

Web Bubba

Should the Greek philosopher Heraclites live today, he would notice that you cannot enter the same Internet twice because between the time you leave and come back, the Internet will have changed considerably. While the Internet was originally designed for a relatively small audience, it has become the principal means of communication on the entire planet. On that section of the Internet we know as the Worldwide Web, you can find anything you want. And much of what you do not want.

The moment you enter the stream you become a target. Not by people who would like to harm you, but by those who would like to sell you something. In the real world, there are market places where various vendors scream to get your attention. There also are places of calm, places of security, places in which you can rest assured that no one will start screaming at you. In the Internet world, however, the vendors have become so aggressive that they pop up at the places you would least expect.

Does that sound bad? Actually, it is much worse. In the real world, only those vendors can try to get your attention that are physically present at the same place you are. So, if you are walking through St. Peter Square in Rome, a man might jump in front of you, show you a spread of slides and say "*Chip, chip!*" (this actually happened to a friend of mine), but no vendor from the Valley of the Kings will be jumping next to him trying to sell you the original death mask of Tutankhamon, nor will another one interrupt to offer you tickets to the latest Broadway show. Of course, if you are standing in front of a Broadway theatre, someone might try to sell you tickets, but then no one will be offering you the Vatican slides. That's the real world.

The Internet world, on the other hand, is where all vendors are competing for all Internet visitors. At the same time. Regardless of where they are. Regardless of who they are. Regardless of how old or young they are. In most cases it costs the vendors nothing to start screaming at you from the moment you have entered the floods of the Internet. They have nothing to lose. They have everything to gain if they do get your attention. And they know that everyone else is screaming at you at the same time. So they all scream louder and louder and louder.

It has come to the point where it is virtually impossible to find a place on the Worldwide Web where some ad does not show up on the top of the screen and where that ad will not be flashing and shaking. Unfortunately, they have become so distracting that even if you try to ignore the ad, it is still flashing in the corner of your eye. And because of that, you are often unable to read what you have come to read in the first place. As a result, many people simply give up on the Internet, or at least do not find the information they have set out to find. That's not good!

Fighting Back

If this happened in the real world, you would be fighting back. You would be training in martial arts and doing all it takes to defend your personal space. In the Internet world, many people just give up because they feel powerless.

Well, I assure everyone who feels powerless on the Internet that knowledge is power. If you know how the Internet works, you will know how the vendors are capable of jumping in front of you and screaming in your face all the time. And then you can defend yourself.

And if you think that the Internet is just so complicated, just so out of your intellectual reach, I assure you that it is not. After all, these vendors have figured it out. They did it by getting together and pooling their knowledge. Well, we can do the same.

Internet: A Crash Course

The Internet is really very simple. It is based on the same principle as the phone network. As long as you have a telephone hooked up to the network, you can call anyone anywhere in the world, presuming that he, too, has a telephone hooked up to the telephone network. All you have to do is dial up his number.

The Internet is one huge network similar to the telephone network. In fact, much of its traffic goes over the same lines as the telephone network. On the Internet everyone has a number, too. These are not called telephone numbers. Instead, we refer to them as the IP addresses, or simply the IPs. By the way, IP stands for Internet Protocol.

The IP address is a series of four numbers separated by a dot. None of these numbers can be smaller than zero or larger than 255. So, 127.0.0.1 would be a valid number, while 385.13.-8.0 would not.

Unlike with phone numbers, many computers do not always have the same IP address all the time. Those of major web sites do, but typical home computers get their IP address assigned every time they connect to

the Internet. They may get the same number as before, or they may get a different one. This is because there are only so many possible combinations of four numbers ranging between 0-255. A little over 4 billion theoretically, but some of the numbers are special, so that excludes them.

Now, all you have to do to connect to a web site, or anything else on the Internet, is to “dial” its number. Of course, it would not be practical for us to remember all those numbers. But almost all of those web sites do have a name, or a domain, as it is called. Such as `WWW.REDPRINCE.NET`, or `WWW.BESTEES.COM`, or `WWW.EASYWEBFILTER.COM`, etc.

Remember Heraclites? You cannot enter the same Internet twice? That also means that new web sites come and old ones go. And some just move. They keep their name but change their IP address. That makes it impossible for a web browser to simply know where every single web site is. Instead, it has to look up their IP numbers one at a time, just as you look up phone numbers of the people you call.

