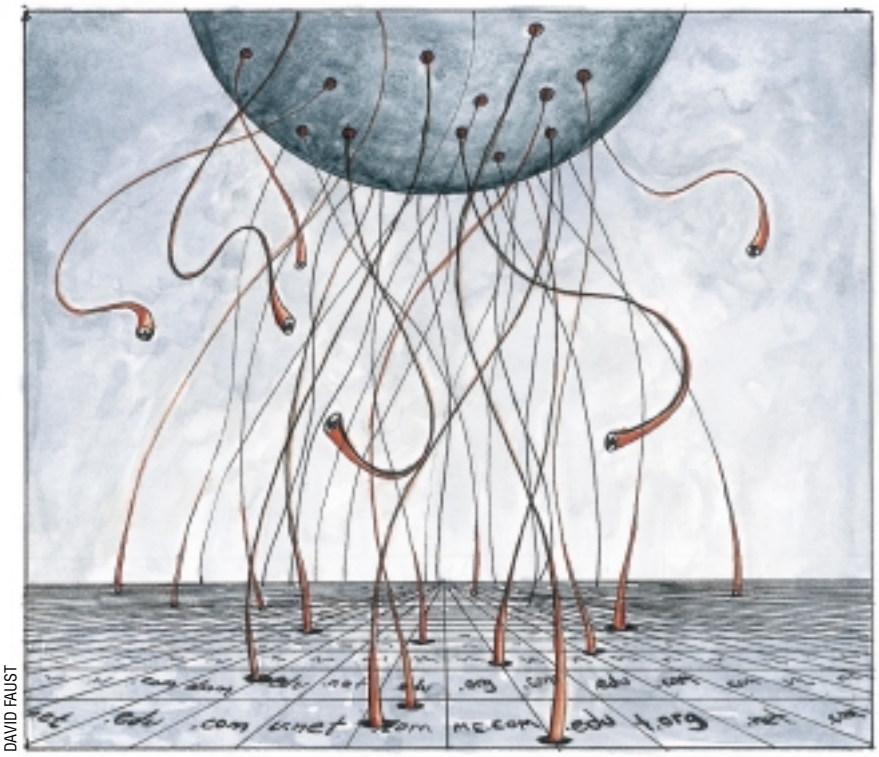
by Peter Collinson, **Hillside Systems**



DAVID FAUST

Domain Names

Many people find it odd that there is no one "in charge" of the Internet. Recently, I heard a radio discussion on the BBC World Service in which the speaker (whose name I didn't catch) said that human society often has complex structures unregulated by officialdom. His primary example was the food supply to New York City. Nobody sits down to ensure that food arrives in any large city in sufficient quantities to feed its population, it just happens. Many different organizations make it happen, each operating for its own benefit. I am sure that it's possible to come up with other examples. The Internet was bootstrapped by the U.S. Government but now has a life of its own, run by many separate organizations whose self interest keeps the whole thing operating.

One area of huge expansion in recent times is the registration and use of domain names. Everybody with a Web site is encouraged to have at least one domain name. Most of the sites that sell them actively encourage you to buy more than one. Since humans prefer to deal with names rather than numbers, the basic function of the name is to act as a human readable tag that addresses a specific machine. The mapping of names is managed by the DNS (the Domain Name System or Service) a distributed database that underpins the Internet.

The familiar dot-separated naming mechanism provides us with a way of constructing unique names for a particular machine and is used to provide us with a way of walking through the name space to find the address of that machine. For example, let's look at the name `frodo.cpg.com`. It is navigated from the right-hand end. We first find the name server that knows about the "top-level" domain `com`, then in that database find the server for the `cpg` domain, then finally we look for `frodo`.

Initially, various "top-level" domains were established, `.com`, `.org`, `.edu`, `.net`, `.mil` and `.gov`. This initial set is currently being extended. Countries outside the U.S. tended to use their standard two-letter country code as their top-level domain.

A key idea of the DNS is that of *delegation*. Once a domain is established, such as `.uk`, then the names within that domain can be run by an organization politically, geographically or economically closer to the users of the domain. When the `.uk` domain was established, the powers-that-be, who were academics, maintained and ran the machines that supported the `.uk` name space. They split the `.uk` name space into `.ac.uk` for universities, `.co.uk` for companies and `.gov.uk` for government.

Some years ago, the whole U.K. operation was dumped in the lap of a not-for-profit company, Nominet Ltd. Some parts of the name space were delegated to other organizations, for example, `.ac.uk` is now run by part of the academic infrastructure. Nominet runs some parts itself. My domain `hillside.co.uk`, for example, is registered with Nominet. In turn, they delegate the

running of my domain to me, so I can create names in it as I wish by altering files on my machines.

