The Ghosts of Banking Past: Empirical Analysis of Closed Bank Websites

Tyler Moore[1] and Richard Clayton[2]

[1] Computer Science and Engineering Department,
Southern Methodist University, Dallas, TX, USA
tylerm@smu.edu
[2] Computer Laboratory, University of Cambridge, UK
richard.clayton@cl.cam.ac.uk

Abstract. We study what happens to the domains used by US banks for their customer-facing websites when the bank is shut down or merges with another institution. The Federal Deposit Insurance Corporation (FDIC) publishes detailed statistical data about the many thousands of US banks, including their website URLs. We extracted details of the 3 181 banks that have closed their doors since 2003 and determined the fate of 2 302 domain names they are known to have used. We found that 47% are still owned by a banking institution but that 33% have passed into the hands of people who are exploiting the residual good reputation attached to the domain by hosting adverts, distributing malware or carrying out search engine optimization (SEO) activities. We map out the lifecycle of domain usage after the original institution no longer requires it as their main customer contact point – and explain our findings from an economic perspective. We present logistic regressions that help explain some of reasons why closed bank domains are let go, as well as why others choose to repurpose them. For instance, we find that smaller and troubled banks are more likely to lose control of their domains, and that the domains from bigger banks are more likely to be repurposed by others. We draw attention to other classes of domain that are best kept off the open market lest old botnets be revivified or other forms of criminality be resurrected. We end by exploring what the public policy options might be that would protect us all from ghost domains that are no longer being looked after by their original registrants.

1 Introduction

Many countries have just a handful of High Street banks, each with branches nationwide. The USA is an exception, in that although there are a number of national or regional brands, there are still many local banks – with perhaps only one branch, or just a couple more in neighboring towns. The US banking sector is underpinned by a government promise that should a bank fail then depositors will get their money back (up to $250,000). The databases created by the administration of this scheme make it relatively straightforward to find data

18th International Conference on Financial Cryptography and Data Security (FC), 2014
Author's pre-publication copy

about US banks – as of 31 March 2013 there were 7 019 institutions that were insured by the Federal Deposit Insurance Corporation (FDIC).

In Spring 2013 we came across what appeared to be a legitimate website, albeit of somewhat dated design, for the Mid Valley Bank, Corning, California, USA. What caught our eye was that on their "News" page they had several stories which appeared to be 'astroturfing' puffs for rare earth metal investments, gold sales and reverse mortgages. Alongside this they had news stories from 2010 on their quarterly financial results, but when we clicked through the pages were dated 2013. In fact, not only were they dated 2013 but some stories even referred to events that would occur several months into the future.

We used a search engine and found Mid Valley Bank listed on white pages websites such as Yelp, Merchant Circle and MapQuest. However we also found links to the FDIC website. This explained that on 23 January 2004 the bank was "merged without assistance" into PremierWest Bank. This is presumably why when we followed another link on the first page of the search results to `lendio.com` (a company founded in 2006 that puts businesses in touch with lenders) their webpage about the Mid Valley Bank marks the details as "not verified".

Examining the history of the `midvalleybank.com` domain we find that it was first registered on 19 July 1996 by the Mid Valley Bank. By 22 February 2008 it was registered in the name of an employee of PremierWest Bank but the domain was allowed to expire on 18 July 2009. It was re-registered on 3 October 2009 by a resident of Novokuznetsk, Russia, a town 500km SE of Novosibirsk and 800 km from the Mongolian border. On 8 October 2010 the registration changed to a proxy service which suggests that its ownership may have changed hands again. It remains registered under a proxy service to the present time.

The Internet Archive `www.archive.org` records that the current website design was put in place sometime between June 2009 and 10 October 2010 – at which time the forward looking statements about financial results now present were dated consistently with the reporting of then recent events. However, the archive shows that identical reports (with exactly the same profit/loss/asset numbers) were posted by the real bank in 2002, and the current website design was in use by the real bank between October 2000 and July 2004, after which a redirection page (to `premierwestbank.com`) was present.

Thus we had determined that one closed bank's website had come back to life with somewhat dubious content. We therefore decided to ascertain how common such resurrections are and then identify how the public might best be protected when domains with a substantial reputation become surplus to requirements.

In Section 2 we discuss the FDIC banking database and how it comes to contain banking domain names. In Section 3 we examine the current state of the domain names of 2 393 of the banks that have merged or been shut down since July 2003. We propose a 'life cycle' for banking domains, with a common progression from each stage of reuse to the next.We find that large banks (as measured by total deposits) are more likely to retain old domains, and that non-banks are nearly always responsible for the resurrection of expired bank domains. We describe methods for identifying when non-banks impersonate banks

on domains from closed banks and for locating at-risk domains that may soon fall out of bank control. In Section 4 we discuss policy options for proactively dealing with the domains of closed banks in order to protect the public interest. We conclude by discussing related work in Section 5 and by summarizing our findings in Section 6.

