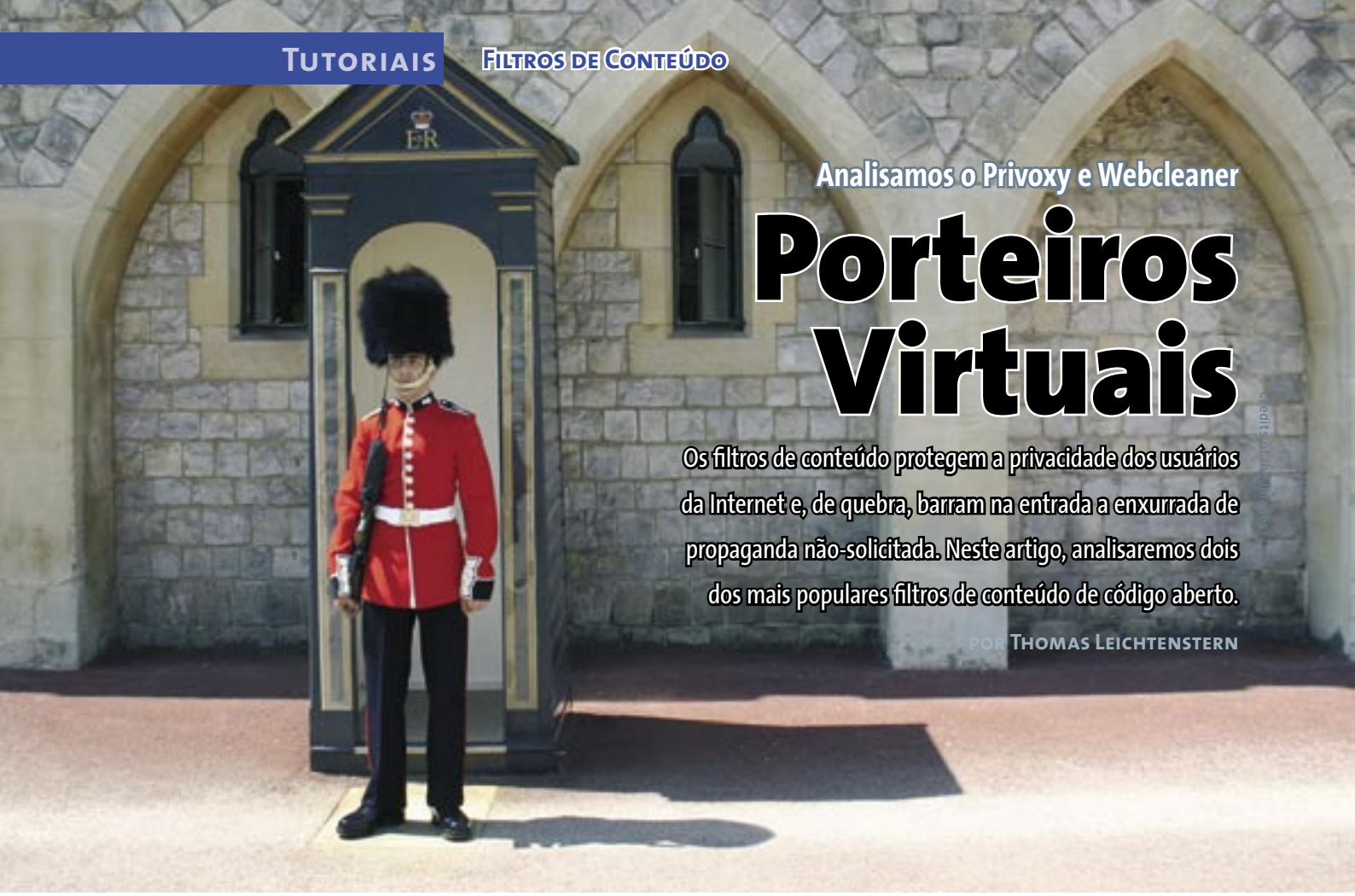

Analisamos o Privoxy e Webcleaner

Porteiros Virtuais

Os filtros de conteúdo protegem a privacidade dos usuários da Internet e, de quebra, barram na entrada a enxurrada de propaganda não-solicitada. Neste artigo, analisaremos dois dos mais populares filtros de conteúdo de código aberto.

