbsn@hotmail.com	1&1 Internet SE	2017-04-07	2017-04-12	5	5
m.willy8097@outlook.com	1&1 Internet SE	2017-05-07	2017-05-08	1	5
fapu@lenovog4.com	Cronon AG	2017-05-09	2017-05-10	1	18
yebawoxe@xperiae5.com	Cronon AG	2017-05-12	2017-05-12	0	5
wirelord19900@gmail.com	NameSilo, LLC	2017-06-11	2018-09-27	473	5
deighton.levee@neusstone.com	1&1 Internet SE	2017-06-22	2017-06-28	6	5
pingkeehong2017@gmail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-06-22	2017-12-08	169	6
arleenprado9090@outlook.com	1&1 Internet SE	2017-07-12	2017-07-13	1	5
mskills@mail.com	1&1 Internet SE	2017-07-18	2017-07-18	0	5
dm31310@mail.com	Domain.com, LLC	2017-09-11	2017-10-09	28	5
guriet769@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-09-19	2017-09-24	5	5
jvilla989@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-09-20	2018-01-21	123	11
msd910@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-10-24	2018-01-10	78	6
jackdans221@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-10-25	2019-02-13	476	10
jstark0385@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-11-06	2017-12-19	43	12
mkane0385@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-11-06	2017-12-05	29	18
jjackson0385@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2017-12-04	2018-01-23	50	5
mharrison0385@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2018-01-05	2018-01-06	1	6
act8989@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2018-01-14	2018-01-23	9	13
jthayes322@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2018-02-11	2018-03-05	22	8
jwoods0385@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2018-02-11	2018-03-12	29	7
info@qhoster.com	NameSilo, LLC	2018-02-20	2019-04-29	433	7
angelcapri0909@outlook.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2019-01-12	2019-02-05	24	6
besta908us@mail.com	PDR Ltd. d/b/a PublicDomainRegistry.com	2019-03-19	2019-04-02	14	6

TABLE VI

REGISTRANTS WITH AT LEAST 5 VIDNS, WITH REGISTRAR, TIME OF FIRST AND LAST DOMAIN REGISTRATION, AND NUMBER OF VIDNS REGISTERED.

	VIDNs per registrant								
	1	2	3	4	5 −10	11 −20	21 −50	51 100	> 100
Regular	2 653	291	135	61	84	12	4	3	1
Privacy	3 382	10	2	2	1	4	2	1	2
Total	6 035	301	137	63	85	16	6	4	3

TABLE VII
NUMBER OF OBSERVED REGISTRANTS BROKEN DOWN BY THE TOTAL NUMBER OF VIDNS AND WHETHER OR NOT PRIVACY/PROXY WHOIS WAS USED.

certainly mounting. Stepping through the rows of the table in this order reveals many additional examples of potentially linked registrants.

Moreover, we can see patterns in which registrars are targeted by malicious registrants over time. For example, `Domain.com` was widely used between May and August 2016, and then only appears once more in September-October 2017. Cronon AG is used several times starting in February 2017 and concluding in May 2017. Apart from a few extremely long-lived user accounts dating to 2014, `publicdomainregistry.com` experienced a surge in registrations beginning in September 2017 and continuing through the end of data collection in 2019.

We also observed that nearly all registrant email addresses were used on only a single registrar. In only 3 cases (`823015516@qq.com`, `s.frayne@dsgschool.net` and `akhzah@outlook.com`), did an email address get used for two registrars. In all other cases, the address was used only at a single registrar. This suggests that attackers found it convenient (or otherwise more cost-effective) to re-use email addresses at the same registrar. When moving to other registrars, then it makes more sense to provide different registrant information. This also suggests that sharing abusive registrant information between registrars is unlikely to be effective in countering the threat.

Finally, as noted earlier, we do not presume that the letter-substitutions we have analyzed are comprehensive. We again use data from DomainTools to quantify the additional scope for VIDNs. We cross-referenced 43 of the registrant emails from Table VI with other historical WHOIS registrations made with the same email address. For example, `bunsour1965@gmail.com` registered 59 domains, the vast majority of which appear to be VIDNs. Some follow the same pattern (e.g., `legrandmarketinq.com`), whereas others follow different patterns (e.g., inserting 1 into long company domain names, or swapping character orderings).

