Laboratório com o Ethereal: HTTP

Version: 1.0

Tendo experimentado o Ethereal no laboratório de introdução, estamos agora prontos para usar o Ethereal para investigar protocolos em operação. Neste laboratório, exploraremos diversos aspectos do protocolo HTTP: a interação básica GET/resposta, formatos das mensagens HTTP, recuperação de grandes arquivos HTML, recuperação de arquivos HTML com referência a objetos, assim como a autenticação HTTP e segurança. Antes de iniciar este lab, você pode querer rever a Seção 2.2 do texto.

1. A interação básica GET/resposta do HTTP

Vamos começar a nossa exploração do http baixando um arquivo HTML bem simples – um arquivo bem curto e que não contém referência a nenhum objeto. Faça o seguinte:

1. Inicie o seu browser Web.
2. Inicie o Ethereal, como descrito no laboratório de Introdução (mas não comece ainda a capturar pacotes). Tecle "http" (apenas as letras sem as aspas) no campo de especificação do filtro de apresentação, de modo que apenas as mensagens HTTP capturadas serão apresentadas posteriormente na janela de listagem dos pacotes. (Nós estamos agora interessados apenas no protocolo HTTP e por isto não queremos ver todos os pacotes capturados).
3. Espere um pouco mais do que um minuto (veremos em breve o porquê), e depois comece a captura de pacotes do Ethereal.
4. Tecle o seguinte no seu browser http://gaia.cs.umass.edu/ethereal-labs/HTTP-ethereal-file1.html O seu browser deve apresentar o arquivo HTML bem simples de apenas uma linha
5. Pare a captura de pacotes do Ethereal.

A sua janela do Ethereal deve estar semelhante à janela apresentada na Figura 1. Se você não for capaz de rodar o Ethereal numa conexão de rede ativa, você pode baixar o registro (trace) de pacotes que foi criado quando os passos acima foram seguidos[1].

O exemplo da Figura 1 mostra na janela de listagem de pacotes que duas mensagens HTTP foram capturadas: a mensagem de GET (do seu browser para o servidor Web gaia.cs.umass.edu) e a mensagem de resposta do servidor para o seu browser. A janela de conteúdo dos pacotes apresenta detalhes da mensagem selecionada (neste caso a mensagem GET do HTTP que está destacada na janela de listagem dos pacotes). Lembre que desde que a mensagem HTTP foi transportada dentro de um segmento TCP, que foi transportado dentro de um datagrama IP, que foi transportado dentro de um Quadro (*Frame*), Ethernet, o

[1] Baixe o arquivo zip http://gaia.cs.umass.edu/ethereal-labs/ethereal-traces.zip e extraia o arquivo http-ethereal-trace-1. Os traços neste arquivo zip foram coletados pelo Ethereal rodando num dos computadores do autor enquanto foram seguidos os passos indicados neste laboratório. Uma vez que você tiver baixado o arquivo você poderá carregá-lo no Ethereal e visualizar o traço usado a opção *Open* do menu *File*, e selecionando o arquivo de traços http-ethereal-trace-1. A tela resultante deve parecer com a da Figura 1.

Ethereal apresenta também informações sobre o Quadro, Ethernet e pacotes IP e TCP. Queremos minimizar a quantidade de dados apresentados que não sejam do tipo HTTP (pois estamos agora interessados no HTTP, investigaremos os outros protocolos em laboratórios futuros). Portanto, assegure-se que as caixas no lado mais esquerdo da informação de *Frame*, Ethernet, IP e TCP tenham um sinal de "mais" (o que significa que as informações não estão apresentadas), e a linha do HTTP tenha um sinal de "menos" (o que significa que todas as informações sobre a mensagem HTTP estejam apresentadas).