por Thomas Leichtenstern

Filtros de conteúdo são úteis para evitar tráfego irrestrito entre os navegadores e os servidores web. Um bom filtro de conteúdo deixa passar apenas o tráfego que o usuário realmente quer, mantendo o navegador a salvo do lixo comercial, *web bugs*, *cookies* e Javascripts chatos. Alguns filtros podem tratar também o tráfego que sai em direção à Internet. Um filtro de conteúdo configurado corretamente pode até mesmo proteger contra falhas de segurança no próprio navegador – coisa que, infelizmente, é algo comum ainda hoje (veja o artigo *Dicas de [In]segurança* na página **16**).

Este artigo investiga os filtros de conteúdo Privoxy [1] e Webcleaner [2]. Ambos oferecem filtragem de conteúdo, mas enquanto o Privoxy concentra-se mais no conteúdo da web, o Webcleaner possui uma infinidade de recursos úteis, como filtro de vírus e um compressor de imagens.

Redirecionando o acesso

Os filtros de conteúdo funcionam mais ou menos como servidores de proxy – ou seja, como sistemas intermediários entre seu navegador e o servidor web. Para usar um filtro de conteúdo, é preciso redirecionar todas as tentativas de acesso do seu navegador para o endereço do filtro, normalmente alterando suas configurações de rede para apontar para o IP e a porta do filtro. Se o filtro estiver rodando no computador do usuário, o endereço será 127.0.0.1, número IP que universalmente sempre aponta para a própria máquina (o chamado *localhost*).

Privoxy

O *Privacy Enhancing Proxy* – *Privoxy* para os íntimos – é um filtro de conteúdo simples, baseado no *Junkbuster*, que não faz nenhum tipo de *cache*. Ao contrário de um simples filtro de URL, entretanto, o programa verifica o conteúdo total do site sendo visitado, seguindo as regras definidas pelo usuário.

Instalação

O Privoxy está disponível para todas as principais distribuições, portanto não será problema encontrar o pacote RPM, TGZ ou DEB apropriado. Os usuários do Ubuntu podem habilitar o repositório `Universe` e instalar o Privoxy com o comando `apt-get install privoxy` (ou usando o *Synaptic*). No Debian Sid, o comando é o mesmo, mas nenhuma alteração na lista de repositórios é necessária. No SuSE 9.3, abra o YaST e instale o programa a partir dos CDs de instalação. Observe que o YaST instala o Privoxy em uma "jaula" *chroot*. O caminho para os arquivos de configuração e registro de eventos ("*log*") é `/var/lib/privoxy/`.

Configuração

Por padrão, o Privoxy entra no ar pelo endereço local (127.0.0.1). Se você quiser que o filtro de conteúdo fique disponível para outras máquinas, altere o valor padrão, no arquivo `/etc/privoxy/config`, de `listen-address 127.0.0.1:8118` para seu endereço na rede local – por exemplo, 192.168.0.1. Se você deixar o endereço IP em branco, o Privoxy vai ficar disponível em todas as interfaces de rede – o que não é recomendável, especialmente se uma delas estiver ligada diretamente à Internet...

A interface web permite que se configurem regras de filtragem. Para acessá-la, vá à página config.privoxy.org – sim, é uma página na Internet, mas se o Privoxy estiver ativo ele interceptará a chamada e mostrará a interface de configuração. Antes de chamar essa página, entretanto, devemos configurar o Privoxy como o servidor de proxy do seu navegador – no Mozilla Firefox, por exemplo, isso estará em *Editar* | *Preferências* | *Geral* | *Conexão*. Nos sistemas baseados no Debian, algumas configurações a mais são necessárias. Procure pelas linhas `enable-remote-toggle` e `enable-edit-actions` no arquivo `/etc/privoxy/config` e atribua a ambas o valor `1`. Depois de concluir as alterações, reinicie o programa com o comando `/etc/init.d/privoxy restart`. Como o Privoxy não possui qualquer rotina de autenticação, qualquer usuário com acesso ao filtro pode alterar suas configurações.