Where do you find people’s phone numbers? Those you call all the time, you probably remember. The others you look up in your phone book. If they are not in your phone book, you dial directory assistance and ask. Your computer works almost exactly the same. It remembers some IP addresses, usually those that you have visited very recently. It does so by storing them in a cache, which is a fancy name for a section of its memory. However, it only stores them there temporarily. This is because of the Heraclitan nature of the Internet: An IP number can change at any time, so remembering it for too long is not a good idea. At any rate, the cache is the first place where your computer looks every time you tell it to visit any place on the Internet.

If the IP address is not stored in the cache, the computer must look it up. Most modern operating systems do have a “phone book” where they look next. Why do they have a “phone book” despite the Heraclitan nature of the Internet? Because many computers are hooked up to a *local area network* (LAN). Earlier I mentioned that certain IP numbers were special. Some of these numbers are reserved for the computers on a LAN, specifically, 192.168.0.0 - 192.168.255.255, 172.16.0.0 - 172.31.255.255, and 10.0.0.0 - 10.255.255.255. Typically, small LANs use the first range, medium ones the second, and large ones the third. For your computer to find other computers within its LAN, these private IPs must be listed in each computer’s own “phone book.” Most modern operating systems do so by listing them in a special file called `hosts`. On Unix systems, this file is usually located in the `/etc` directory. Windows uses a different file organization, so this directory does not exist. As is typical for Windows, each version of the system stores the `hosts` file in a different directory. Windows 95 has it in `C:\WINDOWS`, Windows XP in `C:\WINDOWS\SYSTEM32\drivers\etc`, and so on. At any rate, if the computer does not find the IP of a domain in its cache, it looks into the `hosts` file.

Finally, if the computer cannot find the IP of a domain in either its cache or the `hosts` file, it consults with a computer on the Internet called a DNS server. This is akin to you calling the directory assistance. DNS, by the way, stands for Domain Name System.

There is something else we need to know about the Internet, specifically about the organization of the web pages. A typical web page consists of some text and some images. When your web browser requests that page from its web site, it only receives the text. This text contains directions instructing the web browser what to do. For example, it tells the browser to show you some text and to show it in bold letters, or just as regular text. It also tells the browser where on the page it should display an image. Or several images. It also tells it where it can find those images.

That is right. The images are not included with the text. The web browser has to fetch every single image from its place on the Internet. And that image can be anywhere on the Internet. It can come from the same web site the text of the page is coming from, *or from a totally different one.*

This concludes our crash course on the Internet. We now know enough about the Internet end of the equation. But we still need to take another crash course before we can fight back effectively.

e-Advertising: A Crash Course

The Heraclitan nature presents a major challenge to Internet advertisers. They cannot just contact every web site and negotiate an advertising deal with it, and do so every time they want to post a new ad or change an existing one, or perhaps ditch one that does not work. There are too many web sites, and they change too much. Further, some web sites are popular and visited by millions of people, while others are just sitting there barely noticed by the public. Additionally, some web sites specialize. For example, there are web sites that talk about nothing but reincarnation, while others talk about geometry, yet others about red hair. But the vast majority of web sites do not specialize in anything, they just talk about anything their webmaster wants to talk about.

Advertisers love the sites that specialize in something. These sites attract an audience with a common interest. That means that an advertiser can tell in advance whether the visitors of that web site would be interested in whatever it is he wants to advertise. The advertising placed on these web sites is usually incorporated directly into the contents of the web site itself. It is usually subtle and non-obtrusive. The visitors of the web site (such as you and me) usually do not object to this kind of advertising—after all, the webmaster has to pay for the web site somehow. Often we even welcome the advertising—after all, it tells us about products we may actually be interested in.

Alas, most web sites either do not specialize in anything, or do specialize but in something that is of interest to everyone (for example, a search engine specializes in web search, but everyone needs that). So, here the advertisers just scream at everyone, hoping to catch their attention. And this is the type of web site subject to the Heraclitan nature of the Internet. These are the sites that come, go, and change. These are the sites that are so numerous that the advertisers cannot contact them all and place ads on them all.

An entire industry has evolved to face the Heraclitan challenge. Web sites that do nothing but offer advertising. We will refer to them as the advertising agencies for the lack of a better word. The advertisers make a deal with such agencies. They place their ads on special servers that do nothing but serve the ads all over the Internet. If an ad brings in a customer, the advertiser pays the advertising agency.