The story in the U.S. is similar, as name management moved from being run by volunteers and funded by government research budgets into the hands of a single commercial registrar. It moved on from there in late 1998. The general management of names on the Internet as a whole is now run by a not-for-profit corporation called ICANN (Internet Corporation for Assigned Names and Numbers). The best way to explain this organization is to steal a paragraph from their Web site:

"Formed in October 1998, the Internet Corporation for Assigned Names and Numbers (ICANN) is a non-profit, private-sector corporation formed by a broad coalition of the Internet's business, technical, academic, and user communities. ICANN has been recognized by the U.S. and other governments as the global consensus entity to co-ordinate the technical management of the Internet's domain name system, the allocation of IP address space, the assignment of protocol parameters, and the management of the root server system."

The "root server system" is the set of machines that manage the top-level domains. ICANN sees itself as a technical coordinating body; its job is not to "run the Internet." ICANN has also been responsible for expanding the number of registrars from one–Network Solutions–to many, each able to create domains in the top-level name spaces. Some new top-level domains: `.biz` and `.info` are being created to expand the global name space.

Confusion

Although the people who manage some domains have established rules for who can register in that domain, names have mostly been allocated on a first-come, first-serve basis. While this seems an equitable policy, it's certainly fostered confusion and obfuscation. It's also lead to quite a bit of annoyance in the business world, where names are money.

A good example of confusion is illustrated by the domain `whitehouse.gov`. This, as you might guess, is the official Internet address of the building in Washington, D.C.–the Executive Office of the President of the United States. The Web site found at this address contains interesting facts about the latest incumbents: Spot, Barney, India the cat, and a few humans whose names you might know.

Well, when looking for the White House, you might think to type the address `whitehouse.com`, perhaps by accident, or in complete innocence. If so, you will find a brightly lit site containing a different set of incumbents, having names like Mandi, Celeste and Sandy. If the truth be told, you don't get to learn much about them from the site. You are simply presented with detailed photos of their particular physical attributes and some statistical information about them. You will have no doubt that this is certainly not the official site of government.

It's perhaps more confusing to type `whitehouse.net`, which is certainly another plausible address for that same building. Now, this looks very similar to the `.gov` site and is executed technically with more skill. It seems to have several different front pages (use Reload on your browser). A closer inspection shows this `.net` address hosts a spoof of the "real" `.gov` site.

At the time I am writing, the `whitehouse.net` site contains a page of reader's responses that displays a letter from someone complaining about the `.com` site as if Spot, Barney and India could do something about it. The letter illustrates the confusion. The author is concerned that the soft porn site is only three characters away. She is worried about allowing her kids to see the images by accident, and this is a common concern. She can, if she wishes, subscribe to a Net monitoring service that will filter out this site for her.

Of course, she wants to ban the site. Our joint puritan heritage is strong on making a lot of noise about banning things. The U.S. is spared some of the furor due to the strong sense on the freedom of speech that is embedded in the psyche of the population, which in turn derives from U.S. constitutional rights. Here, in the U.K., the "I don't like it and don't want you to have it" lobby is often loud on a wide range of topics. After all, the Internet is currently a tag word that sells papers. Incidentally, unclothed well-endowed females are deemed to sell papers too. It's certainly the case that U.K. children can see images of unclothed female torsos in at least two national tabloid newspapers, should they be interested.

Regulation

Please don't get me wrong. I am not arguing for regulation of anything. I don't think someone should be able to sit in judgment and say that `whitehouse.com` or `whitehouse.net` should not exist. Politicians are all trainee comedians, and poking the fickle finger of fun at them is pretty much all they seem to be good for. They are so boringly bland, and some laughter at their expense may help us all feel somewhat less scared about what they can do to us, should they be mindful.

However, the confusion engendered by the free-for-all with domain name registration can be considerably more meaningful for companies, large and small. Companies all spend money on promoting their name or trademarks, and we have national and international systems in place predating the Internet by some years that are supposed to protect that investment. Name protection is generally a good thing as it mostly allows us to be certain when we buy a branded product, we are getting what we have paid for.

Many companies have taken to registering all possible permutations and combinations of their names in as many top-level domains they think fit. The intention is to prevent clone sites from being created using the names. There are people out there who will register names to annoy certain companies, people who register names that are close to a typo or a common misspelling of other names (see `http://www.macdonalds.com`), and people who will register names as an investment. They hope that the company will come along, want the name and pay through the nose for the privilege.

Some things are being done about all these problems. The new `.biz` domain, for example, is starting life allowing for trademark registration. In the U.K., Nominet only allows the registered company name to be used in names in the `.ltd.uk` and `.plc.uk` domains. Since we have a national registration of company names, this guarantees uniqueness of names and ensures that they will be used appropriately.

