

;login:

THE MAGAZINE OF USENIX & SAGE

December 2001 • Volume 26 • Number 8

inside:

SYSADMIN

ISPadmin

By Robert Haskins

USENIX & SAGE

The Advanced Computing Systems Association &
The System Administrators Guild

ISPadmin

Web Hosting

Introduction

In this installment of ISPadmin, I examine how ISPs implement their Web infrastructure to support retail (tilde, or "~", accounts) and hosted domains. Web hosting is an integral part of most if not all ISPs, and many companies (Rackspace Managed Hosting and ServInt to name two) focus exclusively on Web hosting as their core business. Web hosting was the first application to be offered in the area now known as "application hosting."

It is worthwhile discussing the typical migration of a Web hosting customer at a retail ISP. A traditional dial ISP customer starts out buying a "standard" dialup account, which usually consists of the following:

- 1 dialup (PPP) account
- 1 mail (POP) account
- 1 Web hosting account (of the form *www.isp.net/~username*, commonly referred to as the "tilde" account)

Figure 1 illustrates a typical migration path of a Web hosting customer. The subscriber starts utilizing their standard PPP account, and the "tilde" Web account if they want to have some sort of a Web presence. If it is a business account, or a retail subscriber with more than a passing interest in hosting Web content, they will probably outgrow the "tilde" account and want to move to a "real" hosted domain (*www.mydomain.org*). The ISP needs to have a Web hosting offering or else they will lose the customer and associated revenue.

Once the hosted domain owner needs to sell something, they will want to have a shopping basket, secure site, credit card payment mechanism, etc. Once again, unless the ISP wants to risk losing the business, they need to make sure they can support electronic commerce.

The final step is the case where the Web site owner has so much traffic, it needs to be hosted on a dedicated server. To keep this customer and their business, the ISP must have a collocation (colo) offering.

In many senses, the Web hosting business is just an offshoot of the real estate business. In order to support a large number of domains, data center space is required. Of course, there is more to Web hosting than just real estate (for example, UPS and backup generator power, fire suppression, network connectivity, monitoring, etc.), but a large component of the cost will be the "bricks and mortar" and similar fixed non-IT-related components.

Web Hosting Infrastructure

In most cases, a small provider would likely have a very similar setup to a larger provider, but it might differ by (1) using a shared machine, in which the Web hosting machine might also perform other functions (such as mail, RADIUS and/or DNS) and (2) having fewer automated Web hosting-related provisioning and billing mechanisms.

by Robert Haskins

Robert Haskins is currently employed by WorldNET Internet Services, an ISP based in Norwood, MA. After many years of saying he wouldn't work for a telephone company, he is now affiliated with one.



rhaskins@usenix.org

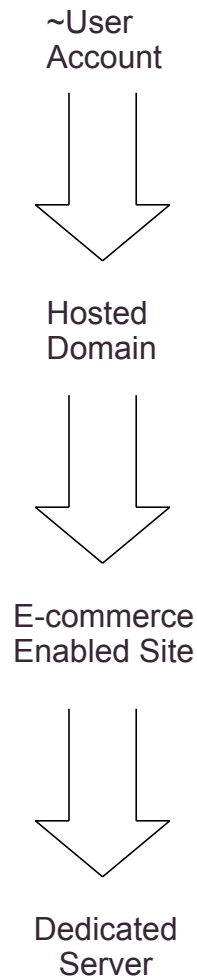


Figure 1

It is useful to have some idea of what it takes to manage a Web infrastructure for a dialup provider.

A large ISP would likely have machines dedicated to each specific type of hosting. For example, a machine or series of machines would be dedicated to each of the following functions: “tilde” accounts, domain hosting, and e-commerce-enabled domain hosting.

It is useful to have some idea of what it takes to manage a Web infrastructure for a dialup provider. Ziplink hosted approximately 5000 “tilde” accounts on a Sun Ultra 10 with 9GB mirrored disk drives. For the hosted domain side of the business, there were two Sun Ultra 10s with mirrored 18GB drives, each server hosting approximately 150 domains. The load, under normal circumstances (i.e., with no bad CGI scripts running and no extremely active pornography sites) was always less than 0.5. This infrastructure was run by the equivalent of half a full-time mid-level staff member. This might seem high, but Ziplink did not have much automation in the area of Web hosting, since the business plan focused on other areas (namely, wholesale dial up).

