

The BIZ Top-Level Domain: Ten Years Later

Tristan Halvorson¹, Janos Szurdi², Gregor Maier³, Mark Felegyhazi², Christian Kreibich³, Nicholas Weaver³, Kirill Levchenko¹, and Vern Paxson³

¹ UC San Diego

² CrySys Lab., Budapest University of Technology and Economics

³ International Computer Science Institute

Abstract. On May 15, 2001 ICANN announced the introduction of the `biz` and `info` generic top-level domains (gTLDs)—the first new gTLDs since the inception of the Domain Name System—aiming to “increase consumer choice and create opportunities for entities that have been shut out under the current name structure.” The `biz` gTLD, in particular, was to become an alternative to the popular `com` top-level domain.

In this paper we examine the current usage of the `biz` gTLD in order to determine whether it has evolved into the role intended by ICANN, and whether concerns expressed in the early discussions of this expansion have been justified. In particular, using DNS zone files, DNS probing, and Web crawler data, we attempt to answer the question of whether `biz` has become a viable alternative to `com`, giving trademark holders who find themselves unable to register a `com` name an attractive alternative; or whether it has merely induced defensive registrations by existing trademark holders who already had equivalent `com` domains.

1 Introduction

The Domain Name System (DNS) serves as the main means of giving names to entities on the Internet. Now over 25 years old, its original goal was to provide a way to consistently map host names to IP addresses. Since then, DNS has grown to identify the Internet presence of brands, companies, products, political campaigns, and individuals.

Compared to other means of navigating the Internet, DNS has a very rigid lookup structure: a user must know the name precisely and completely in order to perform a lookup. While search engines provide a more flexible navigation mechanism, we still perceive the need for a direct mechanism like the domain system: one that remains free of third-party mediation, succinct enough to allow expression in print, and easy for users to act on.

In part for historic reasons, DNS is arranged hierarchically: names associate with a suffix, usually a top-level domain (TLD). The original DNS design included generic TLDs (gTLDs) (`com`, `edu`, etc.) following administrative considerations, matching the classes of organizations with Internet hosts to name. The growth of the Web created the need to name not individual hosts but broader entities. Of the original gTLDs, `com` proved the most open and best suited choice, becoming the de facto TLD for Web addresses.

To the average user, `com` became synonymous with the Web, its iconic status earning it a place in the Oxford English Dictionary in 1994. As a consequence of its popularity, the `com` landscape quickly became crowded. To ease the pressure on `com`, ICANN

moved to create more gTLDs, introducing `biz` and `info` in 2001. The success of this endeavor hinged on how users would perceive TLDs: whether users accepted the notion of TLDs as simply reflecting different potential homes for various Internet entities, or whether they viewed “dot com” as the sole TLD where they would expect to find prominent Web enterprises. Would users find it confusing to encounter the same subdomain in different TLDs?

A decade has passed since, and ICANN now promotes a new, more open system for introducing new TLDs which could dramatically change the Domain Name System. To answer some of these questions, and to re-evaluate some early predictions, this paper takes stock of the `biz` gTLD, which largely promoted as an alternative to `com`. We find that approximately 20% of domains in both `biz` and `com` are *parked*, contrary to original intention to avoid domain speculation in `biz`. Furthermore, between 10% and 25% of `biz` registrations appear only to exist to defend against name infringement.

The rest of this paper proceeds as follows. Section 2 gives a background on the `biz` TLD with a survey of related work. Section 3 describes our data collection methodology, followed by our analysis in Section 4, and a brief discussion of our findings in Section 5. Section 6 concludes the paper.

2 Background

In June 1998, the US Department of Commerce called for the creation of a new not-for-profit organization to assume responsibility for the technical coordination functions performed by the IANA. Later that year, the Internet Corporation for Assigned Names and Numbers (ICANN) was formed to fulfill that role. ICANN obtained “the authority to manage and perform a specific set of functions related to coordination of the domain name system, including the authority necessary to . . . oversee policy for determining the circumstances under which new TLDs are added to the root system” [8]. In 1999 ICANN formed the Domain Name Supporting Organization (DNSO), an advisory body within ICANN, to handle matters concerning the Domain Name System [6], chartering, within the DNSO, Working Group C to study the issues surrounding the formation of new generic top-level domains (gTLDs). The Working Group’s task was to gauge the need for new gTLDs, and, if deemed required, to determine what should be their nature and deployment policy [3]. In March 2000 Working Group C released its final report, addressing these questions [10]. The ICANN board adopted its recommendations in July 2000 [5].