The webmasters of the web sites that cannot carry specialized advertising also strike a deal with the advertising agencies. The webmaster will let the agency post the ad on his web site by serving it directly from its own servers. If someone clicks through and buys, the advertising agency shares its profits with the webmaster of the site where the click-through occurred. The webmaster does not even know what ad is running on his site when you visit it. It is served directly by the advertising agency which can post any ad it wants and it can and does change the ads at any time.

This is the type of advertising that we find so objectionable and that most of us would like to fight. So, let us take a closer look at it again:

This type of advertising consists of an image that does not come to you from the web site you are visiting but from the web servers of the advertising agency. Remember from our first crash course: the images you see on a web page are downloaded separately from the text of the page, and they can come from a different web site than the text does. And that is exactly where these ads are coming from, a different web site.

Defense Against the Dark Arts

Now that we understand how the advertisers send their ads to our computers, we know just what exactly we want to do to defend ourselves. We want our browser not to download the ads from the ad server even if the text of the web page instructs them to. If we could somehow accomplish that, we would not only get rid of all the unwanted advertising, we would also speed up our browsing experience. The fact is that downloading all those ads from the Internet takes time. If you are a dial-up Internet user, the time seems like eternity. If you are on DSL, it is not quite as bad, but it is still slowing you down. And it is just as distracting and annoying.

Now that we know our goal, we need to figure out how to accomplish it. Depending on what web browser you are using, you might configure it not to download the contents served from those domains that the advertising servers use. Alas, not every browser lets you do that. And there are thousands of domains you would have to tell your browser about.

Another solution would be to configure your browser not to download images from sites other than the web site whose web page you are reading. Again, your browser would have to offer such an option. But even if it does, it is not that uncommon for one web site to display images from another web site, so you would not be able to see those. A better solution is needed.

A better solution exists. It is not your browser that you need to configure. After all, every time you upgrade your browser or switch to a different browser, you would have to reconfigure it from scratch. Besides, more and more software is ad-supported. That is, software that runs on your system. Such software downloads its advertising from the ad servers directly, bypassing your web browser. Configuring your web browser would have zero effect on all the other software.

Enough with what does not work. Let us take a look at what does. What we need to do is let the browser and all the other software do its work but change the “phone number” (the IP address) of all the ad servers to “ring” a web server that will satisfy all requests with a tiny, invisible image. That is, an image that is just 1 pixel wide and 1 pixel tall, and that is completely transparent, so the browser (and all the other software) will display absolutely nothing at all where the ad is supposed to be.

At first, this may sound hard to accomplish. If a web server existed that could satisfy all of its web requests with this tiny invisible image, would not everyone configure their computers to visit this web site? Would this web server not be so swamped with requests that it would either not be able to satisfy them all or would do so very, very slooooooowly?

Most likely, so.

But what if we could run such a server on our own computer and redirect all those ad requests to the server on our own computer? This would work perfect, but only if certain conditions were met:

- The server must be extremely small and not take up too much memory.
- The server must be extremely fast and fill all its requests instantly.
- The server must only be visible to the local computer or, optionally, to the local area network (not accessible by the outside world).

While several servers were written to accomplish the task, only one satisfies all three conditions. That server is *Web Bubba*. Written in assembly language (the language of the computer, not a human programming language), it is both extremely small and fast. In its default configuration, it is only visible to the local computer, but you can easily make it visible to your LAN, if that is what you want. When asked for any resource, regard-

less of its name or path (directory), it serves a tiny invisible GIF image. By the way, the image was created in a registered and fully licensed copy of Corel Photo-Paint 8, so it is not in violation of any patent (as many GIF files are alleged to be).

We then need to take two separate and independent steps to block the unwanted advertisements from ever reaching our computer. We need to redirect all of their requests to our own computer, and our computer needs to satisfy any requests from a web browser (and similar software) by serving it the invisible image.