What is What?

How can you determine whether a site you are looking at is run by the organization that purports to own it? Here's a related question: How can I find a contact for a particular domain?

Most of the level domain administrators run a `whois` server to provide information about registrants of names in their domain. The problem is that these servers only contain information about the domain they support and there is no way of asking one server to find another. So your first job is to find the `whois` server for the top-level domain you are seeking.

If the domain you are seeking is in `.com`, `.net` or `.org`, then the `whois` server is `whois.internic.net`. If it's another domain then you may have trouble guessing. There are now several Web sites that will search the appropriate `whois` server for a domain name. On Webmagic's site `http://www.webmagic.com/whois/index.html`, you'll find a form where you can type the domain name, select the top-level domain and the system will interrogate the correct server for a response. It will tell you whether the name is in use and give you the option of seeing the information the server contains. The site `http://www.alldomains.com` does a similar job. Run by the same company, `http://www.allwhois.com` searches the server and shows you its output directly.

I have to confess to being nervous about using Web-based search systems when looking for a new name to be registered for a client. It's entirely possible for the domain names that you type into these forms to be logged, and a pre-emptive registration can then be done, perhaps before you can get back to your client and negotiate about whether they want the name or not. I am not accusing the sites mentioned above of doing this, I am merely stating the possibility.

I prefer to use the `whois` client that's found on my UNIX system to interrogate a server. If you type:

```
$ whois cpg.com
```

you'll get output shown in Figure 1.

The `whois` command interrogates a default `whois` server, and you may have an old and defunct server compiled into the program. You can give the program an explicit server to talk to by using the `-h` option. Your `whois` client will need to connect to `whois.internic.net` to obtain the information above.

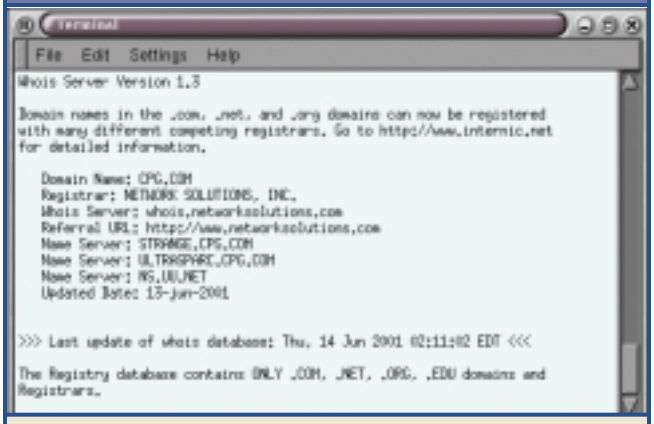
In fact, for the U.S. domains in the default server, you have to give a second command to find the full information about that domain. Take the server name from the `Whois Server:` line in the initial output shown in Figure 1 and type:

```
$ whois -h whois.networksolutions.com cpg.com
```

and you get the full output for the domain. You can obtain several names and addresses: the owner of the domain and the addresses of technical, administrative and billing contacts. The record will also give the names and IP addresses of the name servers that support the domain.

While I was doing research for this article, `whois.internic.net` disappeared from the Internet for a bit. I logged into a friend's FreeBSD machine (that happens to be in California) to see whether it was my ISP that was at fault. It wasn't. However, I found that the version of `whois` on the FreeBSD system has been updated to cope somewhat better with the current situation. It will automatically perform the second `whois` lookup for you.

Figure 1. *Whois* server



Information can include names, addresses and domain owner.

Also, and this is a huge step forward, the FreeBSD `whois` automatically finds the appropriate `whois` server for the domain you are seeking. A special `whois` server run by Centergate Research Group, LLC, makes this work. To look up a U.K. domain name, make a query like:

```
$ whois -h uk.whois-servers.net hillside.co.uk
```

To create a query using `whois-servers.net` take the top-level domain and prepend it to the server name. If you type this command, you will be given the information for my domain that's presented to the world from the main UK `whois` server at `whois.nic.uk`. You'll find that the information is different from that given to you by U.S. servers. You'll only get a registrant name, the company that registered the name in the form of a tag, and the name server information. It's quite typical that different `whois` servers will present different information.