Filtros

O Privoxy distingue entre arquivos de filtro e de ações. Os filtros possuem regras para, por exemplo, remover banners cujo tamanho ultrapasse um certo limite. Já as ações mapeiam as regras a sites específicos. Quando falamos em sites, estamos dizendo tanto sua URL completa como também caracteres coringa que representem fragmentos de endereços. ad*.exemplo.com consideraria todos os subdomínios de exemplo.com que incluíssem a partícula *ad* seguida de qualquer cadeia de caracteres.

Hasta la Vista, Baby

O arquivo padrão de filtros (`/etc/privoxy/default.filter`) vem de fábrica com uma coleção bem grande de regras. Não tente mexer em nada no arquivo de filtros, a não ser que esteja craque em expressões regulares! Como a interface de configuração do Privoxy não permite que se edite o arquivo de filtros, será preciso usar seu editor de textos preferido para a tarefa. O exemplo a seguir mostra mais ou menos como uma regra se parece:

```
FILTER: LinuxMagazine Regra de exemplo ➘
s/chuva(?!.com)/sol/ig
```

`FILTER:` define uma nova classe e contém o nome do filtro (`LinuxMagazine`) e uma descrição resumida (`Regra de exemplo`). A interface de administração web mostra essas informações. A segunda linha contém a regra em si. Neste caso, simplesmente substitui a palavra `chuva` em qualquer texto na página visitada por `sol`. Uma classe pode ter um número ilimitado de regras, que são ativadas por um único clique na interface web. Há inúmeros filtros para download na Internet [4].

Figura 1: A tela principal do *Privoxy* possui muitos menus com várias funções.

Ação!

Você pode ter as regras mais bem-elaboradas do universo, mas se não tiver um alvo elas serão inúteis. É para isso que servem os arquivos de ação. Tanto `user.action` como `default.action` são arquivos de ação e ambos podem ser editados pela interface de administração do Privoxy, basta acessar o endereço: config.privoxy.org/show-status. Embora os dois arquivos tenham uma estrutura idêntica, são usados para propósitos bastante diversos. O arquivo `default.action` especifica o comportamento global, enquanto `user.action` lida com aplicações específicas.

O arquivo `default.action` contém as regras padrão que são aplicadas caso nenhum dos outros arquivos seja aplicável. O Privoxy possui três políticas predefinidas para novos usuários. Na realidade, pode-se clicar em alguns links na interface de administração para mudar o comportamento do filtro, desde *Cautious* (*Cuidadoso*) até *Adventuresome* (*Ousado*).

Figura 2: O *Privoxy* mostra as regras na interface de administração.

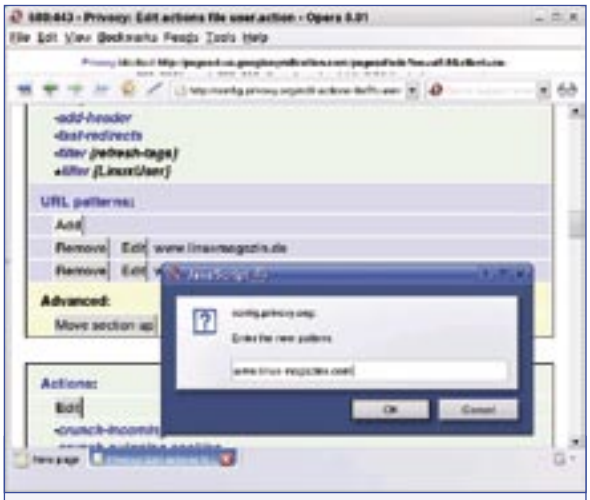

Figura 3: Depois de "pegar a manha", criar suas próprias regras para o *Privoxy* é muito fácil.