Web Server Software

APACHE

By far the most commonly utilized Web server, Apache has outstanding support for service providers and is very configurable. Some of the Apache modules I have found particularly useful at ISPs include: `mod_alias`, `mod_rewrite`, `mod_userdir`, `mod_spelling` (particularly useful for migrating a VMS or other non-case sensitive O/S Web server to UNIX), and `mod_vhost_alias`.

The references at the end of the article contain a link to the virtual domain support page for Apache.

MICROSOFT

Providing Microsoft (MS)-based services (such as Active Server Pages [ASP] or FrontPage Extensions) requires some extra effort if the ISP is a UNIX-centered shop. Some service providers may charge more for an NT-based infrastructure. UNIX-based service providers have two options. They can either (1) install a Microsoft-based infrastructure or (2) install additional software components to enable the required Microsoft functionality.

If option 1 is taken, a parallel MS Web hosting infrastructure must be deployed. This configuration would take the form of a Windows 2000 (Win2K) server running IIS (which now ships as part of Windows 2000). A stock Win2K server (aka NT) would be able to handle both ASP and FrontPage Server Extensions (FPSE) functionality. (Note that one must download and install the FPSE for the Microsoft platforms; Win2K and NT do not ship with FPSE). If the ISP has a back-end billing/provisioning system, then such system(s) must be modified to provision and bill this infrastructure.

If option 2 is taken, Apache (and most non-Microsoft-based Web servers) require additional components. These additional components can often be engineered on top of the provider’s existing Web infrastructure. In order to support FPSE, Microsoft has a software package which in its current version (FrontPage Server Extensions 2002 for UNIX) runs under Apache 1.3.19 and implements the server side of FrontPage. In order to implement MS ASP functionality under Apache, a package such as Sun Chili!Soft ASP must be implemented. There is also a Perl ASP (with Perl scripting only) implementation for Apache available called `Apache::ASP`. I do not have any direct experience with either of these packages.

The other popular Web servers out there (iPlanet and Zeus) are not often deployed at service providers, probably because they are both commercial products, unlike the public-domain Apache.

The Netcraft site has a good graphic showing the statistics of the various Web servers out on the Internet. I was particularly interested in the July 2001 report, which showed a rather large downturn (4.29%) in the number of Apache sites and a significant upturn (5.49%) in the number of Microsoft IIS installations. According to the Netcraft analysis, this is due to two large sites converting to Microsoft, and has been masking a larger trend away from Apache to IIS. The August 2001 report showed a much smaller increase in IIS deployments, so the July 2001 increase appears to be an anomaly.

Web Server Issues

There are many challenges facing service providers who host Web content. I will briefly discuss the issues and solutions surrounding each.

MANAGING DOMAINS

Managing a large number of domains is a problem. Most larger Web hosting companies have a file-naming scheme whereby the top-level namespace is broken down by the first letter of the domain. Thus, directory /www1/a would house all domains starting with “a”, directory /www1/b would house all domains starting with “b”, and so forth. Of course, the top-level directory name might indicate the machine name (www1, www2, etc.) for individual Web server machines). Also, Apache’s URL mapping scheme can be utilized to automatically redirect domains to the correct Web server domain content via a “global” command, without the need for specific per-domain configuration entries in the Apache configuration file.

EMAIL ALIASING

“Email aliasing” refers to the ability for email at a hosted domain (*info@mydomain.com*) to be forwarded to a “real” mailbox, for example *cust@isp.net*, where the customer actually picks up the mail through POP3, IMAP, Webmail, etc. Both the Sendmail and Postfix mail transport agents have good support for virtual mail configurations. The ISP usually adds an interface (typically Web based) for the customer support agent and/or the customers themselves to edit these mail mappings. These mappings must be forwarded to external accounts or the service provider’s “regular” dialup mail accounts. As a result, there is usually an interface between the provider’s billing system and mail infrastructure.

LEGAL LIABILITY OF HOSTED CONTENT

Most providers consider themselves “common carriers” and do not police the content placed on Web sites by their customers. Of course, people can and do complain. If it is an obvious copyright infringement (for example, posting illegal software or copyrighted mp3s) or something similar, the provider usually can and does take swift action without waiting for a court order. However, if the complaint is not as clear cut as that (for example, a site that parodies someone or something), the provider usually will wait for a properly executed court order before taking action. Most if not all service providers do not monitor content, since the provider would then be expected to monitor all content. Please be aware that policies in this area vary quite a bit from one provider to another so it is dangerous to make too many generalities!