2 Data Collection and Analysis Methodology

We first describe our approach to identifying the 'ghost' websites associated with closed banks in Section 2.1. We then describe in Section 2.2 a methodology for classifying how the websites are being used and whether or not the bank still retains control over the domains. The collected data and analysis scripts are publicly available for replication purposes at `doi:10.7910/DVN/26011`.

2.1 FDIC Data Collection

Franklin D. Roosevelt was inaugurated as US President on 4 March 1933 amidst a banking crisis – confidence had evaporated, investors were withdrawing their funds, and banks were closing their doors because they did not have the currency to fund withdrawals. FDR declared a banking holiday on 6 March 1933 and banks were only allowed to reopen once federal inspectors had declared them sound and that they had access to sufficient capital. This restored confidence and investors queued up again to return their funds to the banks that reopened.

This system of federal investor deposit insurance was regularized by the Banking Act of June 1933 which created the Federal Deposit Insurance Corporation (FDIC). The FDIC examines and supervises US banks, including state banks. Should a bank fail then the FDIC will manage it in receivership, and it also has a rôle in ensuring mergers occur so as to prevent a bank from failing.

The FDIC provides an online database in which are recorded all of the institutions that it has supervised, including those which no longer exist, having merged or failed.[1] This database is populated from the quarterly questionnaires that all supervised banks must complete, and one of the optional questions requests the URL of the bank's website. In other words the FDIC has a database that records a substantial number of domains currently being used by US banks and – key to the present study – it often records the domains being used by banks at the point at which they became, in the FDIC's jargon, "inactive".

We fetched a copy of the FDIC's database for 6 June 2013 and extracted from this the website URLs for banks that had closed on or after 1 July 2003 (i.e. over a period of almost ten years). We found that quite a number of these closed banks did not have a website URL entry. However, we located a third-party website (`http://banks.com-guide.org`) which appears to have populated its pages using FDIC data from 2007 – and this provided us a large number of website URLs that the current FDIC database was missing.

[1] Federal Deposit Insurance Corporation Institution Directory: `http://www2.fdic.gov/idasp/warp_download_all.asp`

In total, the FDIC database lists 3 181 banks that were merged or closed between 1 July 2003 and 6 June 2013 and, by the means just described, we were eventually able to obtain 2 302 URLs for their websites matching 2 393 banks (75% of the total).[2]

2.2 Methodology for Identifying Domain Usage

Following an initial sampling of websites, we identified the following categories for how closed bank domains are used:

1. operable bank-held website (old bank, redirect, or interstitial page);
2. domain parking pages with syndicated advertisements;
3. websites used to distribute malware;
4. other forms of reuse (e.g., blog spam, black-hat search-engine optimization);
5. inoperable websites (e.g., blank pages, misconfigured websites);
6. inactive domains (unregistered, or not resolving).

We visited all the closed banks' domains programmatically using a Selenium Firefox client, capturing a screenshot of the rendered website. We manually inspected each screenshot and assigned the domains to the appropriate category, a tedious but straightforward task. We identified malware-distributing websites by observing a blocking page set by the university firewall indicating that the website appears in a malware blacklist. We did not verify that the website still continued to distribute malware.

Inactive domains were identified by DNS lookup failures. WHOIS information was gathered for all domains and parsed using the DeftWhois Perl package.[3]

We also distinguished between domains still held by banks and those controlled by others. We used the following heuristics to confirm that a bank controls the domain:

1. any website whose screenshot is categorized as a bank *and* the domain has been continuously registered since before the bank closed;
2. any website that redirects to a currently open bank website URL that appears in the FDIC list;
3. any domain with WHOIS information indicating ownership by a bank.

Any domain satisfying one of these requirements is classified as being bank-held. This enables us to identify which inoperable domains are controlled by banks as opposed to third parties.

The first heuristic also enables us to identify the rare but insidious practice of impersonating a bank. Some websites look like a bank, but are in fact run by someone other than a bank. We can identify this by looking for bank-like websites where the domain dates from *after* the associated bank has already closed. In these cases, the closed bank allowed the domain registration to lapse, after which it is re-registered by a non-bank entity.

[2] The reason that we found fewer distinct URLs than banks is that some closed banks used the same web address (most likely as a result of merging).

[3] WHOIS Data Extracted from Templates: `http://www.deft-whois.org`

3 Empirical Analysis

We now discuss the data collected on ghost websites. First, in Section 3.1 we break down the prevalence of different forms of reuse. Then in Section 3.2 we present evidence that domains often progress from relatively innocuous forms of reuse to more insidious ones. In Section 3.3 we investigate how different characteristics such as bank size affect the likelihood of banks retaining control over domains. We then identify instances of bank impersonation on ghost domains in Section 3.4, followed by finding at-risk domains currently held by banks but susceptible to changing hands in Section 3.5.

