Domain Registration Policy Strategies and the Fight against Online Crime

Janos Szurdi
Carnegie Mellon University
jszurdi@andrew.cmu.edu

Nicolas Christin
Carnegie Mellon University
nicolasc@andrew.cmu.edu

ABSTRACT

Domain names play an important role in nearly all online crime. Criminals commonly use domain names to evade blacklisting or to confuse users. The most common approaches to combat abusive registrations are proactive detection of the criminal activity, and reactive blacklisting of domain names. To complement these methods, we study the effects of policy intervention on the domain registration ecosystem with a focus on malicious domain names.

Building on our understanding of the domain registration ecosystem, we develop a multi-stage analysis framework for registration policy proposals. As part of our framework, we discuss the biggest challenges to registration policy deployment (e.g., the complexity of the international domain registration ecosystem); when domain registration can or cannot affect online crime; and the inherent limitations of such analyses. We hope to stimulate further policy work and broaden the discussion beyond technical measures to impede online criminal activity.

Our most promising registration policy proposal comes from the observation that online criminals need far more domain names to operate effectively than benign registrants. We propose a dynamic pricing function and stricter identity verification to make bulk domain registrations expensive. Our game-theoretical analysis indicates that this proposal should have a minimal effect on benign registrants and registries while having a significant financial and operational impact on certain criminal activities. Most interestingly, we observe a synergy between blacklisting and domain registration policies, where increasing blacklisting performance disproportionately boosts policy effectiveness.

1 INTRODUCTION

The Internet depends on the Domain Name System (DNS) to resolve names humans can remember to Internet Protocol (IP) addresses understood by computers. While DNS is also used for a few secondary reasons such as load-balancing and geo-targeting, its main purpose has remained to help humans to find websites (e.g., HTTP, HTTPS), communicate with other humans (e.g. SMTP, POP3, IMAP), or to find other services (e.g. FTP, SSH, Gaming servers). Some domain names became extremely valuable brands and sell for millions of dollars [1].

The value and importance of domain names brought with them a wide range of abuse aimed to profit from them. Domain squatters [2], typosquatters [3], combosquatters [4], and soundsquatters [5] hope to profit from their domain names' similarity to a brand name by passively counting on users' mistakes (e.g. typing mistakes) or by actively fooling users (phishing). Phishing and scams frequently use domain names designed to add a veneer of legitimacy. Spammers use domain names to evade blacklisting of their sender email domains or the domain names in the advertised URL. Drive-by-downloads, botnet operators, illegal content distribution sites and many other online criminals need a large number of domain names to evade blacklisting.

Existing efforts have focused on retroactively blacklisting domain names, after evidence of abuse had surfaced, or proactively detecting criminal activity, for instance, by banning domain names known to be automatically generated by bots.

In this paper, we look at the problem from a slightly different angle: can we design registration policies that make it harder for criminals to register domain names in the first place, without impeding benign registrants? Our objective is to improve existing defenses by making domain ownership more transparent, abusive domain registrations more expensive, and raising the operational risk of registering domain names at-scale for abuse.

In other words, we attempt to complement existing technical work on domain abuse detection and remedial with an exploration of the impact of domain registration policies.

Developing and analyzing an anti-abuse registration policy is challenging. First, we must consider the effects of such a proposal at least on benign users, registrars, registries and ICANN. Second, DNS is a global system deployed across political borders, thereby straddling potentially very different notions of "abuse" or "illegality."

Our contributions include:

- We summarize how domain names are used for different types of online crime, how recent research tackles abusive registrations and whether criminals have a distinctive domain registration pattern that could be leveraged to combat them via domain registration policies (Section 2).
- We design a framework to evaluate domain registration policies (Section 3).
- We discuss the potential benefits, drawbacks, and challenges of multiple registration policy proposals (Section 3.3).
- Using our framework and a game-theoretical model, we evaluate one of the most promising proposals to assess its potential effectiveness against online crime (Section 4).

2 BACKGROUND

We provide here the necessary background for the rest of our paper, by surveying the entire area. We start with a quick overview of the domain registration ecosystem, examine the relationship between online criminal activities and domain name registration. We then turn to a discussion of the "WHOIS debate," which is germane to the problem at hand; last, we discuss recent proposals for domain reputation.

2.1 The domain registration ecosystem

As the Internet grew from a few hosts to millions of domains, the Domain Name System, in charge of mapping IP addresses to human-memorable strings, evolved from a simple translation file ("HOSTS.TXT," back in the days of the ARPANET) to one of the largest, if not the largest, hierarchical distributed systems in existence. Internet domain names have become so important that they are frequently interchangeable with brands, and it is not uncommon for valuable domain names to be resold for millions of dollars [1].

Figure 1 depicts a simplified view of the most important entities in the domain registration ecosystem. ICANN was created to manage the Internet's numerical addresses and domain names. Individual top-level domains (TLDs) are operated by *registries*. There are two kinds of TLDs: generic TLDs (gTLDs) and country-code TLDs (ccTLDs). Registries wishing to operate gTLDs need to be approved and follow ICANN's policies. As an example in Figure 1, Radix has an agreement with ICANN to operate gTLDs such as `.fun` and `.space`. [6]. On the other hand, registries operating ccTLDs have varying levels of cooperation with ICANN: agreements are handled on a purely voluntary basis. For example, the Hungarian registry ISZT has an agreement with ICANN about the `.hu` ccTLD, but the Chinese registry CNNIC has no such agreement. Furthermore, some registries (such as Verisign) can operate multiple gTLDs and ccTLDs, where they need agreements with ICANN and multiple countries at the same time.

Registrars are the entities selling domain names to *registrants* (users registering domain names). *Registered domains* are the part of fully qualified domain names (FQDNs) that registrants can buy.

Besides the myriad registrants with whom registrars have agreements, registrars usually have an agreement with registries to be able to sell their domain names. To directly access gTLDs, registrars need to be accredited by ICANN. Some domain resellers, usually hosting companies, further act as middlemen, selling domains to users, and buying them from registrars.

The purple arrows in Figure 1 depict how money is distributed when a user acquires a domain name. For example, when a user buys `example.com` at reseller *A*, part of the payment is divided between reseller *A*, 1&1, Verisign and ICANN. If another user buys `example.cc` at Godaddy, then GoDaddy, Verisign and the Cocos Island government all profit from this transaction.

2.2 A survey of abusive domain registrations

Besides benign registrants, the rise in popularity of the Internet unfortunately attracted domain speculators and miscreants trying to profit from the relative ease of registering domains. Speculators buy domain names for cheap, in hope to profit from users accidentally visiting their sites, or hoping that they can resell some of their domains for a large profit margin. While domain speculation is an unintended byproduct of Internet domain registration policies, it remains legal as long as speculators are not infringing on existing trademarks or supporting criminal activities.

Understanding differences in the registration patterns and behavior between malicious and benign users is important to design a policy which affect the former but not the latter. Table 1 lists the major categories of online frauds, and summarizes how domain registration plays in the furtherance of each fraudulent activity.

Table 1: Malicious domain name registration patterns

	High demand for domains	Distinctive lexical features	Role of domains	Are domains substitutable?
Spamming	yes	no	Evade BL	easy to BL
Generic Phishing / Scams	yes	usually	Evade BL Fool users	easy to BL
Targeted Phishing / Scams	no	usually	Fool users	less effective
Botnets	yes	no	Evade BL	possible
Malvertisement	yes	no	Evade BL	easy to BL
Illegal pharmacies	yes	no	Evade BL	easy to BL
Drive-by-downloads	yes	no	Evade BL	easy to BL
Illegal streaming	yes	no	Evade BL	easy to BL
Squatting variants	no	yes	Siphon trf. Fool users	no

We focus on two main registration patterns that we can leverage. First, miscreants frequently need to register a large amount of domain names to conduct their activities. Second, some domains have distinctive lexical features related to a target domain.

Miscreants use domain names for four main reasons. First, criminals need to evade blacklisting of their domain names and IP addresses, which often leads them to register a large number of domains. Second, they use domains for accounting and business agility (e.g., traffic distribution systems [7]) when offering their services to other miscreants. Third, crooks frequently use domain names to fool users into believing they are representing an official brand or company. Finally, criminals can register specially crafted domain names to siphon traffic from legitimate domains.

Domain squatting, typosquatting and variants. In domain squatting and its variants, profit stems from the domain name itself. Domain squatting (also known as cybersquatting) [2] is the act of registering domain names of brand names in hope to sell them to the brand owners for profit. More notorious domain squatters used to redirect visitors to adult pages to extort money from brand owners [8, 9].

Typosquatters register domain names lexically close to a target domain to profit from users mistyping the target domain name [3]. Soundsquatting domains are domains that sound similar to the target domain [5]. All these squatting techniques are illegal in the U.S., where the Anti-cybersquatting Consumer Protection Act (15 USC x1125(d)) can be used to protect brand owners. Internationally, ICANN provides a Uniform Domain-Name Dispute-Resolution Policy (UDRP) to mediate domain registration disputes.

Typosquatting and combosquatting domains are also often used for phishing and scam attacks [4, 10]. Combosquatting domain names contain the name of a brand to make the fraudulent domain look like a domain owned by this brand (e.g., `famousbrand-security.com`).

Domain squatting differs from the majority of other online criminal activities, since here domain names are the means to an end: Domain squatters can be driven out of business entirely by targeting their domain registrations.

Spamming is defined as unsolicited bulk messaging. The most common form of spamming is email spam, but spammers often

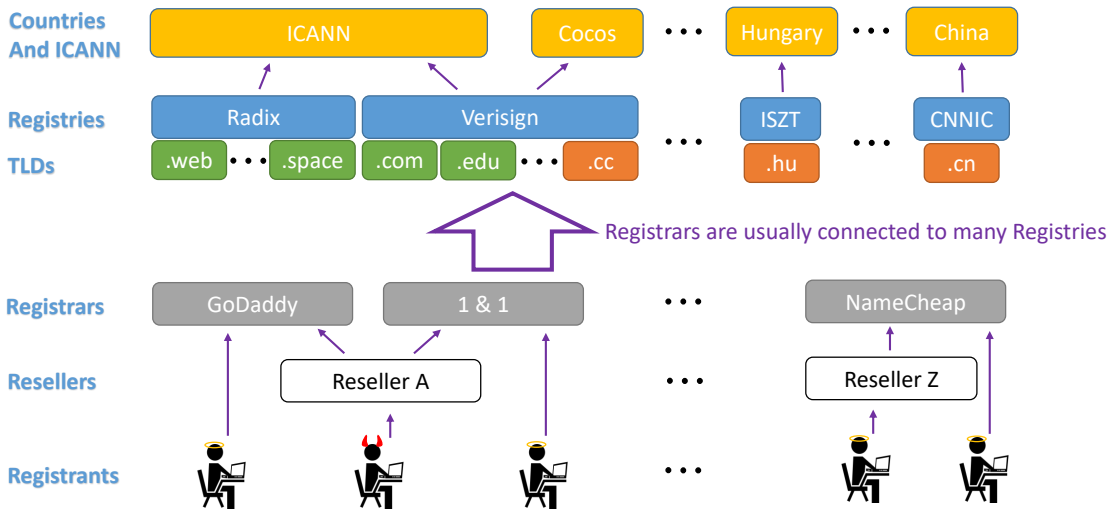


Figure 1: A simplified view of the domain registration ecosystem. gTLDs are in green; ccTLD in orange. Purple arrows denote administrative ownership dependencies—i.e., how money flows from registrants to domain administrators.

target blog comments, tweets, and other messaging systems. Spammers especially need a lot of domain names to evade blacklisting of their email address domains and their spamvertised domains. While spammers could choose to use IP addresses directly instead of domain names, it would raise suspicion leading to blacklisting, since IP addresses are extremely rare as part of URLs in legitimate emails.