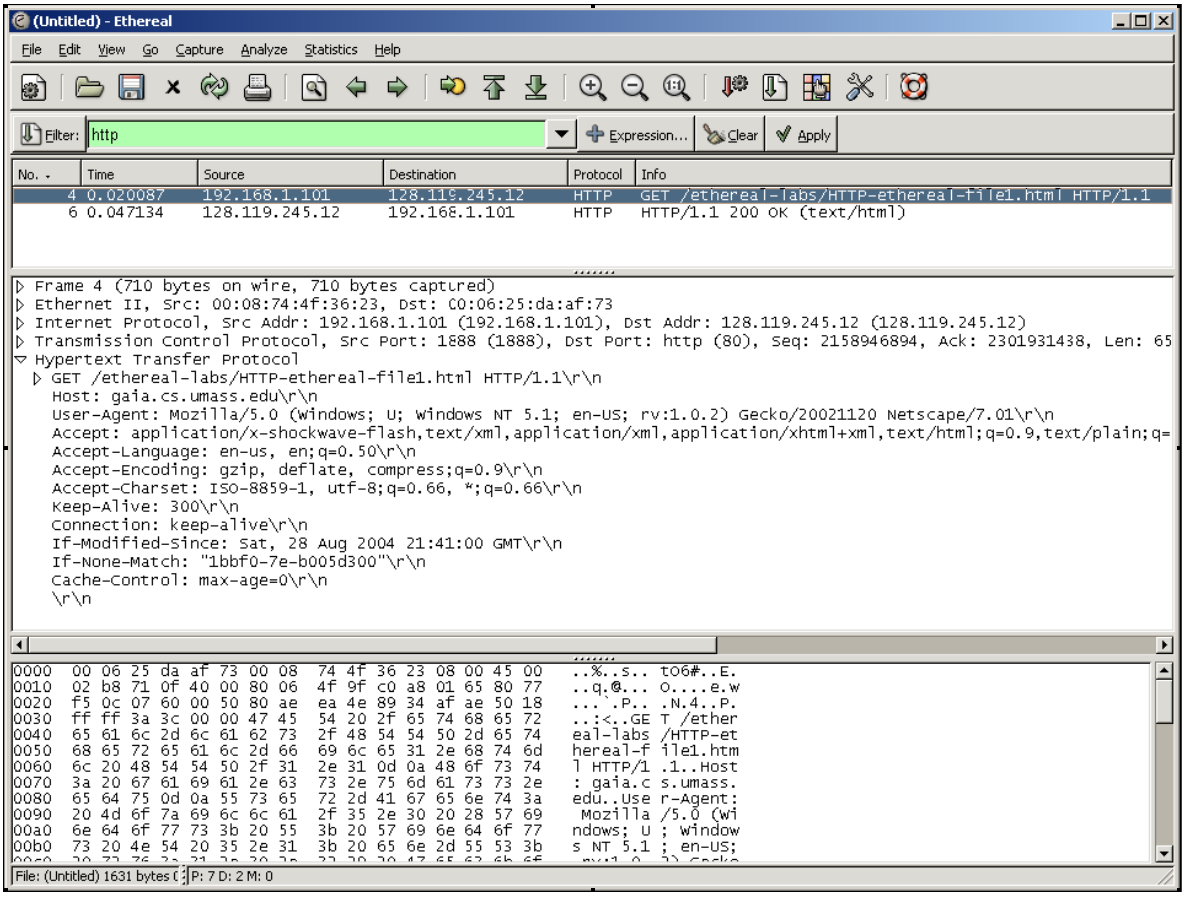

Figura 1 - Tela do Ethereal após a recuperação do arquivo http://gaia.cs.umass.edu/ethereal-labs/ HTTP-etherealfile1.html pelo seu browser

Observando as informações nas mensagens de GET e resposta do http, responda às seguintes questões. Ao responder às questões seguintes, você deve imprimir as mensagens de GET e de resposta (veja o laboratório de Introdução do Ethereal para uma explicação sobre como fazer isto) e indique onde na mensagem você encontrou a informação que responde às seguintes questões.

1. O seu browser está rodando a versão 1.0 ou 1.1 do HTTP? Que versão do HTTP está sendo executada no servidor?
2. Que linguagens (se houver alguma) o seu browser indica ao servidor que pode aceitar?
3. Qual é o endereço IP do seu computador? E o do servidor gaia.cs.umass.edu?
4. Qual é o código de status retornado do servidor para o seu browser?

5. Quando o arquivo HTML que você está recuperando foi modificado pela última vez no servidor?
6. Quantos bytes de conteúdo estão sendo retornados para o seu browser?
7. Inspecionando os dados brutos na janela de conteúdo do pacote, você vê algum cabeçalho dentro dos dados que não é apresentado na janela de listagem de pacotes? Caso veja, cite pelo menos um.