3.1 How Closed Bank Websites are Used

2 393 banks operated 2 302 distinct websites at the time they were closed. The first question one might ask of these orphaned website domains is who controls them. Surprisingly, just 46% (1 059) of the domains are still held by banks. 45% (1 030) are used by others, while 9% of domains (213) are unregistered.

Figure 1 shows how the domains are currently being used. 30% of the closed bank domains are still used as bank websites, by redirecting to another bank's website, displaying an interstitial page, or hosting the old website. 37% of the domains have registered owners but are functionally inoperable.

The most popular repurposing of bank domains is for websites displaying the type of pay-per-click adverts typical of domain parking companies (426 domains, 18% of the total). Malware is distributed by 11 websites (0.5% of the total) and 110 domains (4.6%) have websites used for an assortment of other purposes.

Occasionally these domains are bought by legitimate services interested in the address (e.g., the social technology firm Gab Online registered `gab.com` after the Greater Atlantic Bank collapsed). More frequently, the new purpose bears little resemblance to the original bank. For example, a few websites sell pharmaceuticals or display pornographic content. Perhaps the most curious reuse is `bankoffriendship.com`, which displays a trailer and cast information for the German language film "Nullstex". This could be a symptom of dodgy search-engine optimization, which is a frequent form of reuse.

3.2 The Lifecycle of Closed Bank Websites

Given the many uses for closed bank domains, we now investigate the extent to which bank domains cycle through different phases of usage over time. Figure 2 plots the fraction of domains still held by a bank against year of closure.

We start by noting that when the bank originally registered their domain name they will have been able to choose to register for 1, 2, 5 or 10 years.[4] Thus when the bank closed it may have been several years until the next time at which

[4] Some top level domains do not allow very long registration periods for domain names, but `.com`, which dominated our results, certainly does.

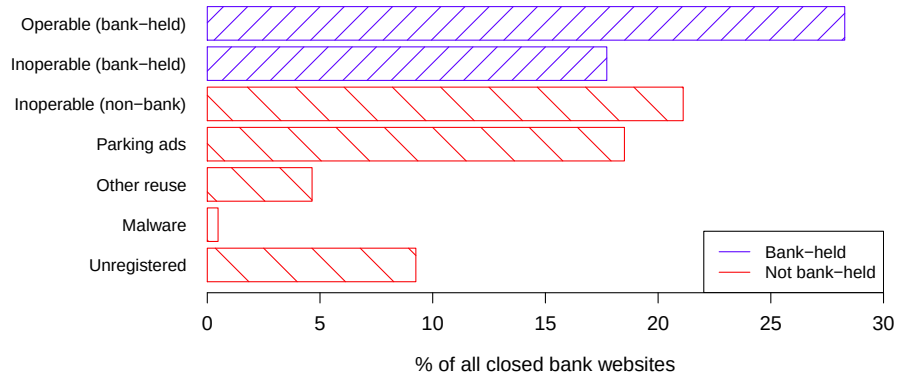

Fig. 1: Current use of domains from closed banks. Blue bars indicate bank-held (46%), red bars indicates non-bank holders (45%) or unregistered (9%).

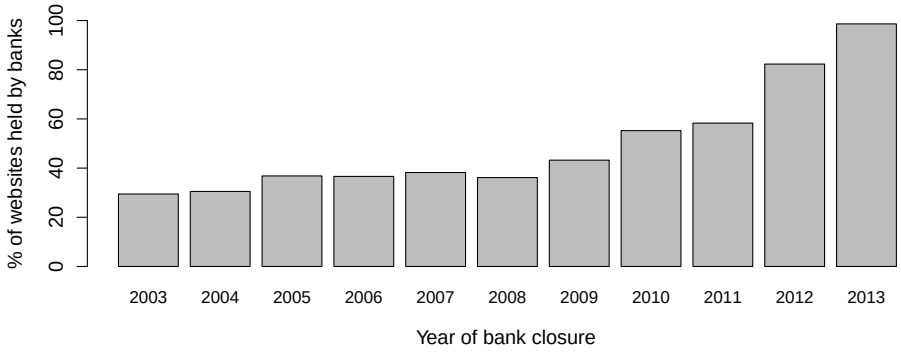

Fig. 2: Fraction of closed banks whose domains are still owned by a bank, by year of bank closure.