Phishing and scamming targeting users. Phishing emails and webpages try to trick users into sharing their personal information with miscreants. Miscreants collect this personal information to sell it to other online criminals who can monetize this information. This personal information includes usernames, passwords, addresses, SSN numbers, identification documents, credit card numbers and other financial information.

Scam operations are very similar to phishing, but instead of tricking users into sharing personal information, scammers try to directly extort money from users.

General phishing and scam attacks try to reach as many users as possible and thus they exhibit similar patterns of domain name usage as spammers to avoid blacklisting.

However, spear phishing attacks and targeted scam attacks use only a couple of carefully selected domain names for a single attack campaign, making registration policies ineffective.

As discussed, typosquatting and combosquatting domains are often used for phishing and scams. Alternatively to these domains, criminals could use an IP or a domain unrelated to the targeted brand name and obfuscate the URL sent in the email or shown in the browser[1]. Using IPs would decrease the success of these attacks just like in the case of spam. Luckily, researchers have created detection systems, such as PhisDef [11], which made URL obfuscation outdated.

Botnets are a collection of infected users' machines controlled by botmasters. Botmasters rent out these machines to be used for a plethora of other illicit online activities.

Botnet operators use techniques called fastflux and doubleflux to hide the location of their command and control centers (C&C). These techniques involve changing the domain names used and changing the NS and A records of these domain names frequently.

Botnet operators are using many other approaches that do not involve domain names to hide their location. These approaches include hard coding IP addresses (often encrypted and obfuscated in binary) or using legitimate cloud service providers' servers to host their C&C. However, these approaches have significant drawback compared to using domain names. If a piece of malware contains hard-coded IPs and is reverse engineered then all samples of the malware can be deactivated. If a piece of malware is using a cloud service provider then either the cloud service provider will be blacklisted after a while or this provider will clean up the malicious activity on their servers. Thus, botnet operators keep enjoying the flexibility and simplicity provided by the domain name system for a low cost.

Malvertisement. Malvertisers post malicious advertisements on benign ad networks to infect, phish or scam users for profit.

Ad network owners such as Google and Facebook continuously try to detect and block malicious advertisements; facing constant blocks, malvertisers thus need a large number of domain names to conceal their activity.

Illegal online pharmacies and other counterfeit stores frequently rely on domain names to provide a veneer of legitimacy to their businesses, making them particulatly vulnerable to blacklisting.

[1]The goal of URL obfuscation is to trick users into believing that they are visiting a known brand's or company's website.

Drive-by-downloads try to infect the victim's browser or computer upon visiting a webpage.

Domains hosting drive-by-download pages are frequently blacklisted (e.g., by Google Safe Browsing and others) as they try to infect users' machines. Drive-by-download pages are also using redirection chains and domain names (Traffic Distribution Systems) to evade blacklisting.

Copyright infringement. When pirated content is shared, online criminals hope to profit from users visiting their website, either through extensive advertisement, or, worse, by infecting user machines or running different scams or phishing schemes [12].

Pages offering pirated content are often blacklisted and taken down. Hence the operators of these pages need domain names to evade blacklisting and are affected similarly to spammers.

In general, if online criminals want user traffic, then they need to either advertise themselves via spamming malvertisement, or malicious search-engine optimization; or siphon traffic via a squatting technique. A common property of these methods is these activities are much easier to block when the bad actors do not rely on domain names, but, e.g., on IP addresses. The only other way for criminals to reach users without domain names is to penetrate a legitimate service's server and carry out the attack on the users of the compromised service.

2.3 The WHOIS debate

The domain registration database (WHOIS) provides an important tool to fight online crime, but the collection of user data also raises privacy concerns. In this section, we summarize how his tension sparked a decade-long debate concerning the WHOIS system and how our paper builds on it.

Brief history of congressional hearings. Since 1998, the U.S. Congress has held more than twenty hearings about ICANN and policies regarding the domain name system [13]. At the first hearing participants discussed the transfer of management of the domain name system to ICANN. Later on, some of these congressional hearings turned into a clash between different stakeholders [14, 15]. On the one hand, the law enforcement and intellectual property communities argued for easier access to WHOIS records, enforcement of accurate WHOIS information and potentially penalizing registrars for allowing malicious registrations. On the other hand, civil right groups would have liked to restrict access to WHOIS information to protect registrants' privacy, to protect political activists, and to protect registrants from spammers and phishing. ICANN's Security and Stability Advisory Committee (SSAC) established in their "Blind men and an elephant" report [16] the need for a better understanding of why WHOIS is needed, what registration information is needed, and who should be able to access certain information.

The first proposed solution. To solve the tension between different stakeholders Operational Point Of Contact (OPOC) was proposed by the ICANN community [17]. The goal of OPOC was to provide a third-party point of contact for registrants and thus shield their personal information from online criminals and provide them a degree of privacy. This proposal achieved a certain balance between privacy and usability. However, the OPOC proposal became quite complex and different stakeholders could not achieve consensus. Therefore ICANN's Expert Working Group decided not to pursue the OPOC solution and instead initiated studies to better understand WHOIS misuse.

The importance of WHOIS. Maintaining accurate WHOIS data is important for several reasons as noted by SSAC [18] and stakeholders [14, 15]. This data is used to pursue violations of intellectual property such as copyright and trademark infringement. Law enforcement agencies frequently use WHOIS to investigate online crime. Security researchers use WHOIS to understand domain ownership and to contact domain owners to clean up compromised websites. Finally, WHOIS can be used by consumers to look up who they are conducting business with on a given domain.

The problems with open WHOIS access. The drawback of free and unlimited access to WHOIS information is that it can be used by spammers and for more elaborate scams or phishing schemes [14, 15]. This was confirmed by Leontiadis and Christin [19], when they found that WHOIS information is leveraged for spamming the registrant's email address, postal addresses, and phone numbers. Furthermore, some registrants might not want to have their personal data available to the public due to privacy considerations; for instance, activists may not want their identities linked to their websites. Inaccuracies can also occur because some registrants mistype their information for the WHOIS database. Finally, malicious registrants do not want to have their real personal data in the WHOIS database to evade law enforcement and legal investigations.

WHOIS privacy and proxy services. All these lead to a significant number of registrants either using WHOIS privacy services or entering fake data as their WHOIS records. Clayton et al. [20] studied in depth the use of WHOIS privacy and proxy services. They found that both benign and malicious registrants often use WHOIS privacy services.[2] In general, registrants that do not use privacy services often cannot be reached via the phone number provided, and, unsurprisingly, malicious registrants can almost never be reached via phone. The Fraudulent Online Identity Sanctions Act (FOISA) was specially created to deter malicious registrants from providing fake WHOIS information [21, 22]. The act doubles the maximum imprisonment if false WHOIS information was provided while committing a felony offense.[3]

ICANN on WHOIS data validation. More recently ICANN's Security and Stability Advisory Committee (SSAC) published a report discussing options for registration data validation [18]. The authors of the document focused on the reasons for WHOIS inaccuracy and the taxonomy of validation. Their taxonomy consists of three levels of validation: syntactic, operational, and identity validations. Syntactic validation refers to making sure the format of the registrant's data is correct. Operational validation means that the contact data provided actually works, for example, emails are received at the provided email address. The goal of identity validation refers to checking if the data provided corresponds to the real world identity of the registrant.

As of 2013, ICANN requires registrars to perform syntactic and operational validation of registrants' data [23, 24]. However even

[2] Malicious registrants use privacy services more often than benign registrants.
[3] At most, FOISA increases maximum imprisonment by seven years.

as of today registration data is often not valid syntactically or operationally [25]. The focus of our research is on identity verification and we assume that syntactic and operational validation is relatively easy and cheap to do well.

ICANN's current proposed solution. Currently, ICANN is working on a new Registration Directory Service (RDS) [25] that would replace WHOIS for new gTLDs. This proposal is still at an early stage where many questions are still under evaluation [26]. What data should be asked from registrants? Who should be able to access what registration data and on what scale? How should different data fields be validated? One proposal under evaluation is to offer partially public and partially gated access (tiered access) to different entities. Another proposal is to use pre-validated identities at registration time maintained by validators.

Connection to our work. Our work has both a different goal and approach compared to the discussion and research around the WHOIS service. Our goal is to systematically find a composition of policy tools that can hurt malicious registrants but not benign registrants. Contrarily, the WHOIS debate is focused on how to provide accurate registration data for security researchers and at the same time provide some privacy guarantees for registrants.

For our proposals, we assume the existence of a registration data service which solves the tensions in the WHOIS debate by providing tiered access and at least operation level validation of data, while in practice this might be challenging to achieve. On the other hand, we explore questions such as how we can provide privacy for sensitive registrants and what are the trade-offs of identity validation. A couple of our proposed policies are closely related to ICANN's new RDS. Related, we discuss the benefits and costs of different identity validation approaches ranging from no identity validation to strict identity validation. More details can be found in section 3.

2.4 Related work on domain reputation, policies, and case studies

Researchers have been working on building domain reputation systems with two goals in mind: 1) to decrease the time it takes to blacklist a domain name and 2) to increase both the precision and recall of these systems.

Antonakakis et al. [27] built one of the first reputation systems for DNS which leverages the characteristics of domain usage specific to online crime. Their system was able to detect malicious usage weeks earlier than traditional blacklists. Recently, Hao et al. [28] showed how registration time features can be leveraged to proactively blacklist domain names further decreasing the time to blacklist domains.