As seções seguintes contém políticas para manipulação de endereços e fragmentos de endereço, com exemplos de como lidar com SPAM baseado em suas URLs – por exemplo, URLs que contenham as palavras`ad*.` ou `*banner*.`. A configuração global restringe-se, na verdade, a selecionar uma das políticas padrão.

As regras definidas pelo usuário são criadas na seção `user.action`. Se o Privoxy estiver bloqueando conteúdo justamente do seu site preferido, vá até http://config.privoxy.org/show-url-info para encontrar qual é o filtro responsável pelo sacrilégio. Clique em *Insert new Section at top* na seção `user.action` e, em seguida, no botão *Add*. Informe a URL da página em questão. O botão `Edit` mostra uma lista de regras no `default.filter` que são aplicadas nesse caso particular. Todas as regras estarão em *No Change*, indicando que deve ser usado o padrão definido em `default.action`. Você pode então ativar ou desativar cada uma das regras a seu bel-prazer – as alterações serão gravadas no arquivo `user.action`. Todas as alterações em `user.action` têm prioridade sobre os valores padrão definidos em `default.action`.

Sinfonia inacabada

O Privoxy é bastante discreto, trabalhando "incógnito" na maior parte do tempo, e raramente atrasa o carregamento de alguma página, mesmo as grandes. Obviamente, o hardware que vai rodar o Privoxy não deve ser exageradamente decrépito. Uma CPU de 500MHz e 256MB de RAM é o limite mínimo.

O programa mostra a maior parte dos sites corretamente, mesmo se você usar a configuração *Cautious* para os filtros. Se algo der errado, pode usar o verificador de URL, que informa quais regras foram aplicadas à página atual. Infelizmente, o verificador não é lá muito organizado nem tem uma estrutura clara, deixando o usuário "com as calças na mão" na hora de encontrar a regra que está arruinando a vista.

Se um site for bloqueado completamente, o Privoxy ainda assim oferece uma rota de fuga: um link espertamente chamado de *go there anyway* (vá lá mesmo assim). Esse link desabilita o filtro de conteúdo temporariamente para a página atual. O programa insere ainda os chamados *bookmarklets* para habilitar e desabilitar os filtros. Para abrir os *bookmarklets* clique no link *Privoxy - Toggle Privoxy* na página inicial do Privoxy: http://config.privoxy.org. Na janelinha que pipoca na tela, clique na opção desejada para ativar ou desativar o proxy local.

Privoxy: Conclusões

Demos três vivas aos desenvolvedores do Privoxy! A documentação é exemplar, descrevendo todos os recursos do programa em detalhes. A combinação de filtros e ações pode parecer confusa no começo, mas se pensarmos um pouco veremos que é uma idéia excelente.

No fim das contas, o Privoxy é bem pensado e maduro. Rodou sem um único senão ou falha em nosso laboratório e desempenhou com galhardia a tarefa a que se destina: manter nosso navegador livre de comerciais indesejados e proteger a privacidade do usuário sem mudar os padrões.

Webcleaner

O Privoxy é muito bom, mas ficou na sombra de nosso segundo candidato. O *Webcleaner* possui uma estonteante lista de recursos. Além da filtragem de conteúdo, os desenvolvedores tiveram o requinte de incluir mimos como a compressão e redimensionamento de imagens, filtragem de vírus e detecção (com correção!) de erros de HTML conhecidos.

Instalação

Como o Webcleaner foi desenvolvido para o Debian, ele é muito mais fácil de ser instalado em sistemas baseados em Debian como o Ubuntu. É claro que ele pode ser instalado em outras distribuições sem problemas, mas talvez isso requeira do usuário um pouco mais de suor de seu rosto. O esforço é maior ainda se você for o feliz usuário de um SUSE Linux.