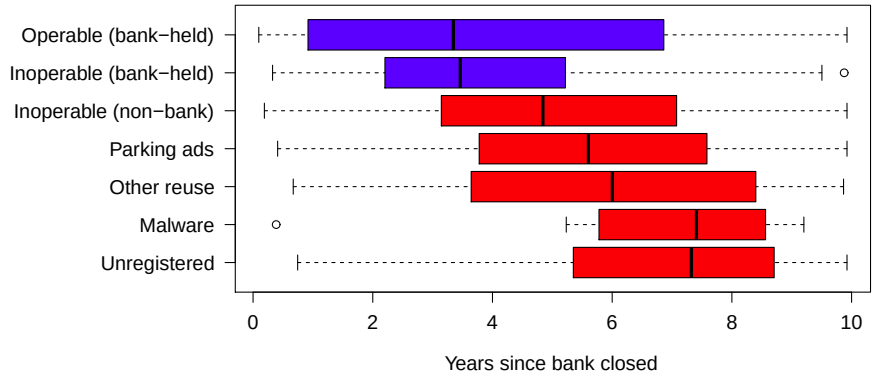

Fig. 3: Box plot of time since bank closed for different website categories.

a renewal decision had to be made. However, we found nothing in our data to suggest that long term registrations obscure shorter-term behavior.

We do find that the domains of recently closed banks are much more likely to remain with a bank – 99% of domains for banks closed in 2013 and 82% of domains for banks closed in 2012 remain under bank control. As time passes and domain registration renewals must be authorized and funded, perhaps several times, more domains fall out of the possession of a bank. The decline is steady, falling from 58% for banks shut in 2011 to 29% for banks closed in 2003.

For those domains that fall into the hands of others, how are they used and when does it happen? Figure 3 sheds some light. Domains that still point to banks have been closed for just under 4 years (median). By contrast, domains now used by parking companies have been closed for 5.5 years on average, with other forms of reuse falling slightly behind at 6 years. The eleven websites distributing malware belonged to banks that had closed 7.5 years prior on average, which is a similar time to domains that are simply unregistered today.

We note that there is substantial variation in the delays observed for each category. Some bank websites are abandoned and repurposed in less than a year, while some banks have held onto domains for more than a decade. But the median values do suggest that most abandoned domains are older, and they shift from use by parking companies to more sinister forms of reuse as time passes.

But are these differences statistically significant? The closure times are not normally distributed, since they are bounded on the left by zero. We also confirmed this using a Q-Q plot against a normal distribution. Hence, we use non-parametric tests to assess the differences in medians across categories.

We first run a Kruskal-Wallis test checking for differences among median values across all categories. This is highly significant, with a χ^2 value of 292.9. We therefore investigate pairwise differences in the closure times for each pair of categories to identify which differences are in fact significant. The results are given in Table 1.

	Not registered	Malware	Other reuse	Parking adverts	Inoperable (not bank)	Inoperable (bank-held)
Malware	✖					
Other reuse	✱	✖				
Parking adverts	✱✱✱	✖	✖			
Inop. (not bank)	✱✱✱	✖	✱	✱✱		
Inop. (bank)	✱✱✱	✱✱	✱✱✱	✱✱✱	✱✱✱	
Operable (bank)	✱✱✱	●	✱✱✱	✱✱✱	✱✱✱	✖

Table 1: Pairwise Wilcoxon rank-sum tests comparing differences in median closure times for different types of reuse. P-values are adjusted using the Holm method. The differences are statistically significant as follows. **Legend**: ✖: not significant, ● : $p < 0.10$, ✱: $p < 0.05$, ✱✱: $p < 0.01$, ✱✱✱: $p < 10^{-6}$.

We can see that the difference in closure times between domains held by banks that are still used for banking and all other categories is statistically significant, with the exception of inoperable domains held by banks. Domains used to display adverts and other reuses are associated with banks that have been closed significantly longer than domains still serving as banks or those that are now inoperable. However, unregistered domains have closed for the longest periods, and the difference between unregistered and both parking and reuse is statistically significant.

3.3 Characteristics Affecting Domain Reuse

We next examine whether certain attributes affect the chances the domains will fall from bank control. Characteristics studied include the time since closure, bank size and the reason the bank closed (e.g., collapse or voluntary merger). We first present descriptive statistics and then construct logistic regressions to more carefully identify factors that affect domain reuse.

Descriptive Statistics Table 2 shows the prevalence of different attributes for domains controlled by banks compared to those which are not. Differences in proportion are checked for statistical significance using χ^2 tests (those categories found to be significant are indicated by + and − signs in the table).

535 bank domains have been allowed to expire at some time after the bank closed. However, 326 of these have subsequently been 'resurrected', that is, re-registered and a new creation date has been recorded in the WHOIS.

The first grouping in Table 2 examines how resurrected domains are used. Only 0.7% of bank-held domains have been resurrected, compared with 30% of domains not held by banks. Once resurrected, very few domains lose their registration again (only 2%). Thus, we can safely conclude that the vast majority of domains abandoned by banks are no longer seen to be valuable for banking, and that non-bank entities are most likely to resurrect an abandoned domain.