In total, for these 43 registrant emails, we found 653 matches in our datasets. But using the cross-referenced data, these registrants actually registered 11 388 domains, which is more than 17 times as many as first uncovered. This provides some indication that the true scope of VIDN abuse is an order of magnitude higher than the totals reported in this paper.

B. Do companies defensively register VIDNs?

With the potential for financial losses both to the companies and to their customers, a reasonable countermeasure would be for companies to defensively register VIDNs.

We detect defensive registration by comparing the registrant email address listed in the company domain WHOIS record to those reported on its corresponding VIDN(s). We used the most expansive definition for matching. We compared all registrant email addresses associated with the legitimate domain since 2008 to all registrant email addresses associated with each VIDN. For example, both `1800flowers.com` and `1800flovvers.com` were registered to `domainadmin@1800flowers.com`. When searching for defensive registrations, we included the domains that were already registered at the start of our .com zone file collection in 2009.

In total, we found 140 VIDNs that have been defensively registered by 136 companies using the same contact email address as for their normal business. In other words, just 0.7% of the VIDNs for which we have historical WHOIS information appear to have been defensively registered directly by the impersonated company. This approach undercounts defensive registrations that are outsourced to third parties or registered using different email addresses. Nonetheless, it still suggests that the vast majority of VIDNs are registered and controlled by people other than the company they are impersonating.

In another respect, the tally of 140 VIDNs may overstate the number of defensive registrations. When we performed this matching, we found 67 company domain names with VIDNs defensively registered using public email addresses that matched the company's registrant email. Four of these company domain names had multiple VIDNs registered. Hence, we are very confident that these domains were in fact defensively registered.

In addition to the company domain names that had public email addresses associated with them, we also found 73 company domain names that had matching emails to associated VIDNs but used private email addresses, such as `noreply@data-protected.net` and `contact@myprivateregistration.com`. For the private registrations, we also matched when the email field was not a valid email address but instead indicated a private registration, such as `REDACTED FOR PRIVACY`. While there is a chance that the hidden registrant emails are different, we conservatively attribute these as defensive registrations, since the pattern of privacy contact information matches exactly.

C. The relationship between registrars and name servers

By default, most registrars will assign their own name servers to domain names when they are initially acquired by a customer. Our data shows that most VIDNs do not change the assigned name server. The historical WHOIS data allowed us to compare registrar and name server information. Ultimately, we found that our results were not substantially changed whether we focused on registrars or name servers.

	1and1-dns	cloudflare	dnspod	domain	domaincontrol	googledomains	hichina	iidns	ipage	microsoftonline	name	name-services	one	orderbox-dns	register	registrar-servers	rookdns	rzone	systemdns	thecloudwebsiteserver	ui-dns	vpweb	wixdns	worldnic	yahoo
1&1 Internet SE	547	1	1	0	0	4	0	0	0	0	1	0	0	0	0	0	0	0	3	0	438	0	0	0	0
Ascio Technologies, Inc	0	0	0	0	0	0	0	0	0	0	0	133	0	0	0	0	0	0	0	0	0	0	0	0	0
Cronon AG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	108	0	0	0	0	0	0	0
Domain.com, LLC	0	0	1	192	2	1	0	0	131	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
eName Technology Co.,Ltd.	0	0	1	0	0	0	0	126	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENOM, INC.	0	3	2	0	0	1	0	0	0	0	0	246	0	0	0	123	0	0	0	0	0	1	0	0	0
FastDomain Inc.	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GoDaddy.com, LLC	0	3	15	0	1 202	0	0	0	0	0	0	0	0	0	0	1	82	0	0	0	0	0	1	0	0
Google Inc.	0	0	0	0	0	1 636	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
HiChina Zhicheng Technology Ltd.	0	0	4	0	0	0	252	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Melbourne IT Ltd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	430
Name.com, Inc.	0	1	0	0	0	0	0	0	0	0	362	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAMECHEAP INC	0	3	0	0	0	0	0	0	0	2	0	0	0	0	0	223	1	0	0	0	0	0	0	0	0
NameSilo, LLC	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NETWORK SOLUTIONS, LLC.	0	0	8	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0	0	0	0	0	160	143	0
PDR Ltd. d/b/a PublicDomainRegistry.com	0	2	9	0	0	0	0	0	0	0	0	0	269	0	0	21	0	0	133	0	0	0	0	0	0
Register.com, Inc.	0	0	2	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	2	1	0	0
TUCOWS, INC.	0	964	1	0	3	3	0	0	2	1	0	1	1	0	0	0	1	0	110	0	0	4 856	0	1	1
Wild West Domains, LLC	0	0	0	0	48	0	0	0	0	519	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XIN NET TECHNOLOGY CORPORATION	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE VIII
COMPARING NAME SERVERS USED BY VIDNS TO THEIR ASSOCIATED REGISTRARS.