O Webcleaner requer o *runit* e a versão 2.4 do interpretador *Python*, incluindo os pacotes de desenvolvimento. Para compilar o Webcleaner a partir do código fonte, vamos precisar ainda de um compilador de C – mas não tema, com `gcc` não há problema!

Para usar todo o potencial do Webcleaner é necessário instalar, antes, os seguintes programas e bibliotecas:

- PIL (*Python Image Libraries*) – bibliotecas em Python para compressão e redimensionamento de imagens.
- *Open-SSL* e *Python-openssl* – para aplicar o filtro de conteúdo a páginas criptografadas com o SSL.
- *Clamav* (`clamd`) – para usar o anti-vírus de forma integrada.

A extensão *psyco* do Python, se instalada, oferece mais recursos ainda. De

acordo com os desenvolvedores, o psyco melhora o desempenho quando compilamos scripts em Python por um fator entre 2 e 100 – embora o processo devore um montão de memória.

Usuários do Debian podem facilmente instalar esses pacotes com *apt* ou o *Synaptic*. Se você possui o SUSE 9.3, será preciso instalar e configurar o *runit*, o *psyco* e o *PIL* manualmente, embora os outros pacotes estejam disponíveis nos CDs e DVDs do SUSE.

Instalação

Digite `tar xfvz webcleaner-2.29.tar.gz` para descompactar o arquivo com o programa. Entre no diretório recém-criado e inicie o processo de compilação com o comando `./configure && make`. Em alguns casos – especialmente no Debian – talvez o script de compilação reclame de alguma biblioteca faltante (como `/lib/cpp`) e interrompa a execução. Nesse caso, instale o pacote `openC++` e reinicie a compilação.

Ao terminar, compile os arquivos em Python com o comando `python setup.py build`. Depois, `python setup.py install` configura e instala o Webcleaner em seu computador.

Se o Webcleaner também for usado para processar sites criptografados, será necessário instalar também os certificados digitais. Para isso, use o comando `webcleaner-certificates install`. Por fim, `make installservice` coloca o *daemon* do Webcleaner em operação. Esse daemon é monitorado pelo *runit* e é colocado em operação imediatamente após emitir um comando. Para o SUSE, será preciso criar o diretório `/var/service/` antes de rodar o script.

Configuração

O Webcleaner pode ser acessado como um proxy direto ou em nível hierárquico superior através do Squid. O caminho para a conexão sai do navegador, passa pelo Squid e chega no Webcleaner – e, a partir dele, atinge a Internet. Para que isso funcione, configure seu navegador para acessar o proxy pela porta do Squid – normalmente `3128`. Para estender o acesso a outras máquinas, adicione as linhas abaixo em sua configuração:

```
....
060 acl localnet src ❷
192.168.0.0/255.255.0.0
....
097 http_access allow localnet
....
```

A rota direta até o Webcleaner usa a porta `8080`. Se preferir usar a conexão direta ao Webcleaner (sem passar pelo Squid) essa é a porta que deve ser informada nas configurações de proxy do seu navegador. O Webcleaner é configurado por uma interface web no endereço http://127.0.0.1:8080 – portanto, acessível apenas na máquina onde está instalado. Antes de iniciar o Webcleaner pela primeira vez, informe a senha que o Webcleaner gerou durante a instalação. Para isso, copie a senha MD5 para o arquivo `/usr/share/webcleaner/config/webcleaner.config`, inserindo-a na linha `adminpass=` e reiniciando o Webcleaner com o comando `kill -HUP` número do processo. Para descobrir o número do processo do Webcleaner, use o comando `ps aux`.

Depois de tudo configurado, acesse a página de administração (http://127.0.0.1:8080) e inicie uma sessão como o usuário `admin`, informando a senha "visível" mostrada logo abaixo da senha MD5. Ambas são mostradas na página.