We can also examine if any characteristics of the bank itself are associated with who ends up controlling the closed bank's domain. Larger banks tend to have greater IT resources, so they are less likely to inadvertently lose control over domain names. Smaller banks may have fewer resources, but their domains may also be less attractive for others to reuse since there would be less incoming traffic and fewer links to the old content.

The second grouping of rows in Table 2 uses the reported total deposits at closure as a measure of bank size. Indeed, large banks are more likely to hold onto their domains. When smaller banks close, their domains are more likely to be abandoned and end up unregistered than mid-sized bank domains.

Finally, we can examine the circumstances of why the bank closed to see if this affects how the domain is later used. Of the 2 394 closed banks, the vast majority shut as a result of a merger or acquisition. 79% merged or were acquired without requiring any financial assistance from federal regulators, while another 18% did so with assistance. 71 banks, 7% of the total, collapsed and were closed

	Bank-held			Not bank-held			Unregistered		
	#	%	Diff.?	#	%	Diff.?	#	%	Diff.?
Not resurrected	1 119	**99.3%**	(+)	739	**70.2%**	(−)	209	98.1%	
Resurrected	8	**0.7%**	(−)	314	**29.8%**	(+)	4	**1.9%**	(−)
Deposits < $100M	353	**31.4**	(−)	365	34.5%		146	**69.2%**	(+)
$100M < Dep. < $1Bn	622	55.4%		591	56.4%		62	**29.4%**	(−)
Deposits > $1Bn	148	**13.2%**	(+)	91	8.7%		3	**1.4%**	(−)
Collapsed	27	2.4%		36	3.4%		8	3.8%	
M/A with assistance	196	17.4%		226	**21.5%**	(+)	12	**5.6%**	(−)
M/A without assistance	904	80%		791	75.1%		193	90.6%	

Table 2: Comparison of characteristics of closed banks to post-closure website use. The first grouping compares websites that are 'resurrected' (i.e., the domain's creation date occurs after the bank closed) to those whose domains have not expired after the bank closed. The second grouping measures bank size in terms of deposits, and the third grouping examines why the bank closed (e.g., due to collapse, versus acquisition or merger made with or without federal assistance). Differences in proportion that are statistically significant at the 95% confidence interval according to a χ^2 test are indicated with a (+) or (−) sign.

by the FDIC. Banks that are merged or acquired with federal assistance (i.e., they were in financial trouble but not enough to lead to total collapse) are disproportionately likely to see their domains fall into the hands of non-banks. These domains are also less likely to be abandoned completely.

Logistic regressions We carry out two related logistic regressions to identify factors that may lead to the abandonment and repurposing of bank websites by others. In the first regression, we create a binary response variable for whether or not the bank relinquishes the domain. This includes domains that are used by others as well as those that remain unregistered.

Our first model takes the following form:

$$\log \frac{p_{abandoned}}{1 - p_{abandoned}} = c_0 + c_1 \log(\text{Deposits}) + c_2 \text{ Troubled} + c_3 \text{ Years closed} + \varepsilon$$

where the variables we examined were:

* **Abandoned:** Boolean response variable set to True if the bank no longer controls the domain (i.e., it is unregistered or not bank-held).
* **Deposits:** Deposits held by the bank when closed (in thousands of dollars).
* **Troubled:** Boolean variable set to True if the bank collapsed or was merged with FDIC assistance.
* **Years closed:** Years since the bank has closed.

Informed by the summary statistics just presented, we hypothesize that troubled banks and smaller banks (as measured by deposits) are more likely to aban-

Regression 1			Response variable: *Abandoned*	
	coef.	Odds Ratio	95% conf. int.	Significance
(Intercept)	0.58	1.79	(0.90,3.63)	-
log(Deposits)	-0.17	**0.84**	(0.80,0.89)	$p \ll 0.0001$
Troubled	0.87	**2.38**	(1.90,2.98)	$p \ll 0.0001$
Years closed	0.29	**1.33**	(1.29,1.39)	$p \ll 0.0001$
Model fit:	$\chi^2 = 322.8, p \ll 0.0001$			

Regression 2			Response variable: *Registered*	
	coef.	Odds Ratio	95% conf. int.	Significance
(Intercept)	-0.84	0.43	(0.13,1.38)	-
log(Deposits)	0.33	**1.39**	(1.27,1.53)	$p \ll 0.0001$
Troubled	0.73	**2.08**	(1.18,3.86)	$p = 0.0151$
Years closed	0.24	**0.79**	(0.73,0.85)	$p \ll 0.0001$
Model fit:	$\chi^2 = 120.7, p \ll 0.0001$			

Table 3: Tables of coefficients for logistic regressions.

don domains. We also anticipate that as more time passes following a bank's closure, the associated domain becomes more likely to fall outside its control.