Step 1: Redirection

As I mentioned earlier, several IP addresses are reserved for special purposes. One serves as a *loopback*. That means that any networking requests made to this address are sent right back to the computer that made them. This special IP address is 127.0.0.1 and is called `localhost`. Take a look at your `hosts` file, discussed above. Most likely it starts with some comment lines. You can recognize them as starting with a `#`. But, comments or not, the file defines 127.0.0.1 as `localhost`. It probably looks something like this:

```
#####  
#  
# Someone's copyright notice.  
#  
# Some explanation of what this is all about.  
#  
#####  
  
127.0.0.1    localhost
```

That's probably it, unless you have changed it to include the information about the various systems within your LAN. For example, on my computer it looks similar to this (you really don't expect me to tell the whole world what my computer is configured like, do you):

```
127.0.0.1    localhost bubba  
192.168.1.1  lan router  
192.168.1.2  potter  
192.168.1.3  weasley  
192.168.254.254 .dsl modem
```

That allows me to enter either `localhost` or `bubba` to my web browser and get the *Web Bubba* server to do its work. If I type either `lan` or `router` I get my router's built-in web page. If I type `dsl` or `modem`, I get the web page build into my DSL modem. My computer's LAN IP address is 192.168.1.2, so if I type `potter`, I get the contents served by the Apache Web Server running on my computer. And if I type `weasley`, I get to my other computer connected to my LAN as 192.168.1.3.

Now suppose I inserted several lines defining some very famous and popular web sites as being located at the IP address of 127.0.0.1:

```
127.0.0.1    localhost bubba
⇒ 127.0.0.1    yahoo.com www.yahoo.com
⇒ 127.0.0.1    google.com www.google.com
⇒ 127.0.0.1    altavista.com www.altavista.com
192.168.1.1  lan router
192.168.1.2  potter
192.168.1.3  weasley
192.168.254.254  dsl modem
```

If I now typed any address within these domains into my web browser, including a path within the domain, e.g., <http://www.google.com/search?hl=en&ie=UTF-8&oe=UTF-8&q=Web+Bubba>, my system would find the entry in my `hosts` file and make a request to the web server on my own computer at the IP address of `127.0.0.1`. As far as my computer would be concerned, all these domains would exist on my computer. Thus, I would effectively wipe those domains off the Internet. Or, to put it in a more technical language, I would filter them out.

Naturally, I am not interested in filtering out the search engines. I have chosen them as an example because I suspect that you do not want to filter those domains out either. I am not trying to tell you which domains to filter out, only showing you how to do it.

Now, although I am not filtering out the search engines, I *am* filtering out more than 18,000 domains by having them listed as `127.0.0.1` in my `hosts` file. How did I figure out which domains to include in the file, and so exclude them from existence as far as my computer is concerned? Most of them come from suggestions made by others, many come from my own visits to web sites whose contents I wanted to read but often could not because of all the ads they were displaying. I examined the HTML code of these pages, and found the domains of the ad servers, then included them in my `hosts` file.

You can do the same. You can download my file from <http://www.easywebfilter.com> if you want. Or you can download someone else's `hosts` file. Many people have made theirs available for the download over the Internet. You can then add your own undesirables to your heart's content.