2.1 The Working Group C Report

The Working Group C report [10] addressed two important questions: whether to create new gTLDs, and, if so, how to introduce them. On the need for new gTLDs, the report argued: “Expanding the number of TLDs will increase consumer choice, and create opportunities for entities that have been shut out under the current name structure.” Moreover, the report observes:

Existing second-level domain names under the `.com` TLD routinely change hands for enormously inflated prices. These are legitimate trades of ordinary, untrademarked words;

their high prices reflect the artificial scarcity of common names in existing gTLDs, and the premium on .com names in particular.

...

If the name space is expanded, companies will be able to get easy-to-remember domain names more easily, and the entry barriers to successful participation in electronic commerce will be lowered. *Addition of new TLDs will allow different companies to have the same second-level domain name in different TLDs* (emphasis added).

Arguing against the consensus position of the report, some members of the Working Group suggested that “an increase in the number of top-level domains could confuse consumers,” and that expanding the domain space “will likely increase trademark owners’ policing costs and the costs of defensive registrations.” Members expressed concerns about “*trademark holders simply duplicating their existing domains*” (emphasis added).

On the nature of the initial rollout, the report weighed two opposing approaches: authorizing “hundreds of new TLDs over the course of the next few years” and introducing new TLDs “slowly and in a controlled manner, and only after effective trademark protection mechanisms had been implemented and shown to be effective.” In arguing for the introduction of many new TLDs, some members argued that “a small number of new gTLDs with no commitment to add more . . . *would encourage pre-emptive and speculative registrations based on the possibility of continued artificial scarcity*” (emphasis added). The report ultimately proposed “deploying six to ten new TLDs.”

2.2 Recent Studies

In its recent move to open TLD registration to the general public, ICANN commissioned several reports analyzing the economic consequences of their new initiative. Among other concerns, the reports address the danger that new gTLDs will compel trademark holders to defensively register their marks in each new TLD. The 2009 Carlton report [2] dismisses this as a problem, arguing that “many registrations that ‘redirect traffic’ to other sites serve productive purposes of attracting and retaining Internet traffic, not merely to prevent cybersquatting.” Furthermore, the report argues: “While some of the registrations for domain names under the new gTLDs may have been made for defensive purposes, the limited number of registrations for new gTLDs indicates that the vast majority of .com registrants did not find a compelling reason to undertake defensive registrations in the new gTLDs.” The 2010 Katz, Rosston, Sullivan report [7] looked back at past TLD introductions: “A broad look at the evidence to date suggests that other gTLDs provide little competition for .com as those gTLDs have neither attracted a large number of domains nor shown any signs of catching up with .com.”⁴

Most closely related to our work is the 2002 survey of biz registrations by Zittrain and Edelman [11]. They found that 90.5% of names registered in biz also occur in com (consistent with our finding). To assess whether biz and com registrants overlapped, the survey authors examined Whois records, comparing records based on registrant postal code, email address, and name server second-level domains. They found that 35.4% of biz-com pairs matched in at least one attribute, 25.8% in two, and 12.9% in all three.

⁴ The report also notes that “the prospects of cyber-squatting within new gTLDs could force parties to undertake defensive registrations.”

3 Data & Methodology

This paper aims to answer two basic questions: how do owners of `biz` domains use their domains, and, if in active use, whether the `biz` domain forms the *primary* domain of the registrant or whether it merely *defends* one registered under another gTLD. To start, we obtained the `biz` and `com` zone files, dated June 27, 2011, to coincide with the 10-year anniversary of the `biz` TLD’s addition to the authoritative root server. We use three sets of domain names: all 2.1 million `biz` domains, their 2 million (94%) `com` namesakes, and a random sample of 2 million `com` domains. We rely on four sources of data to classify domains: zone files, active DNS queries, Whois registration records, and Web content.

3.1 Zone Files

A zone file contains the DNS records used by a name server, typically in BIND format. We obtained the zone files for the `biz` and `com` gTLDs from their respective registries. We used the zone files to get the list of `biz` and `com` domains and determine which domains were registered in both gTLDs. We also gathered their name server information for the DNS crawler.