Indeed, as shown in Table 3 (top), each of these hypotheses are confirmed. Every doubling of the size of deposits at the closed bank reduces the odds that the domain will be abandoned by 16%. For troubled banks, the odds of abandonment are increased by 138%. Finally, each additional year that the bank has been closed increases the odds that the domain will be abandoned by 33%.

We are also interested in finding out which domains that have been abandoned get repurposed by others. Consequently, we performed a second logistic regression on the 1 265 domains that banks no longer control:

$$\log \frac{p_{registered}}{1 - p_{registered}} = c_0 + c_1 \log(\text{Deposits}) + c_2 \text{ Troubled} + c_3 \text{ Years closed} + \varepsilon$$

For this regression, the binary response variable **Registered** is simply set to True if the abandoned domain is still registered.

Once again, as shown in Table 3 (bottom), each of the explanatory variables are statistically significant. However, this time the effects are different. In particular, the abandoned domains associated with closed banks having greater deposits are *more likely* to remain registered. As each year passes, the odds that a website outside of bank control will remain registered falls by 21%. Finally, abandoned domains of troubled banks face double the odds that they will be registered by others.

3.4 Identifying Bank Impersonation

While quite rare, an especially harmful form of closed bank domain reuse is to set up webpages that look like banks but are not in fact banks. We identified such websites by more closely inspecting all the resurrected domains that we had classified as being banks to determine whether their content was branded for an appropriate banking entity.

In all, we found just five dubious domains. Three (`rockbridgebank.com`, `securitystatebank.net` and the aforementioned `midvalleybank.com`) serve copies of the old bank website but have had links to other websites added, likely as part of some blackhat search-engine optimization scheme.

One bank, `plazabank.com`, is a false positive. After Plaza Bank of Texas was acquired, the address `plazabank.com` was allowed to expire. Plaza Bank of California and Nevada, which goes by the address `plazabank.net` according to the FDIC, resurrected `plazabank.com`, which now shows a copy of the content appearing on `plazabank.net`.

The fifth website, `townecenterbank.com`, is more of a head-scratcher. According to the WHOIS information, `townecenterbank.com` is registered to "Domain Listing Agent" and now redirects to `towncenterbank.net`, which is registered to "Town Center Bank". It is plausible that the unrelated Town Center Bank took over the domain after *Towne* Center Bank folded.

3.5 Identifying At-Risk Bank Websites

While the analysis so far has focused on the ways in which expired bank domains are already being reused, we can also identify *at-risk* websites that are more likely to fall from bank control in the future.

We consider a bank-controlled website to be at-risk if, according to the WHOIS record, the domain has not been updated since before the bank closed but has yet to expire. In this circumstance, the bank has not yet had to make a decision whether or not to renew the domain, if indeed they are fully aware that the domain is theirs to renew.

Of the 1 127 bank-controlled websites, 157 are at-risk of falling out of bank control. Figure 4 shows when the registration for these websites is set to expire. Between 30 and 40 websites will expire annually over the next three years. We anticipate that as further banks close, the number for the years 2016 and beyond will rise to the level of 2013–2015.

How many at-risk websites do we anticipate will fall from bank control? We know that the 970 websites for closed banks have been updated and remain held by banks, compared to 1 266 websites that banks no longer control. If the same fraction holds for the 157 at-risk domains that have not yet faced the option to renew, then we would expect that without any change in the approach taken by the banks then 57% of the at-risk domains will be taken over by non-banks.

We next discuss the policy options that might be considered for dealing with these and future at-risk domains.

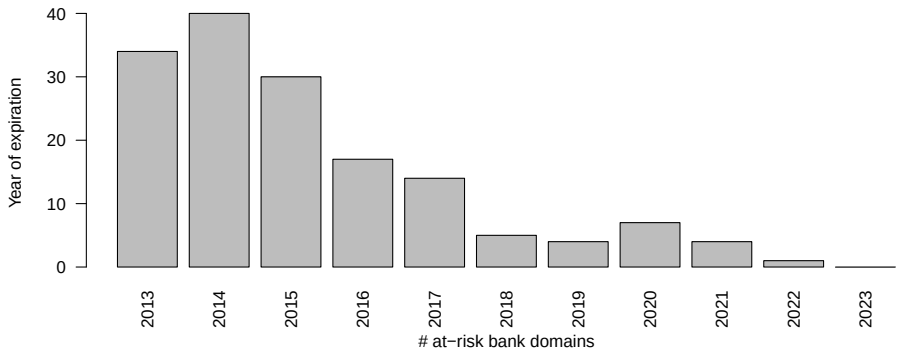

Fig. 4: Number of at-risk websites for closed banks set to expire each year.

4 Policy Options

The domains that were once used by banks are not alone in having a residual reputation that might be exploited once the original owner finds them to be surplus to requirement. It is possible to imagine scenarios in which the domains associated with newspapers, e-commerce or the provision of stock market prices might be used for nefarious purposes by a new owner.

