

Appendices

Digital Dark Matter and the Economics of Apache

NOT FOR PUBLICATION

Appendix A: The Shape of the Server Economy

This section details the sample of web servers in this study by examining two questions. First, is this software widely used in the US economy? Is this evidence consistent with a core premise of this study, that server software plays an integral role in the Internet in many locations and at many companies and in many applications? Second, is there evidence that the use of Apache and Microsoft software differ? Do both Apache and Microsoft software appear in many of the same locations and firms?

How do we observe if the use of server software is widespread or concentrated in a small number of locations? To examine these questions, we correlated the IP addresses for our server software against lists of IP locations maintained by MaxMind.com.¹ We computed both state and county numbers. While no state dominates use of server software far out of proportion with its population, for the sake of brevity, we show only one table. This is for server software and market shares for counties in the US.

¹ <http://www.maxmind.com/app/geolitecity>, accessed December, 2011.

Table 1 shows the results for the top 25 counties, where the number of observations are large enough to lend confidence to the results. It lists the 25 counties with the most servers, ranking them in order. It also shows how large a share the servers in that county comprise of the total number of all servers. It then shows that county's rank in terms of Apache servers, and the share of the total of each of the three most common server software platforms, Apache, Microsoft and ngnix.

Table 1 – Top 25 Counties for server software use

Total server Rank	ST	County	Share of servers	# Servers	Rank Apache	Share of Apache	Share MS	Share ngnix
1	OH	Franklin	0.0642	4129	1	0.0775	0.0442	0.003
2	AZ	Maricopa	0.0636	4091	2	0.0576	0.0784	0.0109
3	CO	Arapahoe	0.0529	3407	4	0.0519	0.0588	0.0018
4	IL	Cook	0.0437	2813	3	0.0548	0.0221	0.0573
5	TX	Dallas	0.0432	2778	6	0.0416	0.0414	0.105
6	TX	Harris	0.0397	2553	5	0.0479	0.0252	0.0296
7	CA	Los Angeles	0.0376	2419	7	0.04	0.0316	0.0573
8	WA	King	0.0301	1937	8	0.0291	0.0255	0.1177
9	CA	Orange	0.0256	1649	10	0.0245	0.0216	0.1056
10	GA	Fulton	0.0234	1503	9	0.026	0.0196	0.0091
11	NY	Kings	0.0196	1261	11	0.0226	0.0144	0.0163
12	TX	Bexar	0.017	1097	14	0.0169	0.0148	0.0519
13	CA	Santa Clara	0.0165	1059	16	0.0141	0.0186	0.0465
14	PA	Allegheny	0.0152	980	12	0.0222	0.0035	0.0006
15	DE	New Castle	0.0132	848	30	0.0068	0.0255	0.0066
16	MA	Middlesex	0.0126	811	17	0.0138	0.0101	0.0151
17	MI	Ingham	0.012	774	13	0.0171	0.0033	0.0042
18	CA	San Bernardino	0.0109	699	15	0.0146	0.0039	0.0115
19	VA	Fairfax	0.0106	684	21	0.0105	0.0109	0.0097
20	MO	St. Louis	0.01	645	19	0.0107	0.0083	0.0151
21	PA	Lackawanna	0.0091	587	20	0.0105	0.0033	0.0507
22	FL	Broward	0.0089	575	23	0.0087	0.0099	0.003
23	CA	San Diego	0.0085	547	22	0.0089	0.0083	0.0012

24	UT	Utah	0.008	513	18	0.0119	0.0012	0
25	PA	Delaware	0.007	449	24	0.0082	0.0053	0

The data in Table 1 show that twenty five counties account for approximately 60% of servers in the United States. While there is some evidence of concentration of servers in large and populous counties, there is no evidence of concentration in a few locations, such as Seattle, Boston, New York, or Santa Clara. Server software is widely used and in many locations, symptomatic of its importance as an integral piece of the Internet.

Table 1 also shows the contribution of each county to the total share of Apache, Microsoft and ngnix use, and lists the ranking of the county in terms of Apache software share. These twenty five counties account for approximately 64% of Apache software, 50% of Microsoft server software, and 72% of ngnix software. Once again, this is evidence that server software is widely used and in many locations.

In addition, the ranking for use of Apache is very similar to the ranking for all server use. This is not surprising, since Apache comprises the largest component of total servers in use. Also contributing is another factor. Microsoft and Apache software do not differ tremendously in the extent of deployment within locations. The results for the top 100 counties are positively correlated. The number of servers deployed to Apache and Microsoft are correlated at .86. This last fact also reinforces the observation that arose in the data about counties about the lack of isolation. The market shares for Apache and Microsoft server software are roughly proportionate to one another in different counties.