3.2 Whois

We retrieved the whois registration information for each `biz-com` pair in our data set. Since whois records consist of free-form text, we use a customized version of `phpwhois`⁵ to parse the whois records and extract the domain registrant (owner) information. Many registrars have limits on the number of whois queries they will answer. Additionally, `phpwhois` could not parse all whois entries. Due to these limitations, we could only extract registrant information from 65% of our `biz-com` pairs.

In order to assess whether the `biz` domain and its `com` namesake share the same owner, we compare the registrant information returned by `phpwhois`. We first exclude domains that use whois privacy mechanisms (e.g., the Domains by Proxy service). We compute the Levenshtein distance between both domains for each of: the registrant’s name, e-mail address, phone, and fax number. We mark each of these fields as missing if they are absent or less than 5 characters long in either domain. We mark a field as a match if present and the Levenshtein distance between `biz` and `com` does not exceed 2 (requiring an exact match does not significantly alter our results). We also mark registrant names as a match if the name from one domain forms a substring of the other domain.

We consider two whois records a strong match if at least two of the four categories match and at most one category is missing. We consider them a weak match if any of the four categories match. We use both types of matches, but differentiate between them in our analysis.

We note that for a significant number of `com` domains, the whois record we retrieved only contained the registrant name but no further fields. We therefore cannot have a strong match for any of these domains.

⁵ <http://sourceforge.net/projects/phpwhois/>

3.3 DNS Probing

We queried the DNS records for a list of all `biz` domains and their `com` counterparts, as well as the randomly selected set of `com` domains. For each name, we queried (starting at the root) to find the authoritative name servers for the `biz` and `com` versions of the domain. We performed the crawling with a custom Python library on September 12th and 13th, 2011.

3.4 Web Crawl

We collected the content of the web pages belonging to the registered domain in our data sets. First, we downloaded the pages for the domains in the `biz` zone file, e.g., `foo.biz`. Then, we crawled the corresponding `com` domain `foo.com` to check the registration purpose for the `biz` domain. When downloading the web pages, we recorded the HTTP status codes for success, redirection, errors and other standardized events. We also recorded unknown errors. Note that at times we could not retrieve the Web pages, either because the domains' owners want to serve no Web content, or due to the time interval (several months) between the zone file creation and our active Web crawling.

3.5 Content Classification

To identify parked domains, we built a simple classifier that searches for a set of regular expressions in the downloaded content. We created highly specific patterns to match templates for the largest known parking sites. We relied on unique features of the page, such as JavaScript libraries or image servers used by the parked pages.

4 Analysis

Recall that we set out to evaluate a number of predictions made about the effects of introducing a new gTLD: whether it would lead to “trademark holders simply duplicating their existing domains” or “will allow different companies to have the same second-level domain name in different TLDs”; whether it “would encourage pre-emptive and speculative registrations based on the possibility of continued artificial scarcity” [10]. To answer these questions, we first group `biz` domains into three functional categories:

Primary. *Primary* domains identify a company, product, service, or organization, either publicly or internally. In other words, the registrant *actually uses* primary domains. For example, NeuStar, Inc. and UC San Diego use `neustar.biz` and `ucsd.edu`, respectively, as the primary domain names by which they identify themselves on the Internet.

Defensive. A registrant uses a domain registration only to *defend* a name while not actively employing the domain to identify itself, its service, or network resources. Examples of defensive registrations are `google.biz` and `gooogle.com`, both of which redirect to `google.com`.

Parked. Registrants *park* domains with the purpose of reselling them or generating advertising revenue from accidental user visits to the site.