There are also domains that were created solely to do harm, which need to be kept out of circulation for a considerable period. Botnet command and control (C&C) domains fall into this category. When malware infects a new machine it will contact the controlling system for instructions so that it can join in with botnet activities. The malware typically locates the controller by resolving a baked-in hostname. A key part of neutralizing this type of botnet is to prevent the hostname from resolving – usually by having the domain name suspended.

However, once the domain name expires (because the botnet operators are unlikely to renew it) then it becomes available for anyone to register. One of the authors of this paper purchased an old botnet C&C domain a year after it had been suspended and found that there were still around 5 000 malware infected machines attempting to make contact. A less civic-minded registrant could have easily resurrected the botnet.

Similarly, some iframe injection exploits in late 2012 were dealt with by taking down the websites hosting the malicious JavaScript. Unfortunately, the websites that had been compromised to add an iframe to fetch the injected code were not all cleaned up. In early October 2013 the domains came under the control of someone who once again supplied malicious JavaScript – and the original security problem had to be tackled for a second time.

We now review a range of mechanisms that might be adopted in order to address the problem of the control of domains that should not be available for just anyone to register for an extended period.

Permanent cancellation: The domain would be permanently canceled and would not be available for anyone to register ever again. This is obviously avoids any possible harm – but preventing all future use will often be overkill; and all sorts of complications would arise if, for example, a bank decided to resurrect a legacy brand and wanted to recover the domain that they used to own. It would also be hard to determine objective criteria for putting a domain name into this state, whether or not the previous owner agreed that it might be for the best.

Prepaid escrow: It could become a requirement for certain classes of domain name to be prepaid for many years into the future. A regulator such as FDIC could require the cost of future domain renewal and management to be escrowed as a prerequisite of operating a customer-facing website. The same effect would be achieved by requiring that banks make the FDIC the registrant of record rather than letting the domain expire – with the FDIC underwriting their costs from their general operating budget, or perhaps by a specific levy on active banks. This policy option would be almost impossible to operate outside of a statute-based regulatory regime, so it does not address maliciously registered domains.

Trusted repository: A neutral body could be created to hold relevant domains in trust and it would be excused annual payment for the domain registration. This body would decide, on the basis of expert analysis of the available evidence, when a domain could be returned to the general pool. Until that point it would be 'sinkholed' – accesses would be logged to assist the decision making process, and perhaps to assist in informing the owners of compromised websites and machines that they had a security problem that should be addressed. Once again, the problem would be to determine the criteria for putting domains into this state – an obvious abuse would be for brand owners to see this as a cheap way of parking domains. Finding some way of funding the necessary infrastructure and of obtaining expert advice would also be somewhat problematic.

Warning lock: Domains that were perceived to have a residual value could be specially tracked so that their imminent expiry triggered warnings to the community. It would then be necessary for public spirited organizations to step up and renew any domains that were not deemed safe to allow just anyone to renew. This policy option is essentially a distributed version of the trusted repository just discussed, and although it could be effective for some types of domain its impact is likely to be extremely patchy. It might be argued that the present ad hoc arrangements for sinkholing maliciously registered domains, operated by organizations such as Shadowserver and Team Cymru, serve as prototypes for this type of approach.

It is perhaps unlikely that the best remedy for preventing the creation of 'ghost' websites will be the same as the ideal solution for blocking the resurrection of

maliciously registered domains, so we do not propose a universally applicable approach. However, for bank websites we conclude that of the available options, prepaid escrow would be the most practical mechanism to adopt since it could easily be added to the winding-down process managed by FDIC.

Because no adequate policy is currently in place, we have elected to defensively register all unregistered bank domain names to prevent further abuse. We plan to reach out to the expired domain's associated bank, in the case of acquisitions, to determine whether they would like to assume control of the domain at no cost. This work is being funded as a research project for ICANN.

5 Related Work

There has been considerable empirical research investigating the nature of phishing attacks impersonating banks [6,7]. To our knowledge, there has been no prior work discussing the re-use of closed banking websites. However, several researchers have observed that spammers sometimes re-register expired domains in order to benefit from the reputation of the old domain [2,3,4]. For instance, Hao et al. found that spammers quickly register recently expired domains, much faster than non-spammers [4].

There has also been detailed research into 'typo-squatting', where domains are registered with similar names to popular websites [8]. The hope is that users will mistype URLs, reach the 'wrong' place and, by clicking on adverts for the 'real' site thereby make money for the domain registrants. In the present case, where there is no 'real' site anymore, someone using the domain names to catch traffic from people who had forgotten about the demise of their bank could only serve up adverts for generic banking or insurance products.