Na sua resposta à pergunta 5 acima, você pode ter ficado surpreso ao descobrir que o documento que você acabou de recuperar foi modificado alguns minutos antes de você recuperá-lo. Isto acontece porque (para este arquivo em particular), o servidor gaia.cs.umass.edu está ajustando o instante da última modificação para a hora atual, e faz isto a cada minuto. Portanto, se você esperar um minuto entre acessos, o arquivo aparecerá como tendo sido modificado recentemente, e portanto o seu browser irá recuperar uma "nova" cópia do documento.

2. Interação do GET Condicional e Resposta do http

Lembre da Seção 2.2.5 do texto, que muitos browsers web efetuam cache de objetos e depois usam o GET condicional quando estiverem recuperando um objeto HTTP. Antes de executar os passos abaixo, tenha certeza de que o cache do seu browser esteja vazio. (Para fazer isto com o Netscape 7.0, selecione *Editar->Preferências->Avançadas->Cache* e limpe a memória e cache do disco. Para o Internet Explorer, selecione *Ferramentas->Opções da Internet->Excluir Arquivos*; estas ações removerão arquivos salvos na cache do seu browser.) Agora faça o seguinte:

- Inicie o seu browser web, e assegure-se de que a cache do seu browser esteja limpa, como discutido anteriormente.
- Inicie o *sniffer* de pacotes, Ethereal,
- Tecle a seguinte URL no seu browser http://gaia.cs.umass.edu/ethereal-labs/HTTP-ethereal-file2.html. O seu browser deverá apresenta um arquivo HTML bem simples de apenas cinco linhas.
- Rapidamente entre a mesma URL no seu browser (ou simplesmente selecione o botão de atualizar do seu browser)
- Pare a captura de pacotes do Ethereal, e tecle "http" na janela de especificação do filtro de apresentação, de modo que apenas as mensagens HTTP capturadas serão apresentadas posteriormente na janela de listagem de pacotes.
- (*Nota*: Se você não for capaz de executar o Ethereal numa rede conectada, você poderá usar o traço de pacotes http-ethereal-trace-2 para responder às questões abaixo; vide nota de rodapé 1. Este traço foi gravado enquanto estavam sendo executados os passos acima num dos computadores do autor.)

Responda às seguintes perguntas:

8. Inspecione o conteúdo da primeira requisição GET do http do seu browser para o servidor. Você vê alguma linha de "IF-MODIFIED-SINCE" no GET do HTTP?
9. Inspecione o conteúdo da resposta do servidor. O servidor retorna explicitamente o conteúdo do arquivo? Como você pode afirmar isto?
10. Agora inspecione o conteúdo da segunda requisição de GET do HTTP do seu browser para o servidor. Você vê alguma linha "IF-MODIFIED SINCE" na

mensagem GET do HTTP? Em caso afirmativo, que informações seguem o cabeçalho "IF-MODIFIED-SINCE"?

11. Qual é o código de status e a frase retornada do servidor em resposta ao segundo GET do http? O servidor retorna explicitamente o conteúdo do arquivo? Explique.

3. Recuperação de Documentos Longos

Nos nossos exemplos até agora os documentos recuperados foram simples e pequenos arquivos HTML. Vamos ver agora o que acontece quando baixamos um arquivo HTML longo. Faça o seguinte:

- Inicie o seu browser web, e tenha certeza de que a cache do browser está vazia, como discutido acima.
- Inicie o sniffer de pacotes Ethereal.
- Tecle a seguinte URL no seu browser http://gaia.cs.umass.edu/ethereal-labs/HTTP-ethereal-file3.html. O seu browser deveria apresentar a bem longa "US Bill of Rights".
- Pare a captura de pacotes e tecle "http" na janela de especificação do filtro de apresentação, de modo que serão apresentadas apenas as mensagens capturadas do http.
- (*Nota*: Se você não for capaz de executar o Ethereal numa rede conectada, você poderá usar o traço de pacotes http-ethereal-trace-3 para responder às questões abaixo; vide nota de rodapé 1. Este traço foi gravado enquanto estavam sendo executados os passos acima num dos computadores do autor.)