Managing a large number of domains is a problem.

Most providers have automated mechanisms for domain registration as well as signup for Web hosting service.

SECURITY

Security of the server as well as security of customer data can be a problem in a shared Web server environment. Many hosting providers will not allow arbitrary customer-written/provided CGI scripts to be run on the machine. Of course, CGI routines are usually made available for standard functions like Web counters, comment sections, Weblogs, etc. However, any non-standard code has to be reviewed by staff prior to implementation. Also, external programs such as CGIWrap are used to help ensure CGI programs are run in a secure manner. The normal file access control mechanisms of the host (UNIX or NT/Win2K) are used as well.

LOGS

Most ISPs generate access and error logs for their customers. Apache has excellent support for automatically generating these logs without human intervention. The Apache configuration commands used to generate per domain logs are “ErrorLog” and “CustomLog” and appear under each VirtualHost section on a per-domain basis. Access to the logs is usually granted via the same FTP interface the customer uses for uploading their content.

BANDWIDTH

Most Web hosting plans include limits on the amount of bandwidth each customer’s pages generate. This bandwidth accounting is meant to limit the ISP’s exposure if a customer’s site should suddenly become very popular (for example, the customer begins to host pornography). If a customer’s site does become too much of a load on a shared server, the service provider will request that the customer move to a dedicated server. Of course, the ISP will charge additional money for that additional functionality to cover costs. These costs are: server hardware (if provided by ISP), staff time (to set up and move the domain) and transit (network bandwidth to the Internet).

BILLING INTEGRATION

A small ISP will usually do everything (setting up DNS, configuring Apache, etc.) by hand; as a result, no billing integration is required or possible. Billing for a larger ISP or Web hosting provider would be much more integrated, as Web hosting might be a larger part of the ISP’s business. Most providers have automated mechanisms for domain registration as well as signup for Web hosting service. This would require an automated interface into the provider’s billing system. How this interface is achieved would be a function of the billing system as well as the provider’s business model.

E-COMMERCE

For the purposes of this article, e-commerce (electronic commerce) includes: a shopping basket, managing an inventory, and processing credit cards. The ISP that wants to retain its customers through the full life cycle must have e-commerce capability. Because of the financial risks something like this poses, e-commerce could be farmed out to a third party or could be done in-house via commercial software or open source software.

Many third-party providers have a complete e-commerce package. In addition, several business to business (B2B) and business to consumer (B2C) solutions exist for ISPs. The most commonly implemented open source shopping basket and inventory management software is RedHat Interchange (formerly Akopia Interchange). Interchange is a full-featured B2C software application written in Perl and is widely used. Credit

card processing is dictated by the kind of e-commerce software the ISP is using. Red-Hat Interchange, for example, supports many credit card payment gateways, including Cybercash/Verisign (Verisign recently acquired Cybercash).

Next time I'll take a look at how ISPs design their backbone networks. In the meantime, if you have a question about ISPs, wondered why something was done a particular way, I'd love to hear from you!

I'd like to thank Vinny Bono of GlobalNAPS for his input.

References

Apache modules: <http://httpd.apache.org/docs/mod/index.html>

Apache Virtual Host documentation: <http://httpd.apache.org/docs/vhosts/index.html>

Apache Web server start page: <http://httpd.apache.org>

Apache::ASP: <http://www.apache-asp.org>

CGIWrap: <http://cgiwrap.unixtools.org>

Cybercash/Verisign: <http://www.verisign.com>

Digex: <http://www.digex.com>

iPlanet: http://www.ipplanet.com/products/ipplanet_Web_enterprise/home_2_1_1m.html

Microsoft FrontPage Server Extensions download: <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnservext/html/fpovrw.asp>

Microsoft FrontPage: <http://www.microsoft.com/frontpage>

Microsoft IIS: <http://www.microsoft.com/iis>

Microsoft Web page: <http://www.microsoft.com>

Netcraft Web Server Survey: <http://www.netcraft.com/survey>

Perl: <http://www.perl.com>

Postfix: <http://www.postfix.org>

Rackspace Managed Hosting: <http://www.rackspace.com>

RedHat Interchange: <http://interchange.redhat.com>

Sendmail virtual domain documentation: <http://www.sendmail.org/virtual-hosting.html>

ServInt: <https://www.servintservers.com>

Sun Chili!Soft ASP: <http://www.chilisoft.com>

Zeus: <http://www.zeus.co.uk>