Blacklisting approaches are made harder by the lack of identity verification and the abundance of cheap domain registration options for users. ICANN recently started its new gTLD program to increase the available options to users for domain name registrations. Halvorson et al. [29, 30] found that new gTLDs have a significantly higher rate of speculative and abusive registrations compared to other TLDs. "Taken together, our findings suggest that new gTLDs, while accruing significant revenue for registrars, have yet to provide value to the Internet community in the same way as legacy TLDs"[30]

Liu et al. [31] analyzed the effects of intervention at a single registry, CNNIC in China. They found that it will help to push abuse from that registry's TLD, .cn but it will not affect criminal endeavors in the long-term. Chachra et al. [32] found that 88% of spam domains are blacklisted in less than two days and thus their revenue is effectively limited. However, blacklisted spam domains continue to monetize because of the high demand for advertised goods, non-universal blacklisting, and delay in deployment. Their economic analysis has shown that the per-domain cost would need to be at least a $100 to make these domain registrations unprofitable. At the same time if domains were to be shut down totally instead of blacklisted less than $3 per-domain cost would be sufficient to deter these registrations. Korczynski et al. [33] studied metrics to characterize abuse at TLDs, they found that the size of the TLD and pricing are positively correlated with abuse and DNSSEC deployment is negatively correlated with abuse. Additionally, they found that TLDs with restricted registration policies are less frequently used for phishing.

Research so far studied how to blacklist domain names more effectively, what affects abuse in TLDs or studied the effects of a couple of registration policy intervention attempts that occurred in the past. Our work is different in that we systematically study how multiple potential registration policy strategies would affect the most important entities in the domain registration ecosystem. By doing this we hope to pinpoint directions that are worthwhile to further explore in the grand battle against online criminals.

3 REGISTRATION POLICY EVALUATION FRAMEWORK

The goal of our policy evaluation framework is to find potentially interesting and viable proposals for further consideration from a large set of policies. Our framework involves a multi-step process towards selecting policies to fight online crime. First, in section 3.1, we compile a set of important considerations for future domain registration policies to be evaluated. Second, in section 3.3, we systematically select and evaluate high-level policy ideas to find the ones that are likely useful against online crime and plausible to be implemented by the community. Third, it needs to be more precisely evaluated how each policy would affect different entities in the eco-system. In section 4, we built a game theoretical model evaluating the effects of one of our promising policies. Finally, if all the previous steps indicate that a policy could be useful then its real-life implementation should be designed and evaluated. This final stage is not in the scope of our paper, because it needs multiple stakeholders to work on it together.

3.1 Policy considerations

The domain name registration ecosystem includes a vast number of entities with complex interactions and connections. In this section, we outline the minimum set of entities one must consider when designing a registration policy.

At the bare minimum, a policy proposal should discuss the effects on the entities we discussed in section 2.1: registrants, registrars, registries, and ICANN.

ICANN. An overwhelming part of ICANN's revenue originates from gTLD domain sales, gTLD applications and maintenance fees

Table 2: Summary of domain name usage

# of domains (order of mag.)	Ref.	Abuse
1,000,000	[34],[35]	Spam
100,000	[34],[35],[36]	Malware
100,000	[34],[35],[36]	Phishing
10,000	[34]	Botnet C&C
1,000,000	[3]	Typosquatting
100,000	[4]	Combosquatting

Table 3: Cost of online crime

Abuse	Ref.	Income magnitude (USD)
Online banking:		
- phishing	[37],[36]	100,000,000
- malware (customer)	[37],[38],[36]	10,000,000
- malware (business)	[37],[36]	100,000,000
Fake antivirus	[37],[38],[10]	10,000,000
Copyright infringment	[37]	1-10,000,000
Illegal Pharmacies	[37],[38]	10-100,000,000
Scams (other than banking)	[37],[38],[39]	10-100,000,000
Spamvertisement	[38]	10,000,000
Click fraud	[38]	10,000,000
Botnet PPI	[36]+[40]	1,000,000

[41]. Consequently, a policy intervention leading to a significant drop in the number of domain registrations or gTLDs operated would adversely impact ICANN, the main governing body of the domain registration ecosystem. At the same time, one of ICANN's goal is to ensure a secure operation of domain name registrations. "The mission of the Internet Corporation for Assigned Names and Numbers ("ICANN") is to ensure the stable and secure operation of the Internet's unique identifier systems as described in this Section 1.1(a) (the "Mission")."[42]

Registries. Registries' sole revenue is the fees from domain registrations. A drop in the number of registrations would obviously impact them negatively. Halvorson et al. [30] found that, at the time of their study, only 10% of new gTLDs were profitable. They estimated, using their most optimistic model, that 10% of new gTLDs would not become profitable even after ten years of operation. Therefore, we need to consider how stricter registration policies might make it even harder to make a TLD profitable. We also need to consider to which extent a specific TLD might contribute to the Internet community at large.

Different registries also have different incentives and rules to adhere to. Registries operating gTLDs remain profit-oriented, but they need to conform to ICANN's policies. Registries of ccTLDs are controlled (or operated, in certain cases) by their government. As such, countries more economically affected by cybercrime might have stronger incentives to adopt stronger defenses. On the other hand, some other governments might not suffer much from online crime, and at the same time may see a significant proportion of their GDP coming from domain registration fees. (Tokelau [43], governing the `.to` domain is one such example.) In short, the *economic* incentives to fight (domain registration) abuse strongly differ from country to country.

Incentives for policy change. Every registry operating a gTLD must follow their agreement with ICANN and therefore ICANN has the power to control their registrations policies. However, ICANN follows a multistakeholder model, where decisions are made based on the inputs of many entities such as governments, registrars, and registries. Countries own ccTLDs thus registries operating these ccTLDs must follow their agreement with the country for the specific ccTLD they operate. In this setting, the ICANN community and different countries have a big weight in deciding which policies will be adopted. Many countries suffer from online criminal activities and therefore they are likely to support policies targeting malicious registrations. As discussed in section 2.3, ICANN is already working on a new registration directory service and so it seems ICANN is also determined to work out some of the current problems with domain name registrations. And while it is possible that ICANN would tolerate some financial loss for social good, it remains unlikely they would support a proposal seriously impacting their revenue.

Registrars. Registrars are responsible for selling domain names to users and therefore registries and ICANN depend on them for their own revenue. This gives registrars an important place in ICANN's multistakeholder model. At the same time, registrars compete for users' business, which limits their profit margin on domain sales. Because of this low profit margin, many registrars use domain registrations as a gateway to increase their customer base and to cross-sell hosting services. For example, GoDaddy offers domains for $0.99, which makes their domain sales unprofitable for the first two years; GoDaddy makes up for the lost revenue by gaining customers for its hosting services. In other words, to be acceptable to registrars, a policy should not result in a decrease in customer volume, which is a more important metric than actual income from domain sales. Additionally, malicious users usually rely on separate hosting infrastructure (compromised hosts, or bulletproof servers, depending on the type of criminal activity taking place), thus a decrease in malicious registrations should only modestly affect honest registrars.

Registrants. Benign registrants value their domain names—be they indicative of a brand, or a mere vanity registration. We can assume that any change to that name, including changing the TLD, would decrease the value of the domain name for them. It is hard to estimate the exact value of a domain name to a user, but it is safe to assume that an increase in price by an order of magnitude would discourage many individual users from registering domain names. At the same time, a more modest increase, e.g. less than doubling the price, would not discourage most users from buying domain names. We discussed how malicious users depend on domain names in Section 2.2: different from benign users, they generally value volume over specific domains (with the exception of the various "squatting" scams).

Sensitive registrants. Many policies proposed to combat miscreants, as a side-effect, could negatively impact registrants' freedom of speech. For instance, "real name policies" used by certain entities such as Facebook, have met with significant community push

back, as they can ostracize entire communities (abuse survivors, for instance).

Fortunately, the problem is not entirely unsolvable, even if we advocate for stronger identification requirements for registrants. First, privacy protection services can shield the identity of a registrant from the general public. This solution is similar to OPOC mentioned in section 2.3 and similar to current privacy services. However, the registrant would still own the domain name and would be responsible for its usage. Additionally, the privacy service would still need to provide data for law-enforcement agencies and security researchers.

Second, sensitive registrants might be able to register domain names at TLDs that are not operating in the jurisdiction of their government. This solution would make it hard or impossible for the registrant's government to associate them with the domain based on registration data.

Third, supporting foreign organizations could offer these users subdomains under their own domain or could even proxy ownership for them. This proposal would shield registrants fearing their own government.

Binding vs. non-binding policies. As we discussed above, the ecosystem is diverse enough that different registries will have different obligations and incentives, thus it is unlikely they would all agree to a common specific registration policy. Consequently, it is beneficial to evaluate three levels of collaboration for each proposed policy: whether only a few registries, most registries, and all registries implement the proposed policy.

In case the proposed policy is non-binding, making abusive domain registrations harder will decrease the abuse at the adopting TLDs, but as observed by Liu et al. [31], malicious registrants will adopt and start registering domains at other TLDs. If, on the other hand, the policy is binding, that is, if ICANN mandates policy implementation, the vast majority of gTLDs will have to collaborate; individual ccTLDs may then be forced to follow suit, as the critical mass of collaborative TLDs would make it easier to blacklist malicious domains registered at shadier, non-collaborative registries.

Hacked domains versus abusive registrations. Often hacked domains and abusive registrations can be used for the same purpose. No matter how successful a domain registration policy is, it will not affect hacked domains used to support online crime. Nevertheless, a successful anti-abuse policy would *force* miscreants to primarily resort to hacked domains—which is more complicated than simply registering a domain. Recent advances in web security (e.g., predictive analytics [44]) may further increase the difficulty of compromising existing domains at scale. In conclusion, we need to tackle both malicious registrations and domain name compromises to solve the general issue with criminals using domains for malicious purposes.

Definition of abuse and illegal across borders. It is important to define the terms "abuse" and "illegal" for domain registrations. We would define a domain name registration to be abusive if it was registered for illegal purposes based on the laws of the country where the TLD's registry resides. For each TLD, the definition of abuse would be different but could have a reasonable common core, which would include illegal activities such as squatting, spamming, scams, phishing, illegal content and goods distribution, botnet operations etc. Building on this common core, registries could take actions against these malicious registrations or could introduce fines or security deposits to make criminal efforts more expensive.

3.2 On the potential of domain registration policies

Based on existing research, Table 2 summarizes the orders of magnitude of blacklisted domains or squatting domains registered every year for each type of abuse. Table 3 shows the estimated order of magnitude of yearly income for different types of online criminality activity.

These estimates must be treated with caution. Criminal income is in particular notoriously difficult to pinpoint and can be either overestimated or underestimated. On the other hand, the number of domains blacklisted is likely underestimated because blacklists try to minimize false positives.