O item *Proxy Configuration* dá acesso aos ajustes básicos do programa. Atente

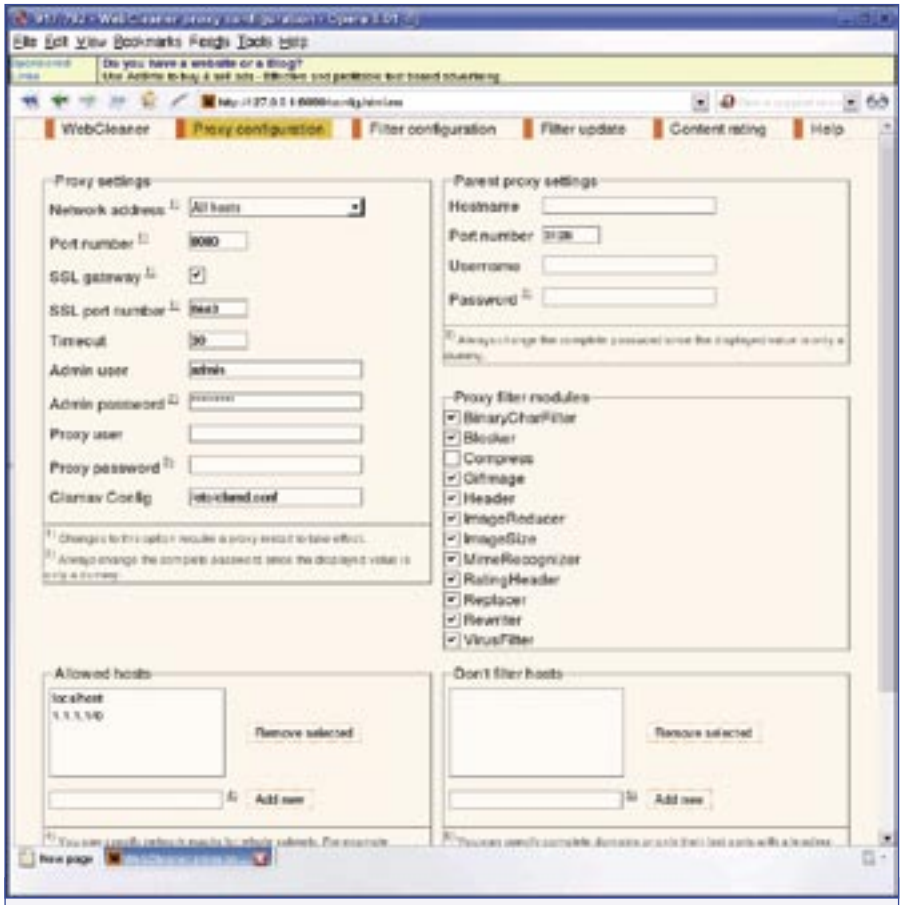

Figura 4: A central de comando do *Webcleaner*

para o quadro *Proxy filter modules*, que traz um resumo dos módulos de filtragem disponíveis. Os mais importantes são:

- *Blocker* – o filtro de URLs.
- *compress* – comprime os arquivos antes de transmití-los.
- *header* – modifica, substitui e apaga o cabeçalho HTTP.
- *Image Reducer* – comprime imagens usando um formato JPEG de baixa resolução.
- *Rewriter* – interpreta e reescreve código HTML e Javascript.

Na parte inferior da tela, sob *Allowed Hosts*, é possível especificar as máquinas com permissão de utilizar o proxy. Informe quaisquer máquinas que tenham salvo-conduto para receber conteúdo **sem** filtragem em *Don't filter Hosts*.

A seção *Filter configurations* inclui regras para arquivos. As linhas na coluna da esquerda representam os diretórios. Quando se clica em um deles, as regras correspondentes a ele são mostradas na coluna central. Se uma das regras for clicada, um menu de configuração para ela aparece na coluna da direita. Para criar uma nova regra, clique no item *New rule*; um menu de configuração para a nova regra aparece à direita, que depende do item selecionado. Nele, informe a ação desejada.

A seção *Content rating* pode ser usada para especificar como classificar (isto é, dar notas) a páginas individuais. Em nosso laboratório, o Webcleaner só esporadicamente aceitou nossas notas nessa seção.