The correlation with nginx is much lower, .48 for Apache and .42 for Microsoft. This is partial evidence that nginx differs from the other two.

Now we consider an additional question: Do the data show evidence of isolated use? The presence of such isolation would be evidence that the deployment of Apache and Microsoft software occurs in vastly different locations or companies, which would arise if these were not substitutes for one another.

To address this second set of questions we match the IP addresses with information about the top level and second level domain names. This is obtained using the nslookup tool, which is a standard feature of Linux. The following tables isolate attention to the three most common servers, Apache, Microsoft and nginx.

Table 2 shows the market share for different server software among different types of users, using the top-level domain names. This table shows server software long ago left its academic and government roots. The table shows that the majority of server software is used by organizations that register under TLD com, the most popular TLD, particularly for firms in the hosting business, who are very common users of this software. The second most common TLD is net, reflecting the importance of networking firms as users of server software in the US economy. The two originators of the Internet, the public military network (arpa) and the research network in universities (edu), account for only 9% of Apache and Microsoft server use.

Table 2 – Server use among Top Level Domain Names

Rank	TLD	Share of Apache	Share MS	Share nginx
1	com	0.5741	0.5131	0.7398
2	net	0.2320	0.2803	0.1714
3	arpa	0.0609	0.0488	0.0197
4	edu	0.0293	0.0434	0.0047
5	org	0.0254	0.0431	0.0236
6	info	0.0184	0.0065	0.0039

Table 3a, 3b, and 3c dig a bit deeper into the market shares for the top deployment of server, ranked by the contribution to the Apache total. As with Table 1, no single firm dominates the deployment of server software, albeit a few firms have especially large server farms. Each of the tables ranks the listing in terms of the organization’s contribution to the Apache total, and in each case it lists the top 15 organizations (16 in edu due to a tie).

Among the top 15 organizations there is only mild evidence of specialization. Many organizations deploy both Microsoft and Apache servers and many use nginx as well. Some firms only use Apache and nothing else, especially within com, but this is not found in net and edu. This appearance may be a partial artifact of showing only 15 organizations. The correlation between Apache and Microsoft server use for the top 100 users within the com group is .75, which is evidence that users of software from one source tend to be users of both, and roughly in similar scales.

Table 3a represents 40% of Apache server use, 17% of Microsoft use and 62% of nginx. That suggests two conclusions. First, nginx users are disproportionately drawn from Apache users. Second, it also shows that server use is quite spread out.

Table 3a – Server use among top 15 Second Level Domain Names among Com

Rank	SLD	Share of Apache	Share MS	Share nginx
1	theplanet	0.0855	0.0316	0.0343
2	softlayer	0.076	0.0616	0.1143
3	amazonaws	0.0559	0.0159	0.2069
4	dreamhost	0.0284	0	0.1154
5	cloud-ips	0.0244	0.0116	0.0766
6	bluehost	0.0229	0	0
7	ubiquityservers	0.0205	0.0057	0.0034
8	Rr	0.0161	0.0451	0.0011
9	myhostcenter	0.0134	0	0.0011
10	Linode	0.0132	0	0.0731
11	ecommerce	0.0132	0	0
12	mailengine1	0.0077	0	0
13	hostmonster	0.0073	0	0
14	nocdirect	0.007	0.0001	0
15	gridserver	0.0065	0	0

Tables 3b and 3c show the results from a similar exercise, now for net and edu. The results are very similar to those found in Table 3a.

The top 15 organizations among net users account for 52% of the deployed Apache software within that group, and 43% and 50% of Microsoft and nginx users within that group. Once again, there is little evidence of specialization. Among the top 100 users the correlations in the deployment of Apache and Microsoft server software is 78%.

Table 3b – Server use among top 15 Second Level Domain Names among Net

Rank	SLD	Share of Apache	Share MS	Share nginx
1	Secureserver	0.2964	0.1266	0.0143

2	Comcast	0.0441	0.0496	0.0095
3	Hostnoc	0.0393	0.0086	0.3429
4	comcastbusiness	0.0241	0.0918	0.0095
5	Verizon	0.0219	0.0476	0.0048
6	Sbcglobal	0.0169	0.0376	0.0048
7	Carpathiahost	0.0146	0.0015	0
8	Lstn	0.0102	0.0031	0.019
9	turnkeyinternet	0.0099	0.0007	0.0143
10	Cox	0.0092	0.0352	0.0048
11	Steadfastdns	0.0089	0.0011	0.0048
12	Securesites	0.0083	0.0013	0
13	Qwest	0.008	0.019	0
14	Scent	0.0075	0.0018	0.0333
15	Slicehost	0.0067	0	0.0429

The top 16 organizations among the edu users account for just 28% and 26% of Apache and Microsoft server software use respectively, reflecting the widespread use among many universities, albeit, universities are not a large fraction of server use in the United States. This group represents 50% of nginx use, however, once again, showing that nginx use is more concentrated, and largely drawn from large Apache users.