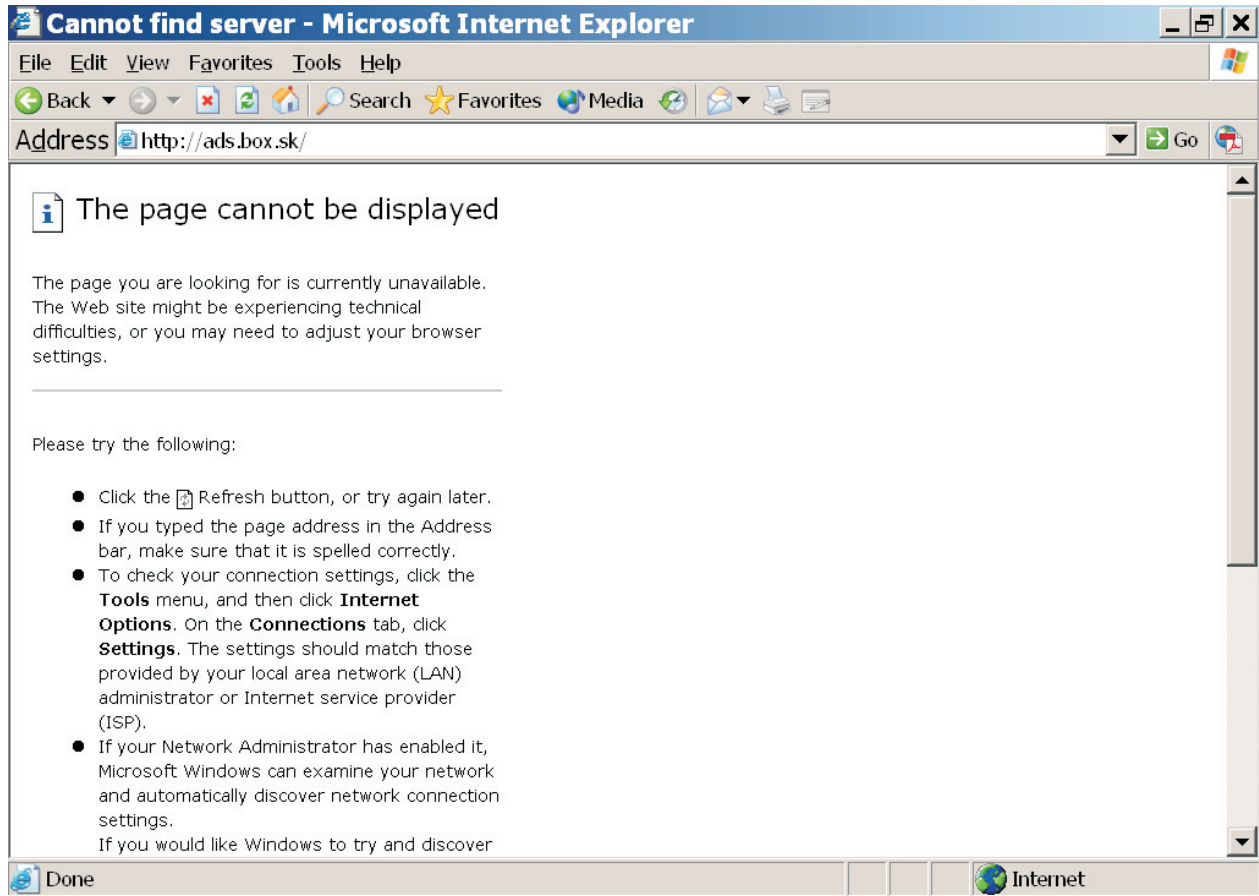
Please note that stopping advertisers is not the only reason you may want to include domains in your `hosts` file. Perhaps you have children and do not wish to expose them to certain web sites that you find offensive. While just about all of us agree on the exclusion of the ad servers, what else we want to exclude can be quite subjective. But we can do it. And we have the freedom to do it. Most importantly, we can do it on our own. There are services that will filter out web sites for you for a monthly fee. Using these services is not only a complete waste of money (paying someone else to do something we can easily do on our own), it also means letting someone else decide which sites we should or should not see. And that is the first step in letting others control our lives.

So, please, feel free to download my file or anyone else's file and to edit it in any way you want.

Step 2: The Server

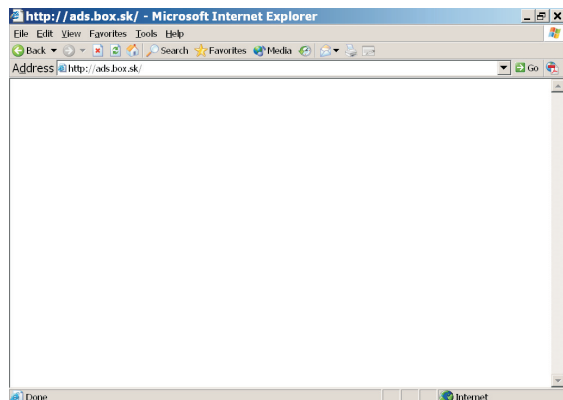
Having done the first step is the most important part of the web filtering process. Any domain you have re-declared as IP `127.0.0.1` will automatically vanish from your computer's universe. Your web browser and other Internet software will ask your own computer to serve it some web contents. What happens next de-

depends on whether you have a web server installed and running on your computer. Most people do not have one. If you are one of them, your browser will wait for the computer to obtain some contents for a while. Eventually, it will realize that nothing is coming and will give up trying. A typical web browser will then draw a little crossed-out box in the place where the image was expected. That is much better than downloading all those ads. But it still slows you down somewhat and still shows you those crossed-out boxes. Or, it can display an error message, which is less desirable. Here is an example of what may happen if you do not have a web server at your computer's 127.0.0.1 IP address:



The solution is to have a tiny web server installed and running on your system. This web server should just give the browser a tiny invisible image. The faster it does so, the better. Your browser will be satisfied that it has received something. Even if it expected something else, perhaps text, or a different image file format, it will know that the transaction is final and closed. It will not waste any time waiting for more data to come.

