

Domain Footprinting for Web Applications and Web Services

Abstract

A wide array of services, from banking and finance transactions to auctions and ticket reservations, are being offered to customers online. This means that an Internet presence for companies may encompass several domains for each of the different services being offered online.

Performing web application or web services assessment with “zero” level knowledge for clients can be a daunting task for the web analyst. It is important to locate and footprint all critical domains running web applications or web services.

One of my previous papers discussed host-level footprinting to find applications pointing to specific IP addresses [<http://www.infosecwriters.com/texts.php?op=display&id=259>]. This paper focuses on domain footprinting and discusses a complete approach to identify and footprint all possible domains running web applications or web services.

Web applications are crawled by all popular search engines. Domains running web applications or web services may have some links that may have been cached and archived by these search engines. This considerably simplifies our task. In this paper, we demonstrate how advanced search options offered by search engines like *Google*, *A9*, *Yahoo*, *Alexa* and others can be leveraged to obtain critical information about domains.

Shreeraj Shah

Founder & Director

Co-Author: "Web Hacking: Attacks and Defense" (Addison Wesley, 2002) and published several advisories on security flaws.



net - square

<http://www.net-square.com>

shreeraj@net-square.com

[blog] <http://shreeraj.blogspot.com>

Table of Contents

ABSTRACT	1
TABLE OF CONTENTS	2
DOMAIN FOOTPRINTING METHODOLOGY: AN OVERVIEW	3
STEP 1: QUERYING “WHOIS”	4
STEP 2: FOOTPRINTING ALL DOMAINS.....	5
STEP 3: FOOTPRINTING CROSS-DOMAINS	6
STEP 4: LINKING “WHOIS” & DOMAINS	8
STEP 5: DOMAIN REFERENCING AND ANALYSIS	9
 SCREENSHOT OF wsPAWN / wsCHES:	10
 CONCLUSION:	10

Acknowledgement

Lyra Fernandes for her help on documentation.

Domain Footprinting Methodology: An overview

This methodology of harvesting domain information on domains running web applications and web services has been developed on the basis of domain information available in public databases. This methodology does not entail making requests to actual machines residing in the domain or IP address range, but rather, query public domains such as *whois databases*¹ or gather information collected by crawlers freely running for many popular web “*search engines*”.

For better understanding, we shall dissect the overall methodology into five steps. Shown below are the steps and objectives.