Looking at Table 2, we can observe that abuse yields earn hundreds of millions of dollars in revenue per year, and corresponds to millions of domain names being registered each year. Straightforward averaging would yield a criminal income per domain name to be around a hundred dollars. Clearly, using the average is not suitable because the effectiveness of criminals and domain registration needs are highly variable. For example, spear phishing campaigns or targeted scam attacks may require only a couple of domain names, each bringing in a very high revenue per domain, and therefore making the designing of policy-based countermeasures challenging. On the other hand, a number of abuses require a lot of domain names and are less effective on a per-domain name basis. In Table 3, spamvertisement jumps out as a potentially good candidate to be affected by stricter registration policies. Typosquatting is also a good example, where most domains would become unprofitable if the cost of malicious registrations increased. In general, previous research has – time and again – shown that online crime is a heavy-tailed business, where a few, major, actors account for the vast majority of the ecosystem [45–47]. Thus a successful registration policy proposal could decrease the number of criminals by further pushing out the less successful ones into bankruptcy.

3.3 High-level policy proposal discussions

We attempt to systematically build a list of policy proposal based on the tools available for registries and ICANN. These basic tools include domain pricing, level of identity verification, fees, security deposits, incentives for good behavior, lexical prediction and combinations of these policies.

Proposal 1: Small increase in the registration price. For the many criminals with a small profit margins, even a small increase in pricing could be discouraging from registering domain names. The question is what price increase would impact malicious registrants but not benign registrants. Future work should attempt to provide accurate estimates of the price-sensitivity of (benign) registrants, to infer possible tuning knobs for such increases.

Proposal 2: Stricter verification requirement. Currently, the overwhelming majority of registries do not have any identity verification in place allowing criminals to register domain names with as many identities as they want. Stricter identity verification would

Table 4: Table evaluating the potential effects of the policy proposals discussed.

	Malicious registrants (One reg. adopts)	Malicious registrants (Most adopt)	Benign registrants	Registrars, registries, ICANN	Sensitive registrants	Adoption probability
Small Price Increase	local	yes	maybe	small	no	possible
Strict Identity Verification	local	yes	small	yes	yes	possible
Fines or Security Deposits	no	no	small	small	no	unlikely
Anti Bulk Registration	local	yes	small	yes	yes	possible
Large Price Increase	no	yes	yes	yes	no	unlikely
Protocol Separation	no	yes	small	yes	no	unlikely
Incentivizing Registries	local	yes	no	yes	no	possible
Anti-squatting	yes	yes	no	yes	no	possible

require miscreants to use high-quality fake or stolen identities, imposing an additional cost on them. The operational risk of criminals would also increase – as procuring (a large number of) stolen identities in itself is a potentially risky endeavor.[4]

Examining further the effects of different identity verification schemes, the most important attributes to look at are their evadability, cost, and accessibility. On the one hand, completely forgoing verification is cheap, accessible, but also easy to evade—the attacker does not need to take any specific precautions to do so. On the other hand, in-person verification is expensive and has limited accessibility, but it is also expensive for an attacker to defeat. SSL-extended validation style of verification is hard to defeat, but it would negatively impact most regular users, as it is both expensive and lacks accessibility.

To find the balance between cost, accessibility, and evadability, one suggestion is to use a combination of identification documents, which are hard to find on black markets, with automated face recognition and liveness detection. Such a system could be affordable and accessible for benign users, but expensive to evade. Matching credit cards and identity documents are scarce on online black markets and are more expensive than requiring non-matching documents. Adding phone number and email verification (potentially from a big email service provider) can also raise the cost of Sybil attacks. While state-of-the-art liveness detection and face recognition can be evaded [49], evasion requires higher technical skills and more investment (per identity) from the attackers. One of the biggest online identity verification provider informed us that they sell their product for $0.5-2 per identification. Our suggestion is very similar to their automated solution and also takes black market pricing (using data from [50]) into account. In other words, this approach could have a low cost, be accessible and would be potentially expensive for criminals.

Standardized registration policy and increased strictness of identity verification would also allow for better defenses detecting Sybil attacks. For example, an IP reputation system could be used to make Sybil attacks harder by tracking the number and kind of registrations from IP addresses. Luckily a lot of work has been done in this space led by tech companies such as Google, Facebook, and Jumio.

Proposal 3: Fines and security deposits. Fines are traditionally used to incent citizens to remain law-abiding. Conversely, criminals already hide their identity or operate in jurisdictions different from where they are located, making enforcement mechanisms such as fines hard to deploy. While security deposits could be useful against criminals, it might dissuade regular users from registering domains. However, fines can indirectly affect malicious registrants, by making "outsourcing" less appealing. Specifically, fines could disincent otherwise law-abiding people from registering domain names with their own identity on behalf of criminals. Finally, security deposits could be used in case of suspicious domain registrations such as typosquatting or a sudden large amount of registration attempts from a developing country.

Proposal 4: Anti-bulk registration policy. The anti-bulk registration proposal builds on the observation that spammers, botnet operators, typosquatters and many other online criminals are banking on the fact that they can access a large number of domain names cheaply to avoid reputation systems and blacklists. By making bulk registrations hard and expensive we target the abundance of cheap domain names for online criminals. Additionally, for most TLDs, the identity of registrants are not validated leading to a lack of transparency in the ownership of domain names.

The policy changes we propose are strict verification of identity at registration time, increasing domain price with the number of domains registered and optionally a security fine/deposit to thwart malicious behavior. Strict identity verification is important to make Sybil attacks expensive and to increase transparency. Increasing domain name price as the function of domain names owned is crucial to make bulk registrations expensive and at the same time allow users to own a few domains for an inexpensive price. This policy proposal leverages the benefits of several previously discussed policy options, while it minimizes their drawbacks.

There are only a handful of legitimate reasons for a registrant to own more than a couple of domains. Domain name speculators buy large quantities of domain names in hope to sell them later for profit or earn money from incoming traffic (e.g. type-in navigation). Sometimes these domains lead to malicious content when domain owners employ more lucrative but more questionable parking services [51, 52]. As explained before, the goal of the domain name system is to give memorable names to resources on the Internet for users. Speculative domain registrations are a parasitic byproduct and are not serving the primary goal of the domain name system. A better example of benign registrations is defensive registrations. Users defensively register many variants of their brand name to protect it from domain squatting, typosquatting and other variants

[4] Ross Ulbricht, the creator of the Silk Road website [48], actually had an initial encounter with the police, when ordering a bunch of fake driver's licenses from his own marketplace.

of name squatting. A simple algorithm could decide if a registration is defensive or not and thus a registrant could register these domains on the base price. Finally, hosting providers also often own their customers' domain names, this could be resolved by proxy ownership, where both owners are responsible for potential misuse of the domain name.

Proposal 5: Considerable increase in the registration price. This proposal's aims at making domain names less desirable for miscreants indirectly. We plan on achieving this by making the hierarchy of ownership in the domain name system deeper. Currently, the hierarchy is only two-level deep: TLDs and people registering domains under these TLDs (most often second level domains). Even though now nearly two thousand TLDs exist, only a handful of them is actually used by most Internet users. This means that domain name reputation systems basically have to work only with registered (mostly second level) domains. The proposal is to use pricing to motivate the usage of lower level domain names for domain ownership and have a different use for the different level of domain names.

More specifically, the proposal is to make domain names very expensive such that only big companies/brands/organization could afford them. This would force personal websites and small business to lower levels (mostly third level domains). We would call these domains first- and second-tier domains respectively. This proposal would make first-tier domains not economical for malicious usage. Additionally, penalties could be put in place to enforce first-tier domain owners to keep their namespace clean.

To discuss how registrants could cope with this change, consider the example of a florist from Pittsburgh named Jane. Jane would not be able to purchase the domain `janetheflorist.com`, which would be out of her price range. Instead, she could join together with small businesses in Pittsburgh and buy `pgh.com` and then use `janetheflorist.pgh.com`. Registration requirement under `pgh.com` would be strict and would require individuals to own a business in Pittsburgh,[5] and for this reason, abusive second-tier registrations would be cumbersome and rare under `pgh.com`. Free-speech advocacy organizations could buy domains such as `freedom.com` to allow anyone to have a web presence anonymously and cheaply by allowing them to use their namespace, e.g., `mypolitics.freedom.com`. To mitigate abusive second-tier domains, `freedom.com` would probably only allow a limited web presence for its second-tier domains.

The main problem with this proposal would be the transition from the current system. Many people have marketed and built out a brand around their domain names, changing them would be highly undesirable and would cause potentially financial losses to these users. In fact, the current trend in DNS – flattening of the namespace by the introduction of new gTLDs – flies very much in the face of this proposal.

Proposal 6: Combining domain registrations with SSL and protocol separation. Here, the idea is to have different levels of trust in domain names based on the level of identity verification and price paid for the domain name. Based on the level of trust, different application protocols would be accepted for different domains. As an example, email protocols would need a higher level of trust than running webpages with certain restrictions. However, to allow webpages to offer files for download and to allow these files to leave the browser's sandbox would also require a higher security level.

These security levels would aim at directly making domain names too expensive for certain types of cybercrime. For a domain to be used in a certain protocol, it would need to be priced according to the protocol's potential for malicious usage. In addition, this approach could be easily coupled with SSL domain validations. The main problem is that this proposal would need to be adopted by most users of the Internet, allowing them to decline connections from low security level domain names.

Proposal 7: Incentivizing domain registries to fight abuse. As observed by Korczynski et al.[53], we could incentivize domain registries or registrars to decrease abusive registrations based on the actual abuse found at these registries. More specifically we could increase or decrease the per domain fee they pay based on the number of domains blacklisted in their TLDs. Participants would need to agree on the definition of "abusive" and would also need to agree on which entities could decide if a domain was abusive, hence a penalty is necessary. This policy has relatively few drawbacks, making it a promising avenue for further investigation.

Proposal 8: Anti-squatting analysis. Typosquatting, combosquatting, soundsquatting and cybersquatting share that they can be identified based on lexical features with good true positive rate and moderate precision. Therefore proven typosquatting domains could be removed and new registrations for these users could be hardened. In case these registrants cannot present convincing proof of their benign intent then a security deposit could be required from them.

This approach would be highly effective against squatting and would also impact phishing and scam attacks which frequently rely on lexicographically-close domain names to fool victims. This proposal is also non-binding, thus it is effective even if only one registry implements it. Furthermore, it does not negatively impact benign registrants – making it also a promising prospect.

Summary of proposals. Table 4 summarizes the effects of the previously discussed policies. In general, we would like to find policies that are effective against malicious registrants but do not hurt benign registrants. Most policies would impact registrars, registries, and ICANN because they decrease the number of domain registrations. The question is how much they would be impacted and can we counter-balance it somehow? If identity verification is required then sensitive registrants will be impacted, we discussed in Section 3.1 what options they have to mitigate the policy's impact on them. Last, we would like to focus on policies that are not unlikely to be implemented.