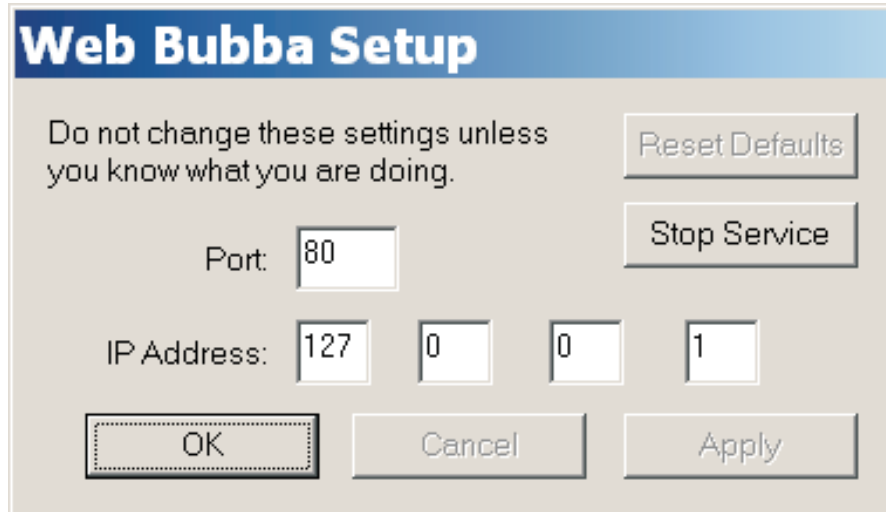
I tried to load the same web site as in the above example but with the *Web Bubba* server active. The result was quite different this time. The browser quickly displayed a blank page. And we only need a small picture to show you (but you can zoom in with your Acrobat Viewer if you want to):



Web Bubba

Web Bubba is a tiny web server which can be installed and run on Windows systems. On the NT-based Windows systems (Windows 2000, Windows XP, etc.) it runs as a Windows service. The earlier (9x) versions of Windows do not support Windows services. However, *Web Bubba* can run on these systems as an ordinary program. Other than having a Windows-based computer system, there are no special requirements for *Web Bubba*. You can download *Web Bubba* from <http://www.easywebfilter.com>. Then just run its installation program.

By default, *Web Bubba* acts as a tiny web server listening at port 80 of the 127.0.0.1 IP address. That is exactly what is needed when redirecting various web domains to that address. Should a special need arise, however, you can change both the port and the IP address from its configuration utility. Even though the majority of users will never need to change these settings, you should load the configuration utility after installing *Web Bubba*. It will show a window that looks like this:



It should list the port as 80 and the address as 127.0.0.1. On an NT-based system the Stop Service button should be visible. Click on it. It should gray out for less than a second, then change to Start Service. Click on it again, and the service should start again. If it does not start (the button continues to say Start Service), that means that some other web server is running on your computer listening to port 80 at the 127.0.0.1 IP address (and possibly other addresses). Hopefully it is a server you have installed. You need to reconfigure it not to listen at the 127.0.0.1 IP address. Please consult its documentation on how to do that.

If, however, you have never installed a web server on your system, it is possible that some spyware program has installed one without letting you know. You can find out what you have running there by loading your web browser and typing <http://localhost> to its URL window. You can also learn more about spyware at <http://grc.com>.

It is also possible that the button is all grayed out and you cannot click on it at all. This happens when *Web Bubba* is not installed as a Windows service (as would be the case if you use Windows 9x).

For most people that's all there is to it. If you ever want to stop the service, just load the configuration utility and click on the button. If you ever need to change the IP address or the port, do it from the same utility. If the service is running, your changes will become effective as soon as you click Apply or OK, or press enter.

