



Guia – HTTP

Objectivos:

- Estudo do protocolo HTTP:
 - a interação GET / resposta básica,
 - formatos de mensagem HTTP,
 - recuperação de grandes arquivos HTML,
 - recuperação de arquivos HTML com objetos incorporados e autenticação e segurança HTTP
- Comparação com protocolo HTTPS

A interação GET / resposta HTTP básica

⇒ Vamos começar por usar o protocolo HTTP para pedir um documento HTML (muito simples) - um que é muito curto e não contém objetos incorporados. Faça o seguinte:

- Inicie o browser.
- Inicie o Wireshark. Digite “http” (apenas as letras, não as aspas) na janela de especificação do filtro de exibição, para que apenas as mensagens HTTP capturadas sejam exibidas posteriormente na janela de listagem de pacotes.
- Espere um pouco mais de um minuto (veremos porque em breve) e, em seguida, comece a captura de pacotes Wireshark.
- Digite o seguinte no browser <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html>
- O browser deve exibir o arquivo HTML muito simples de uma linha.
- Pare a captura de pacotes.

⇒ Analisando a captura, responda às seguintes questões, colocando as respostas num ficheiro de texto:

1. Qual a versão do protocolo HTTP?
2. Quais os idiomas (se houver) que o browser indica que pode aceitar do servidor?
3. Qual é o endereço IP do seu computador? E do servidor gaia.cs.umass.edu?
4. Qual é o código de status devolvido pelo servidor para o browser?
5. Quando é que o documento HTML, que foi descarregado do servidor, foi modificado pela última vez no servidor?
6. Quantos bytes de conteúdo foram descarregados?
7. Ao inspecionar os dados brutos na janela de conteúdo do pacote, vê algum cabeçalho nos dados que não são exibidos na janela de listagem de pacotes? Em caso afirmativo, indique um.

A interação HTTP GET / resposta CONDICIONAL

⇒ A maioria dos browsers executa o armazenamento em cache do objeto e, portanto, executa um GET condicional ao recuperar um objeto HTTP. Antes de realizar as etapas abaixo, verifique se o cache do seu navegador está vazio. (Para fazer isso no Firefox, selecione Ferramentas-> Limpar histórico recente e marque a caixa Cache, ou para o Internet Explorer, selecione Ferramentas-> Opções da Internet-> Excluir



arquivo; essas ações removerão os arquivos em cache do cache do seu browser.). Agora faça o seguinte:

- Inicie o browser e certifique-se de que a cache do browser está limpo, conforme discutido acima.
- Inicie o Wireshark
- Insira o seguinte URL: <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html>
- O browser deve exibir um documento HTML de cinco linhas.
- Volte a visitar o site inserindo o URL ou actualizando.
- Interrompa a captura e digite “http” na janela de especificação do filtro de exibição.

⇒ Responda às seguintes questões:

8. Inspecione o conteúdo da primeira solicitação HTTP GET. Existe uma linha “IF-MODIFIED-SINCE” no HTTP GET?
9. Inspecione o conteúdo da resposta do servidor. O servidor retornou explicitamente o conteúdo do arquivo? Tem a certeza?
10. Agora inspecione o conteúdo da segunda solicitação HTTP GET de seu browser para o servidor. Você vê uma linha “IF-MODIFIED-SINCE:” no HTTP GET? Se sim, que informação estão no cabeçalho “IF-MODIFIED-SINCE:”?
11. Qual é o código de status HTTP e a frase retornada do servidor em resposta a este segundo HTTP GET? O servidor retornou explicitamente o conteúdo do arquivo? Explique.

Descarregando documentos Longos

⇒ Vamos ver a seguir o que acontece quando descarregamos um documento HTML longo. Faça o seguinte:

- Inicie o browser e certifique-se de que o cache do browser esteja limpo, conforme discutido acima.
- Inicie o Wireshark
- Insira o seguinte URL em seu browser <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html>
- O browser deve exibir a extensa Declaração de Direitos dos EUA.
- Pare a captura e digite "http" na especificação do filtro de exibição

⇒ Responda as seguintes questões:

12. Quantas mensagens HTTP GET o browser enviou? Que número de pacote contém a mensagem GET para o projeto de lei ou direitos?
13. Qual número de pacote no rastreamento contém o código de status e a frase associada com a resposta à solicitação HTTP GET?
14. Qual é o código de status e a frase na resposta?
15. Quantos segmentos TCP contendo dados foram necessários para transportar o único HTTP resposta e o texto da Declaração de Direitos?