It is possible to combine multiple policy proposals to increase their effect on online criminals. For example, implementing the anti-bulk registration proposal does not mean that incentivizing registries and registrars to be more due diligent in banning malicious domain could not be effective. Additionally, adding the anti-squatting policy could help remove high-value domains that were not affected by the previous two proposals.

3.4 Policy proposal implementation challenges

Dependence on blacklisting. Domain registration policy efforts need to be in harmony with current blacklisting efforts. Indeed,

[5] Similar, in that sense, to the policies on certain ccTLDs such as `.fr` or `.us`.

without effective blacklisting, malicious users do not need to exhibit a registration pattern substantially different from benign users. The better the blacklisting efforts, the lower the per-domain revenue of online crime, making our registration policy proposals even more effective. At the same time, some of these registration policy proposals could also make blacklisting more effective. Making domain registrations harder for criminals would corner them into fewer TLDs and would decrease the general noise and opaqueness of the current chaotic situation of domain registrations.

Unaffected domains. Certain malicious domains are not affected by changes in pricing and verification requirement. For example, targeted scams or phishing attacks might only use one typosquatting domain name to confuse the customers of a bank. Abusive domain name registrations that do not show distinctive pattern compared to benign registrations are impossible to affect directly via registration policies. Indirectly registries and registrars could be incentivized to clean their domain more diligently.

Data management problems. The current WHOIS system is often used by researchers and security analyst to learn about the ownership of malicious domain names. As discussed in section 2.3, there are two main problems with the current WHOIS database. First, the data is often inaccurate since identity verification is minimal at most TLDs. Second, it has been shown that WHOIS information (when correct) is sometimes used to deliver email, mail or phone spam [19]. Implementing the anti-bulk registration proposal would allow registries to increase the accuracy of their WHOIS information. It would also allow the elimination of WHOIS spam by using pseudonyms. Using pseudonyms would still allow researchers to tie domain names together owned by the same person or company. Alternatively, a gated access to WHOIS data could be introduced as proposed by ICANN (section 2.3).

4 GAME-THEORETIC ANALYSIS OF THE ANTI-BULK REGISTRATION POLICY PROPOSAL

Based on our high-level analysis, we next analyze the anti-bulk registration policy proposal more in depth. The goal is to capture the benefits of increasing domain registration prices while attempting to minimize the drawbacks incurred to legitimate registrants.

4.1 Formal model

The game we design resembles a Stackelberg game variant. In this game, registries are the "leaders," who decide their strategy of pricing and identity verification first. The registrants are the "followers," who decide where and how many domains they want to register.

Our design is different from a classical Stackelberg game in that we have multiple leaders and followers. We model registries as two leading players. One leader is a group of collaborating registries coordinating their registration policy strategy to combat malicious registrations. The other leader is the group of non-collaborating registries, who are not impacted negatively by malicious registrations. Registries are playing a simultaneous move game. Registrants then respond to strategies selected by registries[6].

Our analysis consists of evaluating the pure strategy Nash equilibria between registries and analyzing what would be the best response of non-collaborating registries and registrants to certain strategies chosen by collaborating registries.

Our proposed model simplifies the domain registration ecosystem by only considering registries and registrants. On the one hand, we consider registries as players because they control both registration policies and pricing for the TLDs operated by them, and thus capture the essential mechanisms in the ecosystem. On the other hand, registrar market is extremely saturated, to the point where registrars often sell domain names at or below cost. Therefore, registrars do not significantly affect the final pricing and registration policy for registrants. ICANN and governments could potentially impact registration policies set by registries. We incorporated them in our model indirectly as a parameter in the utility function of registries. ICANN, registries, and registrars have a voice in the registration ecosystem and for a policy to be implemented it is important for their revenue not to be significantly impacted. We can estimate how they are affected by the decrease in the number of domain registrations and the decrease in the number of registrants. In Section 3.3, we discussed how registrants with special needs would be affected and how could they be handled. As such registrants would not significantly impact the game, we assume them away and do not model them.

Players. Let us define benign registrants as $b \in B$, and malicious registrants as $m \in M$. Registries are $r \in R$. For simplicity (and without loss of generality) we assume that there are only two registry players, i.e., $R = \{r_c, r_{nc}\}$, where r_c is the group of collaborating registries and r_{nc} is the group of non-collaborating registries. With this simplification, we do not need to model the interaction between collaborating registries as part of their strategies and utility functions.

Strategies. A malicious registrant m can decide how many domains $n_{m,r}$ they want to register at registry r and how many fraudulent (i.e., fake or stolen) identities $i_{m,r}$ they want to purchase to use at registry r. The maximum number of domains that a malicious registrant can profit from is $n_m^{\max}$, thus $\sum_{r \in R} n_{m,r} \leq n_m^{\max}$. A benign registrant b can decide how many domains $n_{b,r}$ they want to register at registry r. $i_{b,r} = 1$ for all $b \in B$ because we assume benign registrants have only one identity. Similarly to malicious registrants, the following constraint holds for benign registrants: $\sum_{r \in R} n_{b,r} \leq n_b^{\max}$.

A registry r can define its pricing function $C_r(n, i, \alpha_r, \beta_r)$ by setting the base price α_r and the discount (or penalty) for registering more than one domain β_r. The number of domains to be registered n is divided by the number of identities used i, to represent optimal

[6] We assume registrants are best responding to the strategies of the registries and we leverage this to calculate the registries' utilities

fraudulent identity allocation by malicious users.[7]

$$C_r(n, i, \alpha_r, \beta_r) = \sum_{j=1}^{i} \alpha_r \cdot (\frac{n}{i})^{\beta_r}$$

The registry can also define how hard it wants to make identity verification by defining θ_r, the cost of one verification. θ_r will also define the cost of buying a fraudulent identity λ_{θ_r}.

Utility functions.

The utility function of **malicious registrants** consists of four components: the value V_m is derived from the criminal activity, the cost of registering domain names C_r (the same function as defined above), the cost of fraudulent identities F_m and $\theta_r \cdot i_{m,r}$ the cost of verification.

$$U_m = \sum_{r \in R} \Big[V_m(n_{m,r}, i_{m,r}, \gamma_{m,r}, p_{\text{bl}}) - C_r(n_{m,r}, i_{m,r}, \alpha_r, \beta_r) - F_m(i_{m,r}, \lambda_{\theta_r}) - \theta_r \cdot i_{m,r} \Big]$$

The per-domain income from perpetrating a specific type of criminal activity is represented by $\gamma_{m,r}$. For a specific criminal activity, $n_m^{\max}$ is the maximum number of domains that are useful to register.

Finally, p_{bl} is the probability of an individual domain being blacklisted. Using p_{bl} we calculate the expected number of domains blacklisted given the number of fraudulent identities used i. The formula below models how having domains blacklisted and owning too few identities leads to the blacklisting of other domains registered using the same identities. As we do not know the exact value of p_{bl}, we will evaluate a range of possible values.

$$V_m(n_{m,r}, i_{m,r}, \gamma_{m,r}, p_{\text{bl}}) = \gamma_{m,r} \cdot n_{m,r} \cdot (1 - p_{\text{bl}})^{\frac{n_{m,r}}{i_{m,r}}}$$

The cost of buying a fraudulent identity is λ_{θ_r} and multiplying it by $i_{m,r}$ gives the total cost of fraudulent identities for a malicious registrant. When the value of λ_{θ_r} is unknown, we will test several interesting values.

$$F_m(i_{m,r}, \lambda_{\theta_r}) = i_{m,r} \cdot \lambda_{\theta_r}$$

The utility of a **benign registrant** consists of the value of the domain names V_b, the cost of registering the domain names C_r and $\theta_r \cdot i_{b,r}$ the cost of verification.

$$U_b = \sum_{r \in R} \Big[V_b(n_{b,r}, \gamma_{b,r}) - C_r(n_{b,r}, 1, \alpha_r, \beta_r) - \theta_r \cdot i_{b,r} \Big]$$

The average value of a domain name for a registrant is $\gamma_{b,r}$ and the maximum number of domains a registrant can profit from is $n_b^{\max}$:

$$V_b(n_{b,r}, \gamma_{b,r}) = \gamma_{b,r} \cdot n_{b,r}$$

The utility of **registries** consists of two parts: C_r the fees from domain registrations and the cost of online crime.

$$U_r = \sum_{j \in B \cup M} \Big[C_r(n_{j,r}, i_{j,r}, \alpha_r, \beta_r) \Big] - \rho_r \cdot \sum_{b \in B} \Big[V_m(n_{m,r}, i_{m,r}, \gamma_{m,r}, p_{\text{bl}}) \Big]$$

The only new parameter in this equation is ρ_r representing how important the cost of online crime is for a registry. For registries operating in countries where the cost of online crime is higher than the revenue from domain name sales, ρ_r is high. Example of high ρ_r could be countries with high GDP because these countries are more frequently targeted by online crime. In countries where the domain name fees are higher than the cost of online crime, ρ_r is low. For example, $\rho_r = 0$ for Tokelau's ccTLD .tk, because the domain name fees are a significant part of their GDP [43] while they are not affected by these criminal activities. The cost of online crime ρ_r could be influenced by ICANN for gTLDs.

Parameter estimation, simplification and assumptions.

In this section we discuss the parameters of the model and how we can estimate them or what assumptions we have to make.

$\gamma_{b,r}$ **the value of a domain name for** b**.** For most benign users we assume that they would still buy their domains, if domain prices would rise only a little bit (for example less than doubles). But they would not buy their domain, if the price would increase any more than that. This would make $\gamma_{b,r} \approx 20$ for average users. ($\gamma_{b,r}$ is expressed in dollars/domain.) In future work, we hope to estimate Alexa's top 1 million domains' traffic using the Zipf curve we fitted on Alexa traffic estimates and multiply it by how much Google pays per a thousand impressions.

$n_b^{\max}$ **the maximum number of domains that benefits** b**.** We assume $n_b^{\max} = 1$ for the sake of simplicity. An extension to the model could estimate the domain ownership distribution based on WHOIS data.

$\gamma_{m,r}$ **is the value of online crime and** $n_m^{\max}$ **is the maximum number of domains that benefits** m**.** For malicious registrants $\gamma_{m,r}$ and $n_m^{\max}$ is different for each online crime type. We have estimated these values in Section 4.2 for typosquatting and pay-per-install services. We also model a general online criminal with varying values per domain revenue $\gamma_{m,r}$.

θ_r **is the cost of one identity verification.** While θ_r is the choice of registry r, we might want to simplify our model and consider values based on real-life examples. Most registries do not verify the identities of users, which means $\theta_r \approx 0$. For a large identity verification service to do face recognition combined with liveness detection and document verification means $\theta_r \approx 1$. We conjecture that for the rigorous SSL extended verification $\theta_r \approx 100$ in order of magnitude. Our suggestion of combined document verification would be a modification of existing services' verification systems and it should cost approximately the same, conservatively we estimate $\theta_r \approx 4$.