Table VIII shows the number of VIDNs assigned to the top 20 registrars and top 25 name servers. The table is sparsely populated, which suggests that name server is in fact a reasonable substitute for registrar. This is noteworthy, particularly since obtaining bulk WHOIS has become harder for cybercrime investigators.

Digging just a bit deeper, we can see by inspecting the columns that for the vast majority of VIDNs, most name servers were associated with a single registrar. When we study registrars by looking across rows, most activity is concentrated at a handful of name servers. For example, VIDNs served by registrar 1&1 Internet SE used `1and1-dns` and `ui-dns` for name servers. VIDNs registered with Tucows used `cloudflare`, `systemdns`, and `vpweb`. From the perspective of countering cybercrime, this suggests that takedown could focus at either the registrars or the service providers who are the registrars' customers.

V. RELATED WORK

Although there is a widespread perception that phishing (the collection of credentials by means of fake websites) uses VIDNs, this is far from the case. In 2006, McFedries, in an article mainly concerned with the etymology of the phishing jargon mentions replacing 'L' by '1' and 'O' with zero, which he calls "homograph spoofing" [11]. However, the Anti-Phishing Working Group has been publishing summaries of phishing activity since January 2004 and even in the earliest days the main attack vector was so-called 'cousin' domains (such as `bankname-usa.com`) [12]. So although VIDNs clearly were used they were not especially prevalent. More recent work has found that cousin domains have continued to be registered [5].

Krammer, also in 2006, discusses a wide range of URL obfuscation techniques applicable to phishing, including what he calls "single-script spoofing", where he mentions 'O' and zero, 'RN' and 'M' and 'L' and 'T' [1]. We do not consider the L/T option in this paper. He also discusses a range of attacks using non-ASCII (IDN) characters, whilst noting that none had been reported thus far.

Gabrilovich and Gontmakher discussed what they called "homograph attacks" in 2002,[3] explaining how Cyrillic and Greek letters that are identical (in most fonts) to ASCII characters could be used to create VIDNs [13]. They give an actual example, mimicking `microsoft.com`, but almost all registries (including .com) now rule out the mixing of character sets that this requires. In 2019 Quinkert et al. found around 3 000 homograph domains where Unicode glyphs had replaced ASCII characters [2]. They don't provide numbers, but it is clear that a lot of examples they found involve accented characters which the registries do allow to be mixed in with normal ASCII.

A completely different type of malicious domain name registration is typosquatting, where domains are registered in the hope that 'fat fingered' typists will visit a website. In 2003 Edelmann documented how a particular actor had registered 8 800 domains providing sexual content on typosquatted versions of well-known domains [14]. In 2006 Banerjee et al. measured the overall prevelance of the issue [15] and in 2010 Moore and Edelmann showed that the choice of domains to typosquat was not to do with the typing difficulty, but the value of the adverts that could be served from the landing pages [16]. In 2017 Szurdi and Christin investigated email typosquatting, finding that if they registered typosquatting domains they would receive a small amount of misdirected email, however they concluded that this was not actually being used for attacks at that time [17].

[3]The use of the term homograph is potentially confusing. Some of the literature carefully uses the term to mean single characters that are 'homographs' of each other, but other papers expand the term to the whole domain which is said to be a homograph of another (VIDN) domain. Other authors use homoglyph for identical looking characters, but this word has yet to make it into the Oxford English Dictionary. The confusion arises because homograph is a well-known philological term for a word with the same spelling as another but a different origin and meaning (e.g. minute – 60 seconds, or very small). In the case of a homograph attack using a VIDN the whole point is that the spelling is different!