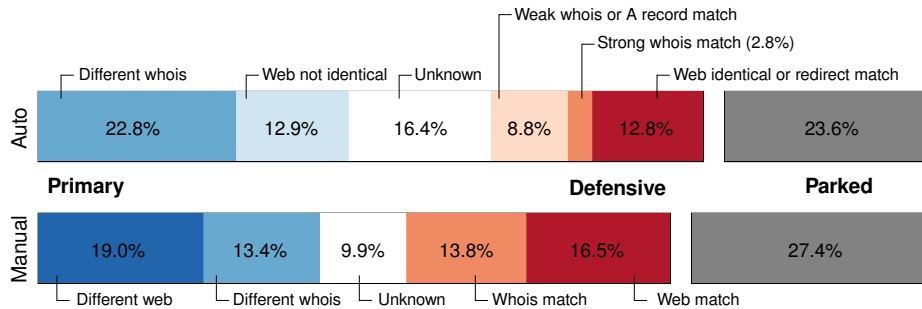


Fig. 1: Disposition of `biz` domain names with respect to their `com` namesakes based on our automatic classification (top) and manual classification of a sample of 485 domains (bottom). Excludes 6.2% of `biz` domains without a corresponding `com` namesake. Darker bands indicate higher confidence.

Defensive registrations prevent another party from misrepresenting itself as the registrant or from simply capturing traffic (intended for the registrant) for advertising purposes. A defensively registered domain is one not used by the registrant to name and identify products, services, or network infrastructure.⁶

Although it is nearly impossible to divine the registrant’s intention with absolute certainty, certain network-visible characteristics of a domain serve as indicators of primary or defensive use. In particular, we consider domain ownership, Web content, and hosting infrastructure sharing as indicators of primary or secondary use. Our results are summarized in Figure 1.

4.1 Parked Domains

A parked domain is not actively used by the registrant, and does not represent a name or brand used by the registrant. Registrants typically hold parked domains with the intention of selling them at a profit or monetizing accidental Web traffic through advertising. Parked domains are easily identified by prominent advertising on the domain’s site that one may purchase the domain and usually includes additional advertising.

We rely on the Web content hosted at the domain as our primary indicator of a parked domain. Table 1 shows the proportion of parked sites in the `biz` and `com` gTLDs. Figure 1 shows the number of parked `biz` domains (23.6%) having a `com` namesake.

4.2 Identical Web Content

Owners of defensively registered domains frequently reroute all Web traffic to the intended (primary) domain, usually via HTTP redirection [4]. Because the browser ac-

⁶ The difference between a defensive registration and either cybersquatting or typosquatting (registering misspellings of popular brands) lies in the identity of the registrant: when the registrant also owns the intended brand name or trademark, the registration is defensive; when the registrant is a third party with no legitimate claim to the name, the registration constitutes cybersquatting or typosquatting.

<i>Category</i>	<i>biz</i>	<i>com</i>
No server	23.5%	17.4%
HTTP Error	3.4%	3.3%
Parked	22.8%	19.4%
Redirect	18.5%	17.3%
On-site	5.1%	8.5%
Namesake	4.1%	0.4%
Other site	9.1%	8.7%
Content served	31.7%	39.9%
Same as com	3.0%	—
Distinct	27.7%	—

Table 1: The Web behavior of domains in the *biz* and *com* gTLDs. The *biz* column shows statistics for the 2.1 million domains in the *biz* TLD, the *com* column for a random sample of 2 million *com* domains.

<i>Category</i>	<i>Abs</i>	<i>Rel</i>
Unknown	693,393	35.1%
Privacy guard	281,417	14.2%
<i>biz</i> only	97,802	5.0%
<i>com</i> only	82,161	4.2%
both	101,454	5.1%
Match	424,683	21.5%
Weak	308,337	15.6%
Strong	116,346	5.9%
No match	573,388	29.1%

Table 2: Comparison of registrants of *biz* domains and their *com* namesakes using Whois records, showing absolute and relative number of *biz*-*com* name pairs in each category. Rightmost column shows value relative to total number of *biz*-*com* name pairs (93.8% of *biz* names).

tively follows HTTP redirects, this method has the advantage of changing the user-visible address bar to reflect the new address (the target of the redirect). The user thus sees the *correct* address, consistent with the branding of the site.

Table 1 includes statistics about this mechanism: 18.5% of *biz* and 17.3% of *com* domains host a Web server that redirects the user. However, 4.1% of *biz* sites redirect to a site hosted at the same domain name in a different TLD, compared to 0.4% for *com*. We also see defensive registrations where both the *biz* and *com* Web servers redirect a user to the same third domain. In our set of domains in both *biz* and *com*, 6.1% have identical redirects.