λ_{θ_r} **is the cost of a fraudulent identity.** When $\theta_r = 0$ then λ_{θ_r} is also zero. However, the value of λ_{θ_r} is questionable if $\theta_r \geq 1$. We estimate its values based on online anonymous marketplace prices. However we use this estimate with caution and we test our model with multiple possible values for λ_{θ_r}.

[7] Our simulation solves the integer version of this problem. For example, $n = 10 \wedge i = 3$ means that two identities will have three domains associated with them and one identity will have four domains associated with it.

ρ_r is the cost of online crime for registries. We discussed ρ_r the cost of online crime for registries earlier. A simplifying but reasonable assumption is $\rho_r = 0.1$ for the collaborative registries and $\rho_r = 0$ for the non-collaborative registries. In case of collaborative registries, we test multiple potential values of ρ_r

p_{bl} is the probability for an individual domain to be blacklisted. We test multiple values of p_{bl}. Setting $p_{\text{bl}} = 0$ means that we are not modeling blacklisting and criminals do not need to worry about it. $p_{\text{bl}} = 1$ means that domains are always blacklisted before crooks can profit from them. A small value for p_{bl} is reasonable because domains are often blacklisted after the online crime was already perpetrated.

The α_r and β_r of the pricing function C_r. To simplify our model we consider only certain values of α_r and β_r, such as $\alpha_r \in \{1, 2, 10, 100, 1000\}$ and $\beta_r \in \{0.95, 0.99, 1, 2, 3\}$. Having a finite number of strategies allows us to compute the game's payoff function based on the registrants' best response.

How registrars, ICANN, and countries are modeled in this game? In this formulation, registrars are represented as part of the registries. If fewer users are registering at a registry or the payoff of the registry decreases due to the strategies chosen by the players, would mean a decrease in the registrar's utility. ICANN and countries are represented in the choice of ρ_r for their TLD and the registrants TLD preferences $\gamma_{b,r} \wedge \gamma_{m,r}$.

4.2 Seeding the model with data

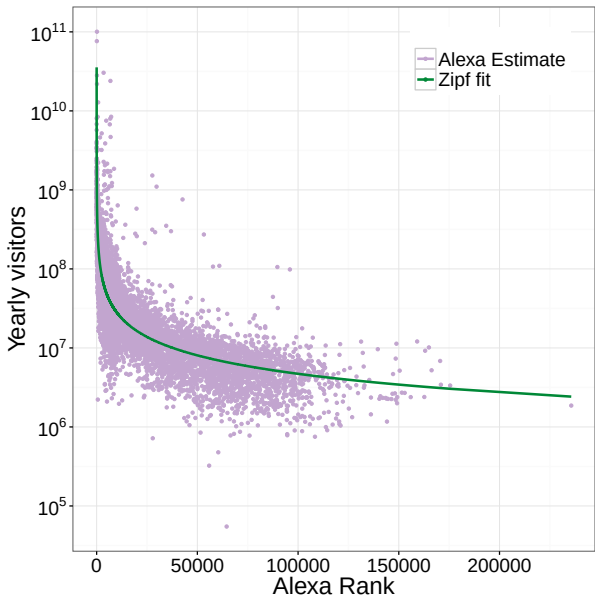


Figure 2: On these plots we can see Alexa's estimate of the yearly visitors at domains as the function of their Alexa rank. The green line represents the Zipf curve we fitted in log space on Alexa's estimate ($R^2 = 0.76$).

Estimating typosquatting domain ownership and revenue. First, we model the number of domains owned by typosquatters based on WHOIS clustering done by Szurdi et al. [54]. It is important to note that this estimate is a lower bound on the number of domains owned by typosquatters because WHOIS data can be easily spoofed, thus one typosquatter might look like many entities in our clusters. We also exclude privacy protected typosquatting domains, which means the probable exclusion of some of the worst typosquatters. As a further precaution, we only consider a registrant to be a typosquatter if she owns at least ten typosquatting domain names.

We estimate the revenue of typosquatters as:

$$\gamma_{m,r} = \text{Traffic}_{orig.} \cdot Rate_{mistype} \cdot CTR \cdot PPC$$

Figure 2 shows our estimate of the number of visitors domain names receive ($\text{Traffic}_{orig.}$) by fitting a Zipf curve on Alexa's estimate of the traffic received by top ranked domains. We use the estimates by Moore et al. [55, 56] directly for PPC and the average percent of $\text{Traffic}_{orig.}$ going mistakenly to typosquatting domains instead of the original domain. We know that a typosquatting domain's quality depends on many factors, therefore we use Szurdi et al.'s [54] observations to estimate $Rate_{mistype}$ for individual typosquatting domains. We use Google's case study [57] to estimate CTR.

For typosquatters, we modeled V_b slightly differently compared to the formula in section 4.1. We took into account that their domains have significantly different values and we assumed they prioritize registering the best of their domain names.

Estimating botnet revenue per domain name. The cost of a thousand unique installs on bots cost from \$7-\$8 to \$100-\$180 [40]. For an upper bound in order of magnitude, we calculate with a cost of \$100 per a thousand bots. We conjecture that a year a machine is sold in this fashion ten times. This leads to our estimate of $R_{bot} =$ \$1/bot/year income for botnet operators. We assume that these bots are solely used for pay-per-click installs.

We also estimate that the time to blacklist domain names is one day on average [28, 33, 58]. This lead to the following function to calculate revenue per domain.

$$\gamma_{m,r} = N_{bots} \cdot R_{bot} \cdot \frac{min(N_{domains}, 365)}{365 \cdot N_{domains}}$$

Representation of malicious registrants. The per domain revenue $\gamma_{m,r}$ can vary by orders of magnitude for different online criminals. In our model, we represent miscreants based on how good or bad their per domain revenue is. We use criminals anywhere in the range $\gamma_{m,r} = (1, 3000)$. Criminals with low per domain efficacy include spamvertisement, small botnet operators, and typosquatters. Examples of decent $\gamma_{m,r}$ are general scam and phishing attackers and better typosquatters or botnet operators. Finally, certain criminals need only a couple of domains with high potential revenue such as spear phishing and banking trojans.

4.3 Analysis

We calculated the Nash Equilibria for a wide range of parameter values, which resulted in many different games and for each game potentially different sets of equilibria. First, we analyzed all these equilibria together to see if we can distill any takeaways that are true for all of them. Second, we evaluated a more precise analysis of specific scenarios. We start with a scenario we believe to be the

most realistic and then we tweak the parameters to see how the results change given different scenarios.

Registry nash equilibria analysis. When we evaluated our model and found the Nash equilibria, we observed that registries set α_r such that they do not lose their customers to other registries. At the same time, they also select the largest α_r where registrants still choose them. Consequently, they primarily use β_r to deter criminals.

Our model does not yield a Nash equilibrium in which the registries can discourage the largest botnet operators from registering domain names. Medium and large botnet operators decide not to register domain names if a combination of high fake identity cost and low utility from registering domain names at non-collaborating registries co-occur. Low utility of registering domains at non-collaborative registries models the situation when most or all registries are actually collaborating and thus non-collaborative registries became isolated.

The probability of blacklisting an individual domain greatly affects our model. Trivially, in the non-realistic case of $p_{\text{bl}} = 1$, no abusive registrations occur. At the other end of the spectrum, if $p_{\text{bl}} = 0$ it becomes harder for registries to do something about malicious registrations, but they are still able to affect certain criminals by increasing prices. When we increase p_{bl} it becomes easier for registries to discourage malicious domain name registrations. This captures the synergy between blacklisting and domain registration policies to combat abusive domain registrations.

Registration policy scenario analysis.

The base scenario. First, we start by making a set of assumptions about the input parameters of the model. While we choose a realistic starting point, we will also analyze the effects of changing these assumptions and the values of these parameters.

In the base scenario we assume that users prefer their current TLDs, but if the pricing increases significantly in their current TLD they are willing to switch to another TLD. This is represented by setting $\frac{\gamma_{b,r_c}}{\gamma_{b,r_{nc}}} = 10$. We assume that collaborative registries care about abusive registrations and non-collaborative registries do not care, leading to $\rho_{r_c} = 0.1$ and $\rho_{r_{nc}} = 0$. We assume that the probability of blacklisting domains is not zero, but it is low $p_{\text{bl}} = 0.01$. Finally, we assumed that benign registrants approximately register a hundred times more domains than malicious registrants. This assumption is reasonable because each year there are millions of domains registered for abusive purposes and there are hundreds of millions of domains registered by benign users.

With these assumptions, we tested 25 different combinations of pricing and identity verification strategies for collaborative registries. The values of α_{r_c} and β_{r_c} tested are shown in table 5. The different values of $\lambda_{\theta_{r_c}}$ for the same θ_{r_c} symbolize different possible costs for defeating the identity verification method suggested by us in section 3.3. We model non-collaborative registries and registrants to be best responding to the strategies of collaborative registries.

Interestingly, in this scenario, benign registrants will always register all of their domains. If $\alpha_{r_c} \in \{100, 1000\}$ then non-collaborative registries will drop their prices to get benign registrants' business. This results in a huge drop in the utility of both the collaborative registries and the benign registrants. Increasing θ_{r_c} leads to a drop

Table 5: Pricing and verification strategies for the base scenario.

Pricing strategy	α_{r_c}	10	10	10	100	1000
	β_{r_c}	1	2	3	1	1
Identitity verification	θ_{r_c}	0	1	4	4	4
	$\lambda_{\theta_{r_c}}$	0	1	10	100	1000

in benign registrants utility, but until $\alpha_{r_c} = 10$ they will keep their domains at the collaborative registries.