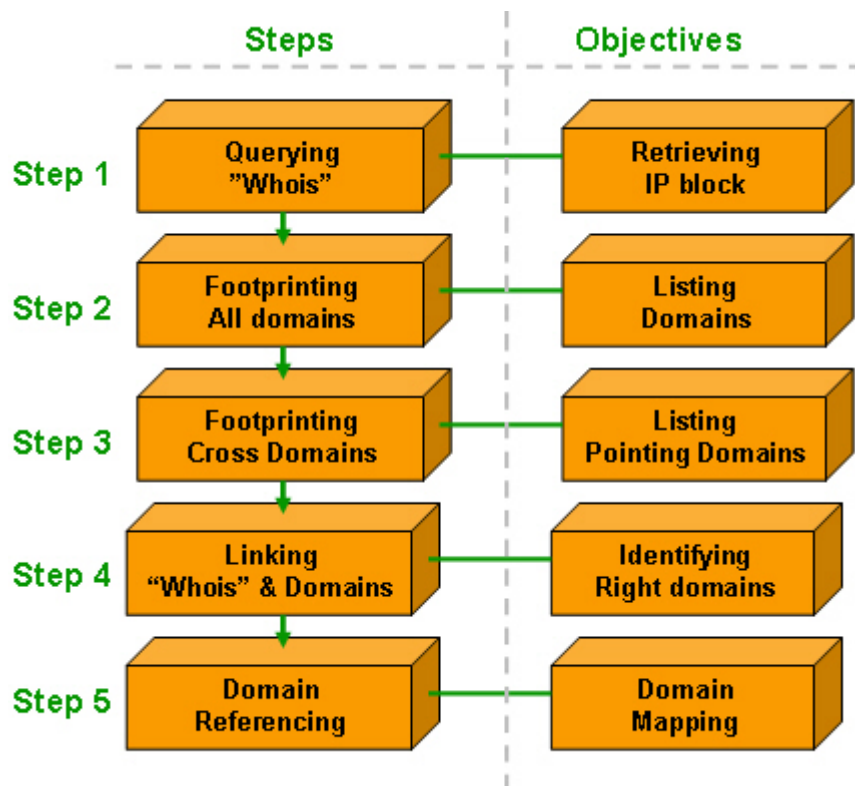


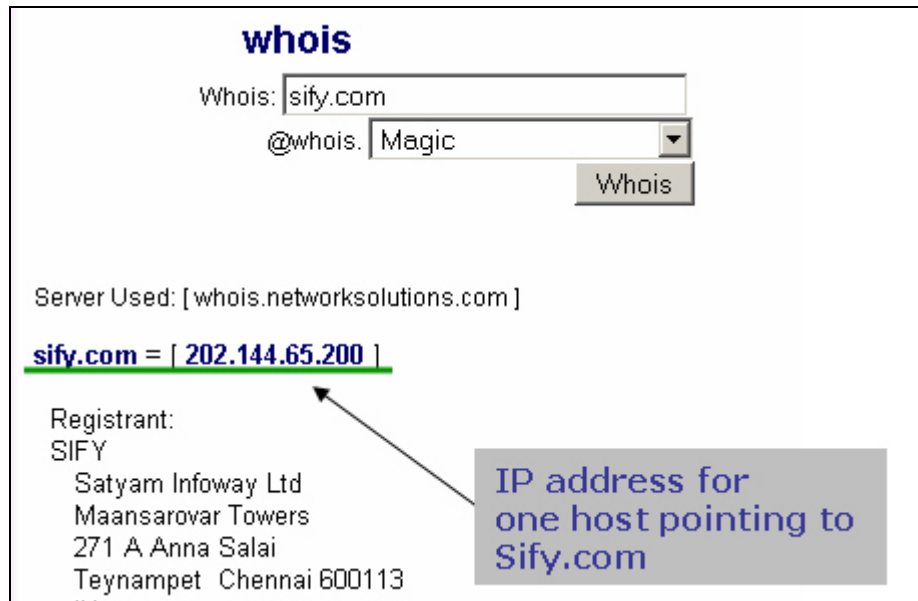
Figure 1: Methodology for domain footprinting

Let us discuss each step in detail with an example. For demonstration purposes, we shall limit our discussion and search to the domain “*sify.com*”.

¹ A searchable database that contains information about networks, networking organizations, domain names, and the contacts associated with them for the COM, NET, EDU, and ISO 3166 country code top-level domains.

Step 1: Querying “whois”

We begin with the *whois* database. If we query *whois* for this particular domain *sify.com*, we get its corresponding IP address. In this case we got the IP address 202.144.65.200 along with some other information such as nameserver, contact etc. The query is made to “samspade.org”, a nice site with a good collection of tools and a web interface. Take a look at the figure below:



whois

Whois:
@whois:

Server Used: [whois.networksolutions.com]