Kalafut et al. examine 'orphan' DNS servers, whose domains have (usually) been suspended due to malicious activity but remain in the DNS as authoritative for some domains [5]. They note that attackers could re-register these domains to take control of otherwise operational domains. This resembles our study in that websites could cause harm if brought back online, though in our study we consider legitimate, trusted resources (banks) instead of illicit websites.

There are counterbalances to the use of confusingly similar domain names and these might conceivably be used to tackle the reuse of domain names in a confusing context. Some jurisdictions have explicit legislation; in the US there is the Anticybersquatting Consumer Protection Act of 1999 (15 U.S.C. §1125(d)) and the Truth in Domain Names Act of 2003 (18 U.S.C. §2252B). Additionally, Uniform Dispute Resolution Procedures (UDRP) are operated by many of the domain registries. The UDRP process is a form of arbitration that allows complainants to recover domains from unworthy registrants [1] but there is no provision for third parties to initiate proceedings, and presumably a bank that has let domains lapse would have limited interest in expensive action under a UDRP regime.

6 Concluding Remarks

We have investigated what happens to domains that were once used for customer-facing banking websites after their owners change or disappear. By inspecting over 2 000 websites associated with banks that have closed in the last decade, we can provide insights drawn from a statistically robust dataset.

We find that while many websites initially remain in the hands of a bank, over time these websites tend to become inactive and their domains are frequently allowed to expire. Large banks tend to hold on for longer but even they frequently choose to relinquish control eventually. When domains do expire they are often quickly acquired by non-banks and repurposed. Logistic regressions have been presented to precisely quantify this behavior.

Often the domains of closed banks are used to serve advertising but other, more sinister, uses may occur. Most reuse is lawful, albeit ethically questionable, such as when advertisers trade on the residual reputation of collapsed institutions. However, older domains are occasionally used to serve malware. Furthermore, in a handful of cases we saw domains that were no longer owned by the original bank but they served up content that made them look as if the original bank was still operating.

What, if anything, should be done? We have examined various policy options that would ensure that the residual reputation of bank domains is not used outside the banking sector. While each approach has drawbacks, placing domains in prepaid escrow as part of FDIC's bank closure process seems most compelling.

Although this paper concentrates on banking domains as an exemplar of domains where controlling future ownership of the domains might reduce risk, we also drew attention to other classes of domain where there is a strong public interest in controlling registration, such as botnet C&C domains and maliciously registered malware exploit domains. Unfortunately, policy solutions for these latter types of domain are rather more limited. Nonetheless, the current solution of having public-spirited organizations hold on to them can already be seen to have occasional failures, so we must expect that many more ghosts will come back to haunt us in the future.

Acknowledgments

The authors thank the anonymous reviewers for their helpful feedback. This work was partially funded by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHSS&T/CSD) Broad Agency Announcement 11.02, the Government of Australia and SPAWAR Systems Center Pacific via contract number N66001-13-C-0131. Richard Clayton's initial contribution was made whilst he was collaborating with the National Physical Laboratory (NPL) under EPSRC Grant EP/H018298/1, "Internet Security". This paper represents the position of the authors and not that of the aforementioned agencies.

References

1. Warren B. Chik. Lord of your domain, but master of none: The need to harmonize and recalibrate the domain name regime of ownership and control. *Int J Law Info Tech*, 16(1):8–72, 2008.
2. Na Dai, Brian D. Davison, and Xiaoguang Qi. Looking into the past to better classify web spam. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, AIRWeb '09, pages 1–8, New York, NY, USA, 2009. ACM.
3. Zoltán Gyöngyi and Hector Garcia-Molina. Web spam taxonomy. In *AIRWeb*, pages 39–47, 2005.
4. Shuang Hao, Matthew Thomas, Vern Paxson, Nick Feamster, Christian Kreibich, Chris Grier, and Scott Hollenbeck. Understanding the domain registration behavior of spammers. *Proc. ACM SIGCOMM IMC*, 2013.
5. Andrew J. Kalafut, Minaxi Gupta, Christopher A. Cole, Lei Chen, and Nathan E. Myers. An empirical study of orphan dns servers in the internet. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, IMC '10, pages 308–314, New York, NY, USA, 2010. ACM.
6. Tyler Moore and Richard Clayton. Examining the impact of website take-down on phishing. In Lorrie Faith Cranor, editor, *eCrime Researchers Summit*, volume 269 of *ACM International Conference Proceeding Series*, pages 1–13. ACM, 2007.
7. Tyler Moore and Richard Clayton. The consequence of non-cooperation in the fight against phishing. In *Third APWG eCrime Researchers Summit*, Atlanta, GA, October 2008.
8. Tyler Moore and Benjamin Edelman. Measuring the perpetrators and funders of typosquatting. In Radu Sion, editor, *Financial Cryptography and Data Security*, volume 6052 of *Lecture Notes in Computer Science*, pages 175–191. Springer, 2010.