Despite the effectiveness of HTTP redirects and their advantage of “correcting” the user, site operators also may simply maintain identical Web sites under both domains. To detect this condition, we compared the content of each *biz* site to that of its *com* namesake. Upon examination, we found that 3.0% of non-parked sites did indeed serve the same content. We only classify pages as identical if the source matches exactly, meaning a match almost certainly indicates a defensive registration, but a mismatch only weakly indicates a primary one.

As shown in the dark red portion of Figure 1, 12.8% of *biz*-*com* domains have identical web content or redirects, a strong indicator of a defensive registration. 12.9% of *biz*-*com* domains have different web content and no stronger classifiers, and so are weakly classified as primary registrations.

4.3 Common Registrant

The identity of the registrant provides another classification feature. One registrant owning both a *biz* domain as well as its *com* namesake likely suggests defensive registra-

tion. To identify such cases, we extracted registrant information from publicly available Whois records, as described in Section 3.2.

We could retrieve and successfully parse both Whois records for 65% of all biz-com pairs.⁷ Of these 65%, 10.1% of biz domains and 9.3% of com domains showed some manner of “privacy protection” mechanism, blocking the registrant information from appearing in the Whois record and leaving 50.6% of all biz-com pairs that could potentially match.

We grouped these pairs into three categories based on the degree to which we believed we identified the same registrants: weak matches, strong matches, and no match. Using the methodology described in Section 3.2, we determine 5.9% of pairs a strong match, another 15.6% a weak match, and 29.1% unlikely to be the same registrant. (Put another way, we found at least some degree of a match for around 40% of the pairs we could assess.)

As shown in Figure 1, we consider whois data to be more reliable than common hosting or different web content, since either of those may be incidental. We consider it to be less reliable than HTTP redirects and identical web content, since those are strong indicators of a defensive registration. After using stronger indicators, we classify 11.6% of biz domains as defensive based on whois and A record data (see Section 4.4) and 22.8% as likely primary.

4.4 Shared Infrastructure

We also used DNS crawling to observe infrastructure sharing between a biz domain and its com counterpart. In particular, we used CNAMEs and common A records as evidence of defensive registrations.

We identify a CNAME match in two different cases: first, when domains in both biz and com have CNAMEs pointing to the same domain; second, when the domain in biz has a CNAME pointing to its com namesake, or vice versa. Of the 2 million domains in both biz and com, 32,431 (1.6%) show common CNAMEs, which demonstrates a clear relationship between the two domains. We include these in the “redirect match” category in Figure 1.

We see many more domains with common A records. Our crawler observed common A records in 439,890 domains (22%) between biz and their com counterparts. We see three plausible explanations for such sharing: first, defensive registrations; second, coincidental common hosting, with unrelated owners of the biz and com employing common hosting infrastructure; third and finally, parking the biz and com domains in the same domain parking infrastructure. Since we classify parked domains first and only distinguish between primary and defensive registrations after considering all parked domains, we can ignore the third case.

While common A records do suggest a defensive registration, we cannot reliably distinguish true sharing and common hosting. Because of this, we consider common A records weak evidence of a defensive registration.

⁷ The delegated nature of the com Whois system means that these 65% necessarily constitute a biased sample, because being able to retrieve and parse a given Whois record depends on the registrar, specifically on their query rate limitations and record formatting.

5 Discussion

Our analysis finds 22.8% of `biz` domains *parked* with a known parking service. We can with certainty classify another 12.8% of `biz` domains as *defensive* registrations, leaving two thirds undetermined. At least 27.7% of these served some kind of content (excluding cases where this content proved identical to the `com` namesake). In addition, of the pairs for which we could assess non-private registrant information, we found at least a degree of match between the `biz` and `com` registration in 40% of the instances (§ 4.3), indicating a substantial level of registrations likely made defensively.

To get at the fundamental value (to registrants) of the `biz` TLD, we can approach the question from the other direction: how many `biz` domains do registrants use actively? We assessed the popularity of `biz` domains in the Alexa [1] Web site rankings, as well as the popularity of `biz` domains in the Open Directory Project [9]. We show the results in Table 3, along with other common TLDs.