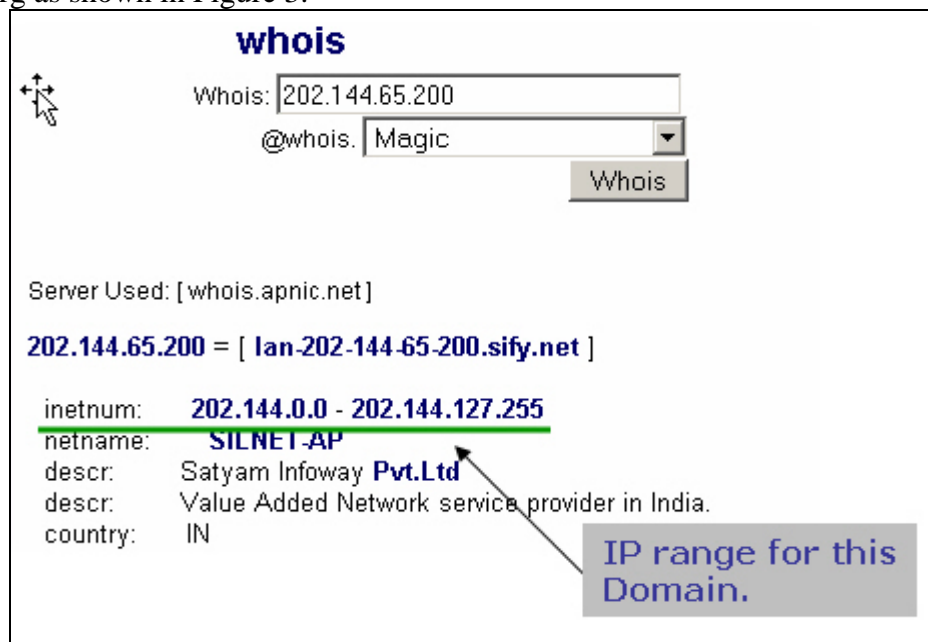
sify.com = [202.144.65.200]

Registrant:
SIFY
Satyam Infoway Ltd
Maansarovar Towers
271 A Anna Salai
Teynampet Chennai 600113

IP address for one host pointing to Sify.com

Figure 2: [URL] <http://www.samspade.org/t/whois?a=sify.com&server=magic>

Having obtained this information, we can now try to identify the IP block assigned to this particular domain. Once again this information can be retrieved from a *whois* record from samspade.org as shown in Figure 3.



whois

Whois:
@whois:

Server Used: [whois.apnic.net]

202.144.65.200 = [lan-202-144-65-200.sify.net]

inetnum: **202.144.0.0 - 202.144.127.255**
netname: **SILNET-AP**
descr: Satyam Infoway Pvt.Ltd
descr: Value Added Network service provider in India.
country: IN

IP range for this Domain.

Figure 3: [URL] <http://www.samspade.org/t/whois?a=202.144.65.200&server=auto>

We have achieved our objective which was to retrieve the IP address range for *sify.com*. This is the IP address block 202.144.0.0 to 202.144.127.255. We will reuse this information while doing analysis and linking in later sections.

Step 2: Footprinting all domains

The objective of this step is to identify all domains or hosts running with web applications on any domain. In our case, we limit our search to “sify.com”.

In our example, therefore, *xyz.sify.com* would refer to a host *xyz*, residing on *sify.com*. Similarly, *abc.news.sify.com* would refer to the domain *news.sify.com* residing on *sify.com*, running with web applications or web services or both on the host *abc*. To determine this information, we can use a search engine like *Google* that would fetch all available information from its database which is having presence on the web and serving with application/services.

Google yields astonishingly good search results when advanced google operators² are used. For instance, if we include [*site:domain*] in our query, Google restricts the results to just those websites available in the given domain.



NOTE: There is no space between "site:" and the domain.

A quick Google for “*site:sify.com*” results in the following query result:

The screenshot shows a Google search interface. The search bar contains the query "site:sify.com". Below the search bar, there are navigation tabs for "Web", "Images", "Groups", "News", and "more". The search results are displayed under the "Web" tab, showing a list of results for "site:sify.com". The first result is "Welcome to Sify Personal Homepages" with a snippet mentioning "Bawarchi.com" and "Posted On: 2005-03-29 23:14:24". The second result is "It's rain havoc time in City all over again - Sify.com" with a snippet mentioning "Sify.com - India's comprehensive breaking news site". The third result is "Olympic Fever - Sify.com" with a snippet mentioning "Sify Sports - India's comprehensive sports site".

Figure 4: Google query with *site:sify.com*

² query words that have special meaning to Google [<http://www.google.com/help/operators.html>]

From the retrieved results, we can see the following domains and hosts.