Advanced Usage

It is possible to use other IP addresses than 127.0.0.1 for your Internet filtering. You simply enter the new address in your *Web Bubba* configuration utility, and it will start listening to requests made to that address instead of 127.0.0.1 (unless, of course, you already have another web server at that address). But remember, your Internet filtering is a 2-step operation. So if you reconfigure your server, you also need to edit your `hosts` file, and replace the various occurrences of 127.0.0.1 with whatever your server is listening to. **You should, however, never change the definition of `localhost` to anything other than 127.0.0.1.** Change anything else, but keep your `localhost` as 127.0.0.1. Did I mention you should never change your `localhost` to an address other than 127.0.0.1? I mean, NEVER.

So, why would you want to use something other than 127.0.0.1? The truth is that most people will not. A good rule of thumb is: *If it ain't broke, don't fix it!* I can think of two very specific reasons under very specific circumstances when you might want to change it (hey, you can always click on `Reset Defaults` if you mess up).

One reason is if you need to run a web server at `localhost`. Perhaps you are a web developer and run a local copy of the Apache Web Server (or some other web server) on your computer to allow you to test your designs by entering `localhost` in your web browser's URL window before you upload it to your web site. No problem. Your computer loops any request to any IP address in the 127.x.x.x range back to itself. So, you can just edit your `hosts` file, change all occurrences of 127.0.0.1 to, say, 127.1.2.3 (all but `localhost`, of course), and configure your *Web Bubba* to listen to port 80 at IP address 127.1.2.3, and all is working properly.

The second reason is if you have a local area network, and want to redirect some of the other computers on the network to your own *Web Bubba*. In that case, you may want to reconfigure *Web Bubba* to whatever your LAN IP address is. Typically, this may be something like 192.168.1.2 (but can be something else, so check it first). And, of course, you will need to edit your `hosts` files accordingly, not just on your own computer but on all those computers on your LAN.

Now, just because you have a LAN, does not mean you should let all the computers on your LAN access your local copy of *Web Bubba*. In fact, I strongly discourage it. *Web Bubba* is so small that you can just install it on every computer on your network. Well, every Windows computer on your network.

If you still want to access *Web Bubba* over a LAN, then please be advised of two cautions:

- ❑ The computer with *Web Bubba* installed must be turned on and running every time another computer on the LAN needs it.
- ❑ Make sure your router or modem does not forward HTTP requests from the outside world (that is anyone on the Internet) to your computer. Otherwise, you will be generously serving empty invisible images to anyone on the Internet. Now, that will not hurt you in any way (all *Web Bubba* does is serve those empty images, so no one can break into your system through it), but why do it?

By the way, if you *want* the entire Internet to have access to your *Web Bubba*, change the address to 0.0.0.0. That is a special address which tells your computer to open up a port to anyone anywhere.

One final thing, lest I forget: Don't ever change your `localhost` to anything other than 127.0.0.1. Some programs might stop functioning properly.

Why I Wrote *Web Bubba*

I got really tired of untargeted advertising appearing on so many web pages. The ads were taking too long to download. They were distracting with all their flashing and shaking. I am the kind of person who finds on his own whatever products he needs.

I applied step 1. I downloaded other people's `hosts` files, combined and cleaned them up, and added to them the servers of whatever unsolicited advertising still kept appearing in my browser. While that stopped the ads, it was still causing some browser confusion. While it correctly started asking my system for the ads, it did not like the missing responses. Quickly I realized that step 2 was needed.

Since I run the Apache Web Server on my computer, I configured it to serve a default empty GIF to all requests to `127.0.0.1`. While that solved the browser problem, my web logs were suddenly growing at an astronomical rate, as the Apache Web Server is logging every request it receives. And present-day web sites just force the browser to make too many such requests.

So, finally, I wrote my own server. I made it extremely fast because I like my browser to get its work done fast. And I made it extremely small because I did not want to waste any system memory. I also made it easy to turn on and off because one of the browsers that I use occasionally (Opera) is too easy to confuse. When it loads its home page which tells it to run some JavaScript, it makes a request for that script. It then appears to assume that it is receiving the script and tries to execute it, even though the web server has told it clearly that it is sending it an image. So, I turn the server off when I use Opera and on when I use Internet Explorer and any other web-related software.

Now, as long as I wrote it, I may as well give it to you. You may use it on your home system for free. If you are using it at any other location, you are required to pay for it by purchasing a special *Web Bubba fightin' back* T-shirt from <http://www.bestees.com/webbubba>, one for each non-home computer that you have the software installed on.