Uma mancha no sol

Nos testes em nosso laboratório, o Webcleaner mostrou algumas deficiências graves. Depois de habilitar o filtro de vírus, alguns downloads não funcionaram de jeito nenhum. Outros até baixaram, mas não eram nada confiáveis. Quando abrimos um site, preparado por nós mesmos e contendo uma falha de segurança atual que afeta o navegador, nossa máquina de testes capotou horrivelmente. O *gateway* SSL recusou-se terminantemente a trabalhar. Ao tentar acessar uma URL criptografada, o navegador mostrou uma página em branco, pura e simples. Sem mensagem de erro nem nada, só o vazio. A interface web tem erros e parece aceitar de forma arbitrária algumas configurações, enquanto recusa outras.

A documentação do Webcleaner não ajuda em nada em nenhuma dessas falhas, já que não dá nenhuma pista sobre como o programa funciona. Se combinarmos isso à configuração nada intuitiva dos filtros, chegamos à conclusão de que a interação com o usuário precisa de um retrabalho geral. Usar o programa é complicado! Nossas tentativas de falar com o iniciador do projeto foram infrutíferas: nossos porquês continuaram sem satisfação até o envio desta edição à gráfica.

Uma vantagem óbvia do Webcleaner em comparação com o Privoxy é permitir autenticação de usuários, tanto para acesso quanto para configuração.

Webcleaner: Conclusões

Embora o Webcleaner se mostre promissor, a profusão estonteante de problemas o torna uma pedra de tropeço. O maior benefício que o Webcleaner oferece, em comparação com seus concorrentes – o filtro contra vírus – falhou desgraçadamente em todos os nossos testes, incluindo aí situações práticas do dia-a-dia. Outros recursos como a compressão de imagens e páginas só ocupam tempo desnecessário de CPU e não fazem muita diferença para a máquina local. ■

Mecanismos de filtragem

As ferramentas de filtragem podem ser divididas mais ou menos nas seguintes categorias:

Filtros de URL

Os filtros de URL (*URL filters*) simplesmente comparam a URL informada no navegador com listas negras (ou brancas) mantidas localmente. Os endereços das listas podem ser completos, com domínio e nome da máquina (por exemplo: urano.sistemasolar.net) ou apenas fragmentos de URL. Os filtros de URL são usados basicamente para regular o acesso dos usuários a determinados sites. O melhor programa dessa categoria é o Squidguard: www.squidguard.org.

Vantagens

- Baixo uso de recursos
- Processamento rápido
- Fácil de configurar

Contras

- Não permite a proteção da privacidade
- Não verifica o *conteúdo* do site, apenas o endereço

Filtro de conteúdo

Nessa categoria, os sistemas lêem todo o conteúdo da página antes de mostrá-lo ao usuário. A decisão do que será mostrado e o que será bloqueado depende de vários critérios. O alcance desses critérios dependem do programa escolhido. Bloqueadores de *pop-up*, de banners e filtros de *cookies* são bastante comuns. Há alguns programas realmente inteligentes, como o DansGuardian dansguardian.org, que validam os sites e páginas baseados em pesos. Se um termo em alguma página excede um determinado "peso", regras predefinidas são aplicadas.

Prós

- Controle mais apurado do conteúdo mostrado
- Alto nível de segurança

Contras

- Dependendo da configuração, as páginas demoram para carregar
- Alto consumo de recursos do sistema
- Falsos positivos (conteúdo legítimo pode ser, mesmo assim, filtrado)
- Configuração complicada

Casos especiais

Alguns programas, como o próprio Webcleaner, possuem mecanismos adicionais, incluindo compressão de páginas e imagens e caça a vírus.

Informações

[1] Privoxy: www.privoxy.org

[2] Webcleaner: webcleaner.sourceforge.net

[3] Junkbuster: internet.junkbuster.com

[4] Regras para o Privoxy: www.neilvandyke.org/privoxy-rules