- food.sify.com
- blogs.sify.com
- sify.com

We can programmatically retrieve all domains and hosts for *sify.com* from Google. Shown below is the list of domains along with their IP addresses.

```
sify.com [202.144.65.200]
food.sify.com [210.210.109.4]
blogs.sify.com [202.144.65.200]
scores.sify.com [210.210.109.11]
discussions.sify.com [202.144.65.7]
www.sify.com [210.210.109.4]
promosnew.sify.com [202.144.65.16]
way2talk.sify.com [202.144.65.28]
customercare.sify.com [202.144.77.113]
login.sify.com [210.210.109.22]
alphacms.sify.com [202.144.65.54]
iw.sify.com [202.162.227.14]
tamil.sify.com [202.144.65.200]
search.sify.com [202.144.65.10]
headlines.sify.com [210.210.109.4]
www.tamil.sify.com [202.144.65.200]
groups.sify.com [210.210.109.7]
sitesearch.sify.com [202.144.65.193]
```

With this step, we have moved closer to achieving our final objective of retrieving all possible domains for *sify.com*. We can now move to the next step in our methodology.

Step 3: Footprinting cross-domains

Cross-domains are the list of domains which are pointing to each of the domains found in the previous step. What interests us is important information such as which domains are pointing to the domain “sify.com”

In the earlier section, we have used the Google query word *site* to locate websites in a given domain. Another of google’s query word is *link*. This lists web pages that have links to the specified webpage. Since we are looking for domains linked to *sify.com* we run a search using the query “*link:sify.com*”. Figure 5 shows the results of this query:



Figure 5: link:sify.com

The screenshot indicates that we were able to get some domains which are pointing to *sify.com*. Here is a small subset of domains which are pointing to “*sify.com*”

- [+]www.samachar.com [210.210.109.1]
- [+]www.khoj.com [210.210.109.22]
- [+]www.monsoonheritage.com [203.199.75.46]
- [+]www.sitagita.com [202.144.67.23]
- [+]hinduwebsite.com [208.56.95.81]
- [+]in.geocities.com [66.218.77.68]
- [+]www.sifyrealestate.com [202.144.65.16]
- [+]www.finance-informant.com [63.247.90.113]
- [+]retraite.blogspot.com [66.102.15.101]
- [+]www.omniglot.com [216.193.201.57]
- [+]www.paidcontent.org [209.59.174.167]
- [+]outsourcing.weblogsinc.com [206.252.155.9]
- [+]p.moreover.com [170.224.8.51]
- [+]www.searchmarketingindia.com [70.84.234.254]
- [+]lists.w3.org [128.30.52.16]
- [+]au.dir.yahoo.com [202.3.14.197]
- [+]www.out-law.com [195.188.8.75]
- [+]sifymax.com [202.144.65.200]
- [+]www.return2india.com [210.210.109.12]
- [+]food.sify.com [210.210.109.4]
- [+]tamil.sify.com [202.144.65.200]

We now have all domains which are a part of sify.com and other cross-domains pointing to each of these domains. This is a complex nested structure that we now have in place.

Continuing in this manner, we can get all domains linked to each of the hosts or domains found in the previous step along with their IP address.