TLD	Alexa 1M	Alexa 500	ODP
<code>com</code>	55.3%	64.6%	41.7%
<code>net</code>	6.26%	4.60%	3.74%
<code>org</code>	4.01%	2.80%	9.00%
<code>ru</code>	3.75%	2.40%	1.46%
<code>de</code>	3.70%	1.40%	9.33%
<code>info</code>	1.82%	0%	0.480%
<code>biz</code>	0.396%	0.200%	0.188%

Table 3: TLD frequency in the Alexa listings and the Open Directory Project. In the Alexa 1,000,000, `biz` ranks (in frequency of occurrence) between `com.cn` and `ir`, while in the ODP, it falls between `cat` and `za`. Only one `biz` domain, `livedoor.biz` (a blogging site), resides in the Alexa 500.

The `biz` TLD occurs 140 times less frequently than `com` in the Alexa 1 million, 323 times less frequently in the Alexa 500 (based on 1 occurrence), and 218 times less frequently in the Open Directory Project. Note that the `com` zone is about 46 times larger than `biz`. Although not a formal assessment of usage, these statistics suggest a disproportionately lower popularity of `biz` compared to `com`.

6 Conclusion

In this paper we examined the current state of the `biz` TLD on its ten-year anniversary. We found that in many respects, most notably in the prevalence of domain speculation (parking), `biz` resembles `com`. And while one could conclude that it has failed to rival `com`, `biz` did extract defensive registrations from existing domain owners. Although registering these domains costs no more than \$10 each, this cost is dwarfed by the additional costs of defending trademarks (via resolution procedures and litigation) in a new TLD.

On the eve of a bold new initiative by ICANN to open TLD registration to the general public, biz provides a valuable lesson in the costs and benefits associated with expanding the DNS name space. In particular, we anticipate a level of defensive registrations that, in aggregate, will reflect significant imposed expenditures.

7 Acknowledgements

This work was supported in part by the National Science Foundation under grants CNS-0905631 and CNS-1111672; by the Office of Naval Research under MURI Grant No. N000140911081; and a fellowship within the postdoctoral program of the German Academic Exchange Service (DAAD). This work is connected also to the scientific program of the “Development of quality-oriented and harmonized R+D+I strategy and functional model at BME” project. This project is supported by the New Széchenyi Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002). Opinions expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

1. Alexa: Alexa top 500 global sites. <http://www.alexa.com/topsites> (June 2011)
2. Carlton, D.: Report of Dennis Carlton regarding ICANN’s proposed mechanism for introducing new gTLDs. <http://www.icann.org/en/topics/new-gtlds/carlton-reproposed-mechanism-05jun09-en.pdf> (June 2009)
3. Domain Name Supporting Organization: Working Group C – creation of new gTLD. <http://www.dnso.org/dns/notes/19990625.NCwgc.html> (June 1999)
4. Fielding, R.T., Gettys, J., Mogul, J.C., Nielsen, H.F., Masinter, L., Leach, P.J., Berners-Lee, T.: Hypertext Transfer Protocol — HTTP/1.1. RFC 2616, The Internet Society (Jun 1999)
5. Internet Assigned Numbers Authority: Iana report on establishment of the .biz and .info top-level domains. <http://www.iana.org/reports/2001/biz-info-report-25jun01.html> (June 2001)
6. Internet Corporation for Assigned Names and Numbers: Domain Name Supporting Organization formation concepts. <http://www.icann.org/en/meetings/singapore/dns-formation.htm> (March 1999)
7. Katz, M.L., Rosston, G.L., Sullivan, T.: An economic framework for the analysis of the expansion of generic top-level domain names. <http://www.icann.org/en/topics/new-gtlds/economic-analysis-of-new-gtlds-16jun10-en.pdf> (June 2010)
8. National Telecommunications and Information Administration: Statement of policy on the management of internet names and addresses. <http://www.ntia.doc.gov/federal-register-notice/1998/statement-policy-management-internet-names-and-addresses> (June 1998)
9. ODP – Open Directory Project. <http://www.dmoz.org> (September 2011)
10. Weinberg, J.: Report (part one) of Working Group C of the Domain Name Supporting Organization Internet Corporation for Assigned Names and Numbers. <http://www.dnso.org/dns/notes/20000321.NCwgc-report.html>
11. Zittrain, J., Edelman, B.: Survey of usage of the .BIZ TLD. <http://cyber.law.harvard.edu/tlds/001/> (June 2002)