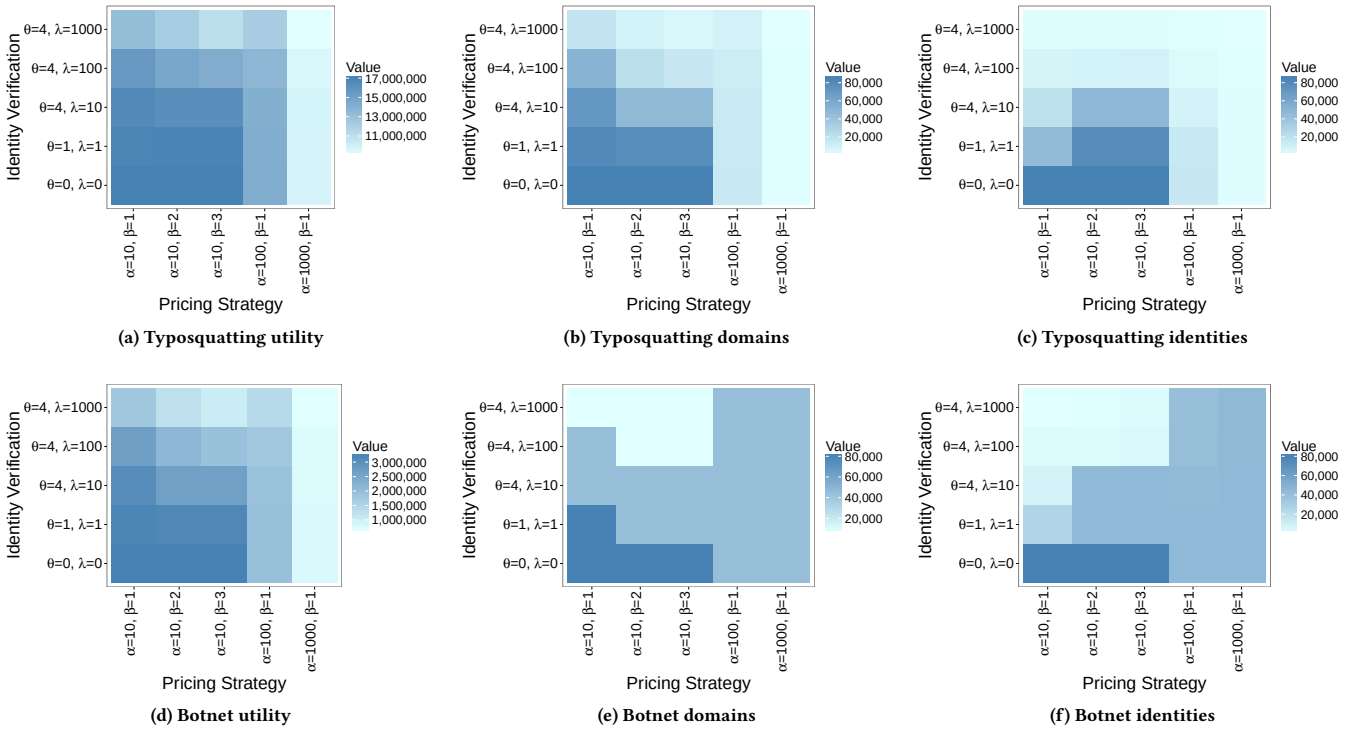
Figure 3 shows the effects of different registrations strategies on malicious users. Setting α_{r_c} high has a significant impact on malicious registrants but it also negatively affects other entities in the ecosystem. A better solution is to keep $\alpha_{r_c} = 10$ and increase β_{r_c} and θ_{r_c}. We can see that even a small increase in θ_{r_c} can affect the utility and the domain registration behavior of malicious registrants slightly. Most interesting is setting $\theta_{r_c} = 4$ and analyzing how different possible $\lambda_{\theta_{r_c}}$ affect miscreants. We can see in Figures 3a and 3d that any value of $\lambda_{\theta_{r_c}}$ has a significant effect on the utility and domain registration behavior of crooks. Analyzing Figures 3b, 3e, 3c, and 3f, we observe that at cells corresponding to $\lambda_{\theta_{r_c}} \in \{1, 10\}$ and $\beta_{r_c} \in \{2, 3\}$ some miscreants still keep their domain names but they need to buy a lot of stolen identities decreasing their profit. When $\lambda_{\theta_{r_c}} \in \{100, 1000\}$, most criminals need to switch registries or give up their domain registrations. However, even in the most adversarial settings, the most successful criminals will continue using their domain names.

Table 6: The effects of $\lambda_{\theta_{r_c}}$ on malicious registrants, when $\alpha_{r_c} = 10$ and $\beta_{r_c} = 3$.

	Typosquatters			Botnets		
$\lambda_{\theta_{r_c}}$	Utility	# doms	# iden	Utility	# doms	# iden
1000	64.9	7.2	2.1	32.9	9.9	3.5
100	83.5	18.4	10.0	58.9	9.9	5.1
10	94.8	55.2	55.2	80.0	55.0	55.0
1	99.0	90.3	90.3	96.3	55.0	55.0
0	100	100	100	100	100	100

Table 6 shows that even a small increase in $\lambda_{\theta_{r_c}}$ has an effect on the number of domains registered and the utility of criminals. $\lambda_{\theta_{r_c}} = 100$ appears to be where the number of malicious domain registration drastically drops. Interestingly, certain combinations of personal documents available on online black markets hover around $100, based on empirical data [50]. However, even drastic drops in domain registrations do not affect the most successful criminals, and have thus a slightly more limited impact on total (aggregate) miscreant utility.

Figure 4 shows that only the very few top typosquatters are not affected by $\lambda_{\theta_{r_c}}$ the cost of stolen identities. Surprisingly to us, a few of the top ten typosquatters are also significantly affected by the increase in $\lambda_{\theta_{r_c}}$. It is likely that these typosquatters own many low or average quality typosquatting domain names and therefore they are increasingly affected by changes in $\lambda_{\theta_{r_c}}$. The



(a) Typosquatting utility (b) Typosquatting domains (c) Typosquatting identities

(d) Botnet utility (e) Botnet domains (f) Botnet identities

Figure 3: Effects of different pricing and identity verification strategies on malicious registrants.

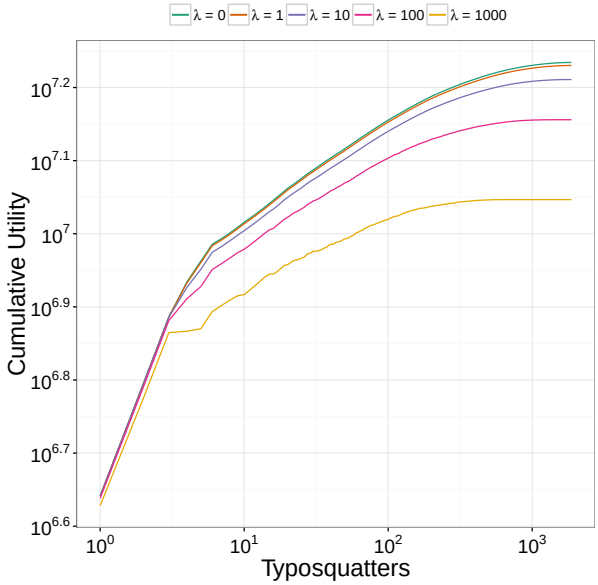


Figure 4: This plot shows the cumulative sum of typosquatters' utility for different values of $\lambda_{\theta_{r_c}}$ ($\alpha_{r_c} = 10$ and $\beta_{r_c} = 3$).

anti-squatting policy would be an effective complement against typosquatters not affected by the anti-bulk registration policy.

Changing the probability of blacklisting.

We originally assumed a low blacklisting probability. Here we answer the question of how effective registration policies are if there is no blacklisting of domain names $p_{\text{bl}} = 0$ or blacklisting will become much more effective $p_{\text{bl}} \in \{0.1, 0.5\}$. Not surprisingly, if there is no blacklisting then malicious registrants' utility is strictly higher than before. While registration policies are still effective, interestingly, the more effective a policy was in the base scenario the biggest impact $p_{\text{bl}} = 0$ had on increasing miscreants payoff. We observed the opposite effect when $p_{\text{bl}} = 0.5$. We find the synergy between the effectiveness of blacklisting and the effectiveness of registration policy strategies interesting: When a policy proposal was more effective, its effects on abusive domain registrations were disproportionally boosted by the increased performance of blacklisting.

Changing the cost of switching TLDs.

For the base scenario, we assumed that switching TLDs is costly for registrants, but if collaborative registries impose a high registration fee then registrants will switch. We test what happens if the cost of switching is higher or lower. When it does not cost anything to switch TLDs for registrants ($\frac{\gamma_{b,r_c}}{\gamma_{b,r_{nc}}} = 1$), then our results indicate that both benign and malicious registrants will switch TLDs. When $\frac{\gamma_{b,r_c}}{\gamma_{b,r_{nc}}} = 100$ registrants will not switch instead they stop registering domain name altogether when collaborative registries increase their prices.

The ratio of benign registrants.

For the base game, we assumed that there are about a hundred times more benign domains than abusive ones. We found that if the ratio of abusive registrants is higher, non-collaborative registries will be hungrier to gain the business of these malicious registrants and will drop their prices, sacrificing in the process income from regular users.

Registry utility.

As we discussed earlier, if collaborating registries set α_{r_c} high, benign registrants will not register their domain names with them. This leads to an extreme drop in the utility of registries, thus it is unlikely for them to adopt such a strategy. If instead, they increase β_{r_c} their utility also increases slightly because they decrease the utility of malicious registrants, thus they decrease the penalty weighted by $\rho_{r_c} = 0.1$. In summary, if registries are motivated (high enough ρ_{r_c}) then their best response will be to decrease malicious registrations while not hurting benign registrations. For future work, we hope to collect more data on malicious and speculative domain registrations. In the current model, we did not include speculative registrants, while the lack of their registrations is likely to significantly decrease the utility of registries if bulk registrations would become expensive.

5 CONCLUSION

We started with an overview of the domain registration ecosystem focusing on the political and financial dependencies of the most important entities. Building on this understanding, we summarized what decision-makers should consider when designing a domain registration policy. We then discussed the potential of several policy proposals. We found that a) anti-bulk registration, b) incentivizing registries and registrars, and c) anti-squatting were all potentially useful policy proposals. We believe leveraging all three of them together could potentially benefit the domain registration ecosystem the most.

We created a game-theoretical model to analyze the anti-bulk registration policy – using a variable pricing model – in more detail. The best strategies we found to fight online crime, for collaborating registries, are to increase the effectiveness of identity verification and to penalize bulk registrations. However, registries never want to increase their base price considerably because it would lead to a loss of customers. Because of the very strong asymmetry in miscreant success (where only a few miscreants succeed in earning their keep), we discovered that even the most successful domain registration policies would not significantly affect the most successful criminals and thus, may not considerably change the total revenue produced by miscreants. However, they could be particularly useful to remove from the pool the unsuccessful criminals, and drastically decrease abusive domain registrations overall. This result emphasizes the importance of combining registrations policies and to use them together with other lines of defenses.

6 ACKNOWLEDGMENTS

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. This work benefited from extensive discussions with Steve Sheng, Maciej Korczynski, and the Carnegie Mellon University cybercrime group. We also thank our anonymous reviewers and the reviewers from class for their useful feedback. Janos Szurdi would like to thank the National Science Foundation for the provided Student Travel Grant.