Documentos HTML com objetos incorporados

⇒ Vamos ver o que acontece quando o browser descarrega um documento com objetos incorporados, ou seja, um documento que inclui outros objetos (por exemplo, imagens) que estão armazenados em outro(s) servidor(es). Faça o seguinte:

- Inicie o browser e certifique-se de que o cache do browser está limpa, como discutido acima.
- Inicie o Wireshark
- Insira o seguinte URL <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html>
- O browser deve exibir um pequeno documento HTML com duas imagens. Essas duas imagens são referenciadas no arquivo HTML de base. Ou seja, as próprias imagens não estão contidas no HTML; em vez disso, os URLs das imagens estão contidos no arquivo HTML descarregado. O browser terá que recuperar esses logotipos dos sites indicados. O logotipo do editor é obtido no site gaia.cs.umass.edu. A imagem da capa está armazenada no servidor caite.cs.umass.edu. (Esses são dois servidores web diferentes dentro de cs.umass.edu).
- Interrompa a captura de pacotes e digite "http" na janela de especificação do filtro de exibição, de forma que apenas as mensagens HTTP capturadas sejam exibidas.

⇒ Responda as seguintes questões:

16. Quantas mensagens HTTP GET o browser enviou? A qual Endereço da Internet foram esses pedidos GET enviados?
17. Você pode dizer se o browser baixou as duas imagens em série ou se elas descarregadas dos dois sites em paralelo? Explique.

Autenticação HTTP

⇒ Vamos tentar visitar um site protegido por password e examinar a sequência de mensagens HTTP trocadas por esse site. O URL http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html é protegido por password. O nome de utilizador é “wireshark-students” (sem as aspas), e a senha é “network” (novamente, sem as aspas). Aceda ao site. Faça o seguinte:

- Certifique-se de que o cache do browser está limpa, conforme discutido acima, e feche o browser. Em seguida, inicie o seu browser.
- Inicie o Wireshark
- Insira o seguinte URL no browser http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html
- Digite o nome de utilizador e a senha solicitados na caixa pop-up.
- Interrompa a captura de pacotes do Wireshark e digite "http" na janela de especificação do filtro de exibição, de forma que apenas as mensagens HTTP capturadas sejam exibidas posteriormente na janela de listagem de pacotes.

⇒ Responda às seguintes perguntas:

18. Qual é a resposta do servidor (código de status e frase) em resposta à mensagem HTTP GET inicial do seu navegador?
19. Quando o seu navegador envia a mensagem HTTP GET pela segunda vez, qual novo campo é incluído na mensagem HTTP GET?



Acesso em HTTPS

⇒ O protocolo HTTP tem vindo a ser substituído por uma nova versão com o nome HTTPS. Ao contrário do que acontece com acesso HTTP o acesso à captura não permite perceber o conteúdo da mesma. Aceda a <http://www.ua.pt/pt/estga>, fazendo o seguinte:

- Certifique-se de que a cache do browser esteja limpa, conforme discutido acima, e feche o browser. Em seguida, inicie o seu browser
- Inicie o Wireshark
- Insira o seguinte URL no browser <http://www.ua.pt/pt/estga>
- Interrompa a captura de pacotes do Wireshark e digite “http” na janela de especificação do filtro de exibição, de forma que apenas as mensagens HTTP capturadas sejam exibidas posteriormente na janela de listagem de pacotes.

⇒ Responda às seguintes perguntas:

20. Identifique o motivo que levou a que o URL final seja diferente do que digitou.
21. Identifique o qual uma vantagem do HTTPS que leva a que substitua HTTP.
22. Tendo em conta o que aprendeu relativamente às proxies HTTP identifique qual o problema desta alteração tecnológica