It's simple to create a script that will do the work for you and use the older `whois` client that should be available on your machine:

```
#!/bin/sh
cmd=/usr/bin/whois
if [ $# = 1 ]
then
 # get the last component of the name
 tld=`expr $1 : '^.*\.\([a-zA-Z]*\)$'`
 # run the whois command
 $cmd -h $tld.whois-servers.net $1
else
 # call the real whois if we have more than
 # one argument
 $cmd "$@"
fi
```

I want to put this in my private `bin` directory and call it `whois`, so I arrange that if the script is called with more than one argument, I will call the standard command with those arguments. If I have one argument, then I am assuming that it's a domain name. This is not quite complete because `whois` servers can contain information other than just domain names, and I am neglecting the possibility of this.

If I have one argument, say `cpg.com`, then I want to pull off the last component of the name, `com`, to prepend to the server name. I've chosen to do this with the pattern matching capabilities of the `expr` command. There are many other ways. The `expr` command is run inside back quotes so that its output is captured and placed in the shell variable `tld`. There are three arguments to `expr`. The `$1` replaced by the first argument to the shell script, the colon says "do a pattern match that matches the regular expression in the third argument against the first argument." There's some magic here that lets us pull out part of the source string and output it. Let's pull that regular expression apart (see Table 1).

So if I am matching against `cpg.com`, then the match string will pick out `com` and `expr` will output that string. It will be loaded into the `tld` variable and injected into the command on the next line. This regular expression could be made more robust and it is possible to test the value of `tld` to see if anything has been matched.

How Useful is *whois*?

It all depends on the amount of information you can retrieve. The U.K. server hardly gives you any information about the name registrant. On talking to Nominet, I find that they have to operate within the provisions of the Data Protection Act in the EU and are constrained by that law. However, they do offer a free dispute resolution service and will help to make contact between people should that be needed.

In the U.S., the `whois` servers supply more information. This can be a good thing, and it can help you obtain a contact for whoever owns the name. It can also be a bad thing. There's a huge temptation for people to use the e-mail information the servers contain to send bulk e-mail or even one-off advertising. Many `whois` servers will transmit a disclaimer that is often ignored. It's hard to prove the exact source used by the junk e-mailer to get their information.

You can certainly use the `whois` information to find some of the truth about the three White House sites. I say *some* of the truth, because it seems that none of the address information presented by `whois` is validated by the name registrars. They have historically depended on people supplying correct information. First, the information can be stale. It could have been hanging around for a number of years. Second, it's possible to register with a legitimate address, pay the bill, which acts as some form of validation, and then change the stored information to obfuscate your origin.

I had an incident recently where I was trying to find out who had registered a name that was essentially the same as the company name of one of my clients. The `whois` information for the domain was entirely bogus, and I used the DNS to find a contact in the reseller company who hosted the domain for the DNS. The person there did reply to my e-mail query, but said I needed a court order to obtain the contact information I was looking for.

After some advice from the people who had actually registered the name, and were directly responsible to ICANN, I complained to their compliance department. The name owner was forced to supply an address, but I don't believe that it's real. It's uncheckable since it's in Kuwait. The phone numbers are certainly bogus. It's all a dead end, which is unsatisfactory.

The real problem with the privatization of the domain registration business is that the actual registration has been placed in the hands of anyone with a Web browser and there are no checks on the data that is supplied. ICANN may regulate its clients, but actual name registration is often being done by organizations that are three (or sometimes four) steps away from ICANN. Pricing is so low there is no money to provide for support staff to validate the information. It's trivial to install cut-offs and blind alleys in the `whois` information so that it's not possible to find the owner of a name. ISPs and name registration companies hide behind the law and refuse to assist you, even if you have a legitimate query. It's a free-for-all where the unscrupulous flourish. I find it all rather worrying.

Table 1. Magic in the Script

`^`	matches the start of the source string
`.*`	followed by anything–this is being discarded
`\.`	followed by a dot; it has to be quoted with a prepended backslash to match a real "dot" in the source expression
`\(`	the start of match section; if the expression inside the match string is found, then the contents of this match section are output by `expr`
`[a-zA-Z]*`	The match section in this case will be matched by any combination of letters in upper and lower case repeated as many times as we need
`\)`	the end of the match section
`$`	the end of the string

Finally

As I've said, the DNS can offer a way of tracing the real owners of names. At least you can try to track the people who are hosting the names and it can help–sometimes. There are also two commands of interest–`nslookup` and `dig`. I obtained some of the information about `whois` from *DNS and BIND* by Paul Albitz and Cricket Liu, published by O'Reilly and Associates Inc., fourth edition (ISBN 0-596-00158-4). You'll find more information about how to interrogate the DNS in this book.

Peter Collinson *runs his own UNIX consultancy, dedicated to earning enough money to allow him to pursue his own interests: doing whatever, whenever, wherever ... He writes, teaches, consults and programs using Solaris running on an UltraSPARC/10. Email:* `pc@cpg.com`.