The evidence for specialization is stronger for this special group than for either com or net. Among the top 50 edu users the correlations in the deployment of Apache and Microsoft server software is only 17%. This arises because many universities tend to be small (the fiftieth ranked university in this data is CUNY and it has only 19 servers). Most universities tend to have large investments in either one or another server, albeit it often is no more than a few dozen.

Table 3c – Server use among top 15 Second Level Domain Names among Edu

Rank	SLD	Share of Apache	Share MS	Share nginx
------	-----	-----------------	----------	-------------

1	utpa	0.0549	0.0162	0
2	utexas	0.0224	0.0068	0
3	mit	0.0191	0.0014	0.1667
4	wisc	0.0179	0.0122	0
5	stanford	0.0168	0.0054	0.1667
6	psu	0.0157	0.1664	0
7	northwestern	0.0157	0.0027	0
8	columbia	0.0146	0.0041	0.1667
9	vt	0.0135	0.0041	0
10	umn	0.0135	0.0054	0
11	duke	0.0135	0.0014	0
12	umich	0.0123	0.0081	0
13	harvard	0.0123	0.0054	0
14	uchicago	0.0112	0.0027	0
15	ucsd	0.0112	0.0041	0
16	usc	0.0112	0.0149	0

In summary, server software is widely used in the US economy, as one would expect if it plays an integral role in the Internet in many locations and at many companies and in many applications. In addition, both Apache and Microsoft software appear in many of the same locations and at the same using organizations. The absence of evidence showing isolated use is consistent with the premise that the two are substitutes for one another.

Appendix B: Substitutability of Apache and IIS

The insights into the data in Appendix A do not end the discussion about substitution between Apache and IIS. When considering substitution, it is also important to compare the boundaries and functionality of the products.

When facing a decision to utilize a web server other than the Apache HTTP Server, businesses must consider a number of other costs associated with this substitution. Such costs often result from any switching between open and closed systems (Scacchi, 2002, Zhu, Kraemer, Gurbaxani, and Xu, 2006), but are especially relevant for a technology as important as a web server.

Although there are other free options for web servers, the Apache community is by far the largest community supporting any of the open source web servers (and one of the most widely used open source projects after Linux). Substituting a different open source web server for Apache HTTP Server alters the ecosystem that comes along with the software. A change in the software results in a loss of the large network of users and contributors who can be called upon for support. Additionally, because web server products exhibit network effects and Apache has already gained dominance in the web server market, most system engineers are only familiar with the Apache HTTP Server, and utilizing a different open source product can lead to a need to retrain engineers.

Another difference between the two is that IIS only runs on Microsoft Windows, while Apache HTTP Server can run on a variety of different operating systems, including Windows. This results in the added expense of purchasing the Windows operating system, as discussed above, to run IIS, whereas HTTP Server can be run on any operating system.

Compatibility with development languages is another area of differentiation for the two web servers. Active Server Pages on the .NET Framework (ASP.NET) is a web application framework produced by Microsoft that allows for the development of dynamic web sites and applications. It is integrated by design into IIS, whereas it can be run on HTTP Server via an add-on module called PHP: Hypertext Preprocessor (PHP), an open source web application framework used for developing dynamic web sites and applications. PHP is designed to be easily integrated with an Apache server, but it can also be run on IIS. ASP.NET and PHP have different pros and cons as well; however, the choice of a web server often depends on the preferences of the web application developer, with ASP.NET being optimized for IIS and PHP being optimized for Apache.

Additionally, IIS is generally considered easier to use due to its graphical interface when compared to the command line interface of Apache. However, the graphical interface also utilizes a greater deal of system resources than a command line interface, and therefore it is difficult to configure a Windows system running IIS to run in a very lean fashion, while it is very easy to do this for a Linux system running Apache. Therefore, a large percentage of Windows system resources and power are often devoted to tasks other than serving web pages, whereas a Linux/Apache system can be configured to spend the majority of resources and power serving web pages.