In 2011 Dinaburg described 'bitsquatting' where hardware errors cause bits to 'flip' between 1 and 0 within domain names [18]. His Black Hat talk clearly caught the attention of criminals because in 2013 Nikiforakis et al. showed that this had caused a spike in relevant domain registrations [19]. None of the transliterations we consider could be caused by bit flips.

The previous work we have described so far is generally concerned with attacks against a relatively small number of targets (banks, mailbox providers, cryptocurrency exchanges and major brands) although Szurdi et al. did find typosquatting attacks against a 'long tail' of far less important domains [20]. Their definition of typosquatting considers all single character changes so it includes bitsquatting and also some of our VIDN generation methods (though they discuss neither). Their study period (Oct 2012 to Feb 2014) is when we see VIDNs just starting to rise in popularity – so our results will go some way to explaining theirs.

As we have explained, we link the use of VIDNs to Business Email Compromise (BEC) attacks and particularly to scams in which invoice payments are redirected to the criminal following an email correspondence where the victim sees a VIDN and believes that the interaction is genuine. Cross and Gillett provide a survey of the literature on all types of BEC fraud and highlight many gaps in our understanding of why it works so well and how to counter it [21]. Loss figures for BEC that have been reported to them are collated by the FBI – but of course this is only a subset of actual losses worldwide [3].

VI. Ethical considerations

This paper discusses the detail of criminal activity and so there is the need to make the usual ethical decisions as to whether the benefit of explaining how the activity is performed outweighs the risks of informing a new generation of criminals how to commit crimes [22]. We believe that the attackers are already pretty well informed, whereas helping defenders understand what is going on is of real value.

We have chosen to provide numerous examples of actual domain names, which we feel is important in order to clarify what we are talking about and to allow others the opportunity to reproduce and expand upon our work. We have also chosen to provide a large number of email addresses for the registrants of malicious domains. It might be thought that there were data protection issues in doing this, but it is entirely clear from the way in which these email addresses are used that the criminals were aware that WHOIS records are publicly available (and that Law Enforcement would use this data as a starting point in any investigation). Hence we believe that we are only documenting pseudonyms (or noms de guerre) and that the criminals have done their best to ensure that we are not identifying anyone.

VII. Conclusions

We have identified the domain names within the .com registry that are in use by medium and large-size companies. We have then looked for registrations of visually impersonating domain names (VIDNs) that are hard to distinguish from the real domain names, but will not be flagged as containing suspicious characters. We find clear evidence that these registrations are malicious and we can track a rise and fall in the incidence of these domains. In particular we see how there are concentrations of activity at particular domain name registrars over time.

The rise in the number of registrations corresponds with the initial growth of Business Email Compromise (BEC) fraud – which accords with our understanding that this fraud is what the domains are used for. However, BEC fraud has continued to grow whereas the particular types of VIDN that we consider have reduced in number, to a thousand or so a year. At present we do not have a good explanation for this – it may be that other types of VIDN are now in use, or it may just be that the criminals have moved away from specially purchased domain names and are using other methods to fool people as to who they are corresponding with.

We do find a handful of VIDNs that have been defensively registered, but this does not appear to be a particularly widespread practice, This may be because companies have not considered the benefits, or it may just be that with the combinatorial explosion of possible VIDNs for longer domain names it is just too expensive to register them all.

Our research is of course limited by only considering .com domain names (whereas internationally many companies use country code top level domains), and we did not consider whether there were attacks against small businesses. Nevertheless, we believe we have done enough to show that although VIDNs were of considerable importance in 2015–2016, they remain a threat today.

Acknowledgements

Moore and Simpson are supported by US National Science Foundation Award No.1652610. Clayton is supported by the EPSRC [grant number EP/M020320/1]. We gratefully acknowledge data contributions from the Cambridge Cybercrime Centre, John Conwell from DomainTools, and Frank Nagle from Harvard Business School.