Step 4: Linking “whois” & Domains

To simplify and reduce domain lists residing at sify.com we can compare resolved IP address to IP block assigned to sify. After comparing it here is the final list we got where [+] represents pointing domains.

```
sify.com
[+]sify.com [202.144.65.200]
[+]www.sitagita.com [202.144.67.23]
[+]www.sifyrealestate.com [202.144.65.16]
[+]sifymax.com [202.144.65.200]
[+]tamil.sify.com [202.144.65.200]
[+]www.search.sify.com [202.144.65.10]
[+]www.sify.com [202.144.65.18]
[+]login.sify.com [202.144.65.16]
[+]www.wizone.sify.com [202.144.65.18]
[+]blogs.sify.com [202.144.65.200]
[+]way2talk.sify.com [202.144.65.28]
food.sify.com
[+]sify.com [202.144.65.200]
[+]sifymax.com [202.144.65.200]
[+]tamil.sify.com [202.144.65.200]
[+]www.sify.com [202.144.65.18]
blogs.sify.com
[+]sify.com [202.144.65.200]
[+]blogs.sify.com [202.144.65.200]
scores.sify.com
[+]sify.com [202.144.65.200]
discussions.sify.com
[+]sify.com [202.144.65.200]
[+]sifymax.com [202.144.65.200]
[+]sify.in [202.144.65.200]
[+]www.sify.com [202.144.65.18]
www.sify.com
[+]sify.com [202.144.65.200]
[+]www.sitagita.com [202.144.67.23]
[+]www.sifyrealestate.com [202.144.65.16]
[+]sifymax.com [202.144.65.200]
[+]tamil.sify.com [202.144.65.200]
[+]www.search.sify.com [202.144.65.10]
[+]www.sify.com [202.144.65.18]
[+]login.sify.com [202.144.65.16]
[+]www.wizone.sify.com [202.144.65.18]
[+]blogs.sify.com [202.144.65.200]
[+]way2talk.sify.com [202.144.65.28]
```



```

promosnew.sify.com
way2talk.sify.com
  [+]way2talk.sify.com [202.144.65.28]
  [+]myaccount.way2talk.com [202.144.75.135]
customercare.sify.com
  [+]way2talk.sify.com [202.144.65.28]
  [+]www.sifycorp.com [202.144.65.28]
  [+]iway.com [202.144.75.101]
  [+]myaccount.way2talk.com [202.144.75.135]
  [+]www.sifygold.com [202.144.65.18]
login.sify.com
broadband.sify.com
  [+]sifymax.com [202.144.65.200]
  [+]sify.com [202.144.65.200]
tamil.sify.com
  [+]sify.com [202.144.65.200]
  [+]sifymax.com [202.144.65.200]
  [+]tamil.sify.com [202.144.65.200]
search.sify.com
  [+]search.sify.com [202.144.65.10]
  [+]www.search.sify.com [202.144.65.10]
www.tamil.sify.com
  [+]tamil.sify.com [202.144.65.200]
groups.sify.com
  [+]sifymax.com [202.144.65.200]
siterearch.sify.com
headlines.sify.com
iw.sify.com
ads.sify.com

```

Step 5: Domain referencing and analysis

Now that we have all possible domains residing on *sify*'s range, we can analyze linkages between these domains as shown below.