REFERENCES

[1] Wikipedia. List of most expensive domain names. https://en.wikipedia.org/wiki/List_of_most_expensive_domain_names, 2018.
[2] John D Mercer. Cybersquatting: Blackmail on the information superhighway. *BUJ Sci. & Tech. L.*, 6:290, 2000.
[3] Janos Szurdi, Balazs Kocso, Gabor Cseh, Jonathan Spring, Mark Felegyhazi, and Chris Kanich. The long" taile" of typosquatting domain names. In *USENIX Security Symposium*, pages 191–206, 2014.
[4] Panagiotis Kintis, Najmeh Miramirkhani, Charles Lever, Yizheng Chen, Rosa Romero-Gómez, Nikolaos Pitropakis, Nick Nikiforakis, and Manos Antonakakis. Hiding in plain sight: A longitudinal study of combosquatting abuse. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 569–586. ACM, 2017.
[5] Nick Nikiforakis, Marco Balduzzi, Lieven Desmet, Frank Piessens, and Wouter Joosen. Soundsquatting: Uncovering the use of homophones in domain squatting. In *International Conference on Information Security*, pages 291–308. Springer, 2014.
[6] ICANN. Icann registry agreements. https://www.icann.org/resources/pages/registries/registries-agreements-en, 2017.
[7] Zhou Li, Sumayah Alrwais, Yinglian Xie, Fang Yu, and XiaoFeng Wang. Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 112–126. IEEE, 2013.
[8] Paul Festa. Domain squatters losing out. https://www.cnet.com/news/domain-squatters-losing-out/, 1998.
[9] Kieren McCarthy. World's most notorious cybersquatter arrested. https://www.theregister.co.uk/2003/09/04/worlds_most_notorious_cybersquatter_arrested/, 2003.
[10] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. Dial one for scam: A large-scale analysis of technical support scams. In *Proceedings of the 24th Network and Distributed System Security Symposium (NDSS 2017). Internet Society*, 2017.
[11] Anh Le, Athina Markopoulou, and Michalis Faloutsos. Phishdef: Url names say it all. In *INFOCOM, 2011 Proceedings IEEE*, pages 191–195. IEEE, 2011.
[12] M Zubair Rafique, Tom Van Goethem, Wouter Joosen, Christophe Huygens, and Nick Nikiforakis. It's free for a reason: Exploring the ecosystem of free live streaming services. In *Proceedings of the 23rd Network and Distributed System Security Symposium (NDSS 2016)*, pages 1–15. Internet Society, 2016.
[13] Icann's many trips up capitol hill, part 1. https://www.bna.com/icanns-trips-capitol-b17179927466/, 2015.
[14] Congressional hearing: Internet domain name fraud - the u.s. government's role in ensuring public access to accuracte whois data. https://babel.hathitrust.org/cgi/pt?id=mdp.39015090379986;view=1up;seq=1, 2003.
[15] Congressional hearing: Icann and the whois database: Providing access to protect consumers from phishing. https://www.gpo.gov/fdsys/pkg/CHRG-109hhrg31537/pdf/CHRG-109hhrg31537.pdf, 2006.
[16] ICANN. Sac055, whois: Blind men and an elephant. https://www.icann.org/en/system/files/files/sac-055-en.pdf, 2012.
[17] ICANN. Operational point of contact final report. https://gnso.icann.org/sites/default/files/filefield_6454/icann-whois-wg-report-final-1-9.pdf, 2007.
[18] ICANN. Sac058, ssac report on domain name registration data validation. https://www.icann.org/en/system/files/files/sac-058-en.pdf, 2013.
[19] Nektarios Leontiadis and Nicolas Christin. Empirically measuring whois misuse. In *European Symposium on Research in Computer Security*, pages 19–36. Springer, 2014.
[20] Richard Clayton and Tony Mansfield. A study of whois privacy and proxy service abuse. In *13th Workshop on the Economics of Information Security*, 2014.
[21] Fraudulent online identity sanctions act (foisa). https://www.congress.gov/bill/108th-congress/house-bill/03754, 2004.
[22] Cybertelecom: Whois policy summary. http://www.cybertelecom.org/dns/whois.htm, 2017.
[23] ICANN. 2013 registrar accreditation agreement. https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#

whois-accuracy, 2013.
[24] ICANN. Whois primer. https://whois.icann.org/en/primer, 2017.
[25] ICANN. Final issue report on a next-generation gtld registration directory service (rds) to replace whois. https://whois.icann.org/sites/default/files/files/final-issue-report-next-generation-rds-07oct15-en.pdf, 2015.
[26] ICANN. Registration directory services review fact sheet. https://community.icann.org/display/WHO/Fact+Sheet, 2018.
[27] Manos Antonakakis, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster. Building a dynamic reputation system for dns. In *USENIX security symposium*, pages 273–290, 2010.
[28] Shuang Hao, Alex Kantchelian, Brad Miller, Vern Paxson, and Nick 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*, pages 1568–1579. ACM, 2016.
[29] Tristan Halvorson, Janos Szurdi, Gregor Maier, Mark Felegyhazi, Christian Kreibich, Nicholas Weaver, Kirill Levchenko, and Vern Paxson. The biz top-level domain: ten years later. In *International Conference on Passive and Active Network Measurement*, pages 221–230. Springer, 2012.
[30] Tristan Halvorson, Matthew F Der, Ian Foster, Stefan Savage, Lawrence K Saul, and Geoffrey M Voelker. From. academy to. zone: An analysis of the new tld land rush. In *Proceedings of the 2015 Internet Measurement Conference*, pages 381–394. ACM, 2015.
[31] He Liu, Kirill Levchenko, Márk Félegyházi, Christian Kreibich, Gregor Maier, Geoffrey M Voelker, and Stefan Savage. On the effects of registrar-level intervention. In *LEET*, 2011.
[32] Neha Chachra, Damon McCoy, Stefan Savage, and Geoffrey M Voelker. Empirically characterizing domain abuse and the revenue impact of blacklisting. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)*, page 4, 2014.
[33] Maciej Korczynski, Samaneh Tajalizadehkhoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten. Reputation metrics design to improve intermediary incentives for security of tlds. In *Security and Privacy (EuroS&P), 2017 IEEE European Symposium on*, pages 579–594. IEEE, 2017.
[34] ICANN. Domain abuse activity project report icann 60. https://www.icann.org/en/system/files/files/presentation-daar-31oct17-en.pdf, 2017.
[35] Maciej Korczynski, Maarten Wullink, Samaneh Tajalizadehkhoob, Giovane CM Moura, and Cristian Hesselman. Statistical analysis of dns abuse in gtlds final report. 2017.
[36] Tyler Moore, Richard Clayton, and Ross Anderson. The economics of online crime. *Journal of Economic Perspectives*, 23(3):3–20, 2009.
[37] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel JG Van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. In *The economics of information security and privacy*, pages 265–300. Springer, 2013.
[38] Kurt Thomas, Danny Yuxing Huang, David Wang Elie Bursztein Chris GrierD, Thomas J Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. Framing dependencies introduced by underground commoditization. In *Workshop on the Economics of Information Security*, 2015.
[39] Brett Stone-Gross, Ryan Abman, Richard A Kemmerer, Christopher Kruegel, Douglas G Steigerwald, and Giovanni Vigna. The underground economy of fake antivirus software. In *Economics of information security and privacy III*, pages 55–78. Springer, 2013.
[40] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. Measuring pay-per-install: the commoditization of malware distribution. In *Usenix security symposium*, pages 13–13, 2011.
[41] ICANN. Icann audited fianncial statement. https://www.icann.org/en/system/files/files/financial-report-fye-30jun17-en.pdf, 2017.
[42] ICANN. Icann mission statement. https://www.icann.org/resources/pages/governance/bylaws-en/#article1, 2017.
[43] Wikipedia. Tokelau. https://en.wikipedia.org/wiki/Tokelau, 2018.
[44] Kyle Soska and Nicolas Christin. Automatically detecting vulnerable websites before they turn malicious. In *USENIX Security Symposium*, pages 625–640, 2014.
[45] N. Leontiadis, T. Moore, and N. Christin. Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In *Proceedings of USENIX Security 2011*, San Francisco, CA, August 2011.
[46] K. Levchenko, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, A. Pitsillidis, N. Weaver, V. Paxson, G. Voelker, and S. Savage. Click trajectories: End-to-end analysis of the spam value chain. In *Proceedings of IEEE Security and Privacy*, Oakland, CA, May 2011.
[47] C. Herley and D. Florêncio. Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy. In *Proceedings (online) of the Worskshop on Economics of Information Security*, June 2009. Available from http://weis09.infosecon.net/.
[48] N. Christin. Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd World Wide Web Conference (WWW'13)*, pages 213–224, Rio de Janeiro, Brazil, May 2013.
[49] Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose. Virtual u: Defeating face liveness detection by building virtual models from your public photos. In *USENIX security symposium*, pages 497–512, 2016.
[50] K. Soska and N. Christin. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *Proceedings of the 23rd USENIX Security Symposium (USENIX Security'14)*, pages 33–48, Washington, DC, August 2015.
[51] Sumayah A Alrwais, Kan Yuan, Eihal Alowaisheq, Zhou Li, and XiaoFeng Wang. Understanding the dark side of domain parking. In *USENIX Security Symposium*, pages 207–222, 2014.
[52] Thomas Vissers, Wouter Joosen, and Nick Nikiforakis. Parking sensors: Analyzing and detecting parked domains. In *Proceedings of the 22nd Network and Distributed System Security Symposium (NDSS 2015)*, pages 53–53. Internet Society, 2015.
[53] Maciej Korczynski, Maarten Wullink, Samaneh Tajalizadehkhoob, Giovane C.M. Moura, Arman Noroozian, Drew Bagley, and Cristian Hesselman. Cybercrime after the sunrise: A statistical analysis of dns abuse in new gtlds. In *Proceedings of the 2018 ACM Asia Conference on Computer and Communications Security*. ACM, 2018.
[54] Janos Szurdi and Nicolas Christin. Email typosquatting. In *Proceedings of the 2017 Internet Measurement Conference*, pages 419–431. ACM, 2017.
[55] Tyler Moore and Benjamin Edelman. Measuring the perpetrators and funders of typosquatting. In *International Conference on Financial Cryptography and Data Security*, pages 175–191. Springer, 2010.
[56] Tyler Moore and Benjamin Edelman. Online appendix for measuring the perpetrators and funders of typosquatting. http://www.benedelman.org/typosquatting/pop.html, 2010.
[57] Google. Efficient frontier's automotive clients receive twice the conversion rate as search with domain ads. http://www.google.com/adwords/casestudies/EfficientFrontierAFDCaseStudy.pdf, 2010.
[58] Andreas Pitsillidis, Chris Kanich, Geoffrey M Voelker, Kirill Levchenko, and Stefan Savage. Taster's choice: a comparative analysis of spam feeds. In *Proceedings of the 2012 Internet Measurement Conference*, pages 427–440. ACM, 2012.