References

[1] V. Krammer, "Phishing defense against IDN address spoofing attacks," in *Proceedings of the 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services*, ser. PST '06. New York, NY, USA: Association for Computing Machinery, 2006. [Online]. Available: https://doi.org/10.1145/1501434.1501473

[2] F. Quinkert, T. Lauinger, W. Robertson, E. Kirda, and T. Holz, "It's not what it looks like: Measuring attacks and defensive registrations of homograph domains," in *2019 IEEE Conference on Communications and Network Security (CNS)*, 2019, pp. 259–267.

[3] Federal Bureau of Investigation, "2019 Internet Crime Report," 2020, https://pdf.ic3.gov/2019_IC3Report.pdf.

[4] E. Gabrilovich and A. Gontmakher, "The homograph attack," *Commun. ACM*, vol. 45, no. 2, p. 128, Feb. 2002. [Online]. Available: https://doi.org/10.1145/503124.503156

[5] K. Tian, S. T. Jan, H. Hu, D. Yao, and G. Wang, "Needle in a haystack: Tracking down elite phishing domains in the wild," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 429–442.

[6] Bureau van Dijk, "Orbis — Company information across the globe — BVD," http://orbis.bvdinfo.com/.

[7] S. Ellis, "Business email compromise scams on the rise," 2015. [Online]. Available: https://www.markmonitor.com/mmblog/brand-protection/business-email-compromise-scams-on-the-rise/

[8] R. Broida, "Get a free domain name and Web hosting for one year," 2012, https://www.pcworld.com/article/2010520/get-a-free-domain-name-and-web-hosting-for-one-year.html.

[9] R. Clayton, T. Moore, and N. Christin, "Concentrating correctly on cybercrime concentration," in *14th Workshop on the Economics of Information Security*, 2015. [Online]. Available: http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_clayton.pdf

[10] R. Böhme and T. Moore, "The "iterated weakest link" model of adaptive security investment," *Journal of Information Security*, vol. 7, no. 2, pp. 81–102, 2016. [Online]. Available: https://tylermoore.utulsa.edu/jis16.pdf

[11] P. McFedries, "Technically speaking: Gone phishin'," *IEEE Spectrum*, vol. 43, no. 4, pp. 80–80, 2006.

[12] Anti-Phishing Working Group, "Phishing Activity Trends Reports." [Online]. Available: https://apwg.org/trendsreports/

[13] E. Gabrilovich and A. Gontmakher, "The homograph attack," *Commun. ACM*, vol. 45, no. 2, p. 128, Feb. 2002.

[14] B. Edelman, "Large-scale registration of domains with typographical errors," 2003. [Online]. Available: https://cyber.harvard.edu/archived_content/people/edelman/typo-domains

[15] A. Banerjee, D. Barman, M. Faloutsos, and L. N. Bhuyan, "Cyber-fraud is one typo away," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 2008, pp. 1939–1947.

[16] T. Moore and B. Edelman, "Measuring the perpetrators and funders of typosquatting," in *Financial Cryptography and Data Security*, ser. Lecture Notes in Computer Science, R. Sion, Ed., vol. 6052. Springer, 2010, pp. 175–191. [Online]. Available: https://tylermoore.utulsa.edu/fc10typo.pdf

[17] J. Szurdi and N. Christin, "Email typosquatting," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 419–431.

[18] A. Dinaburg, "Bitsquatting: DNS hijacking without exploitation." 2011, BlackHat Security.

[19] N. Nikiforakis, S. Van Acker, W. Meert, L. Desmet, F. Piessens, and W. Joosen, "Bitsquatting: Exploiting bit-flips for fun, or profit?" in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 989–998.

[20] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The long "taile" of typosquatting domain names," in *23rd USENIX Security Symposium)*, 2014, pp. 191–206.

[21] C. Cross and R. Gillett, "Exploiting trust for financial gain: An overview of business email compromise BEC fraud," *Journal of Financial Crime*, pp. 1–14, April 2020.

[22] T. Moore and R. Clayton, "Ethical dilemmas in take-down research," in *Financial Cryptography and Data Security*, G. Danezis, S. Dietrich, and K. Sako, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 154–168.