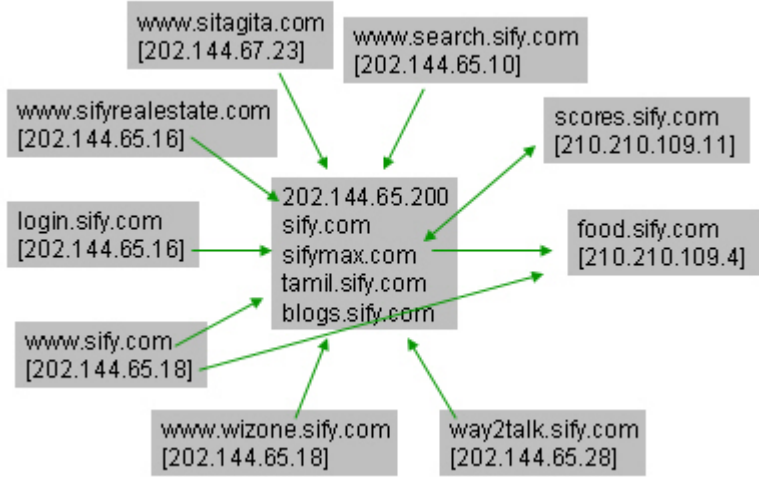


Figure 6: Snapshot of domain mapping

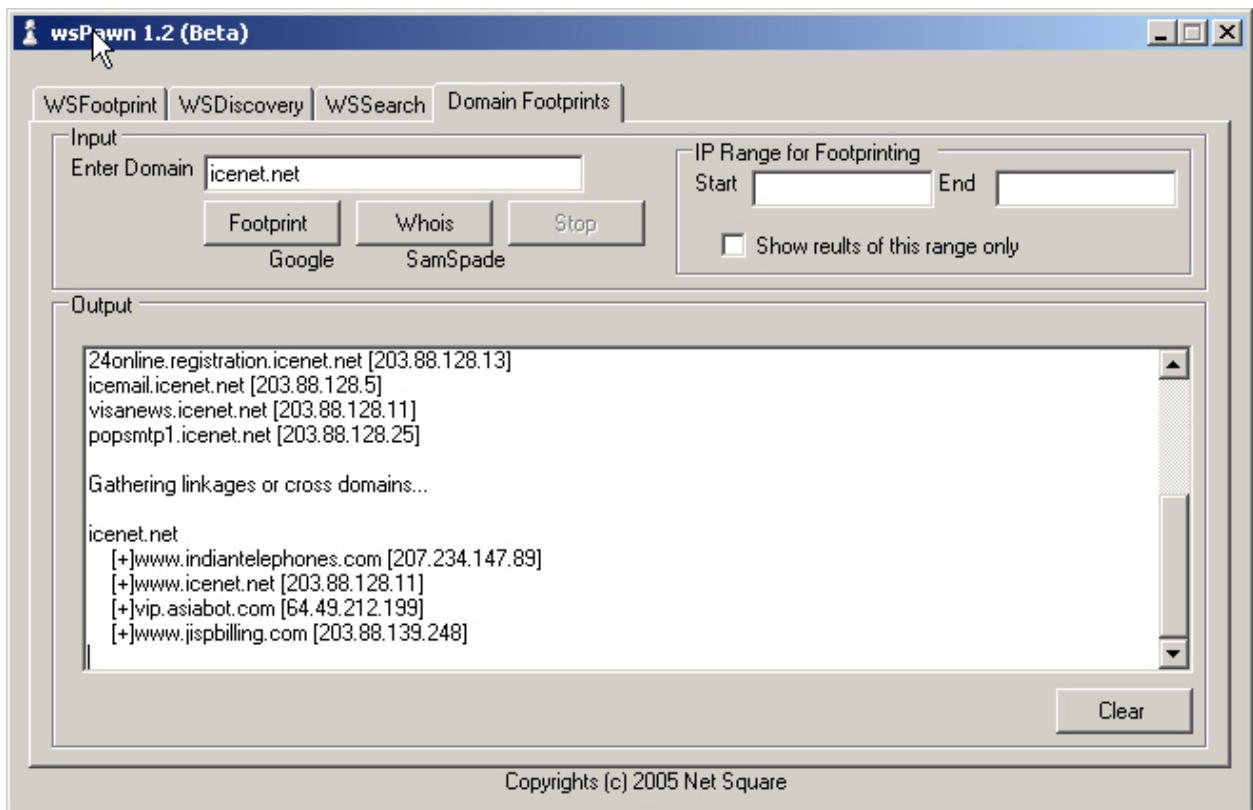
Some critical peripheral information that can be deduced from domain referencing and analysis includes –

1. Multiple web applications running on a single IP address.
2. Interlinking of these domains using cross-domain cookies to identify one large application running on all these hosts.
3. Using a single cookie across multiple web applications in cross-domains.
4. How one domain is linked to other domains in the same range.
5. Several new domains can be found in addition to the original. In our case, we began with *sify.com* and proceeded to footprint other related domains like *sifyrealestate.com*, *sifygold.com*, *ipay.com* and many others.



NOTE: The above methodology has been partially automated in beta 1.2 as part of the web services assessment toolkit called *wschess*. The toolkit is available at <http://net-square.com/wschess>

Screenshot:



Conclusion:

Domain footprinting is a reconnaissance methodology that allows domain data residing in *whois* databases to be gathered using search engine queries. It attempts to identify the Web-based applications residing in the company domain namespace.

One of the challenges of performing an external web application assessment and audit with only the web application URL or domain name, is to first identify the rest of the information piece by piece and then set about analyzing the vast amount of data and completing the assessment exercise. Simply put, this means determining primary domains and other related domains from a single domain name or web application URL using the methodology outlined in this paper.