Na janela de listagem de pacotes, você deveria ver a sua mensagem de GET do http, seguida de uma resposta de múltiplos pacotes para o seu pedido GET do HTTP. Esta resposta de múltiplos pacotes merece uma pequena explicação. Lembre da Seção 2.2 (veja a Figura 2.9 no texto) que a mensagem de resposta do HTTP consiste de uma linha de status, seguida por linhas de cabeçalho, seguida por uma linha em branco, seguida pelo corpo da entidade. No caso do nosso GET do HTTP, o corpo da entidade na resposta e *todo* o arquivo HTML solicitado. No nosso caso aqui, o arquivo HTML é bem grande, e o tamanho de 4500 bytes é muito grande para caber em um único pacote TCP. A mensagem de resposta do http é, portanto, quebrada em diversos pedaços pelo TCP, sendo cada pedaço contido num segmento TCP diferente (veja Figura 1.22 no texto). Cada segmento TCP é registrado como um pacote diferente pelo Ethereal, e o fato de que a mensagem de resposta HTTP foi fragmentada entre múltiplos pacotes TCP é indicado pela frase "Continuation" apresentada pelo Ethereal. Nós chamamos a atenção de que não existe nenhuma mensagem de "Continuation" no http!

Responda às seguintes questões:

12. Quantas mensagens de pedidos GET do HTTP foram enviadas pelo seu browser?
13. Quantos segmentos TCP de dados foram necessários para transportar a mensagem de resposta http?
14. Qual é o código de status e a frase associada com a resposta ao pedido GET do HTTP?

15. Há alguma linha de status HTTP nos dados transmitidos associados com uma "Continuação" induzida pelo TCP?

4. Documentos HTML com Objetos

Agora que vimos como o Ethereal apresenta o tráfego de pacotes capturados para grandes arquivos HTML, podemos ver o que acontece quando o seu browser recupera um arquivo com objetos referenciados, i.e., um arquivo que inclui outros objetos (no exemplo abaixo, arquivos de imagens) que são armazenados em outro(s) servidor(es).

Faça o seguinte:

- Inicie o seu browser web, e assegure-se que a cache do seu browser foi liberada, como discutido acima.
- Inicie o pacote Ethereal.
- Entre a seguinte URL no seu browser http://gaia.cs.umass.edu/ethereal-labs/HTTP-ethereal-file44.html. O seu browser deveria apresentar um arquivo HTML curto com duas imagens. Estas duas imagens são referenciadas no arquivo base HTML. Ou seja, as imagens por si mesmas não estão contidas no HTML. Como discutido no livro-texto, o seu browser terá que recuperar as logomarcas a partir dos sítios web indicados. A logomarca da nossa Editora é recuperada a partir do sítio web www.awl.com. A imagem da capa do nosso livro está armazenada no servidor manix.cs.umass.edu.
- Pare a captura de pacotes do Ethereal, e tecle "http" na janela de especificação do filtro de apresentação, de modo que apenas as mensagens HTTP capturadas serão apresentadas.
- (*Nota*: Se você não for capaz de executar o Ethereal numa rede conectada, você poderá usar o traço de pacotes http-ethereal-trace-4 para responder às questões abaixo; vide nota de rodapé 1. Este traço foi gravado enquanto estavam sendo executados os passos acima num dos computadores do autor.)

Responda às seguintes perguntas:

16. Quantas mensagens de pedidos GET do HTTP foram enviadas pelo seu browser? Para quais endereços Internet foram enviadas estas solicitações GET?
17. Você pode dizer se o seu browser baixou as duas imagens serialmente, ou se elas foram baixadas a partir dos dois sítios web em paralelo? Explique.

2.5 Autenticação do http

Agora vamos tentar visitar um sítio web que seja protegido por senha e examinar a seqüência de mensagens http trocadas para este sítio. A URL http://gaia.cs.umass.edu/ethereal-labs/protected_pages/HTTP-ethereal-file5.html é protegida por senha. O usuário "eth-students" (sem as aspas), e a senha é "networks" (mais uma vez, sem as aspas). Portanto, vamos acessar este sítio protegido por senha. Faça o seguinte:

- Assegure-se de que a cache do seu browser esteja limpa, como discutido acima, e feche o seu browser. Em seguinda inicie o seu browser.
- Inicie o sniffer Ethereal

- Tecle a seguinte URL no seu browser http://gaia.cs.umass.edu/ethereal-labs/protected_pages/HTTP-ethereal-file5.html Digite o usuário e a senha no quadro que se abrirá.
- Pare a captura de pacotes, e entre "http" na janela de especificação e filtragem, de modo que apenas as mensagens HTTP capturadas serão apresentadas posteriormente na janela de listagem dos pacotes.
- (*Nota*: Se você não for capaz de executar o Ethereal numa rede conectada, você poderá usar o traço de pacotes http-ethereal-trace-5 para responder às questões abaixo; vide nota de rodapé 1. Este traço foi gravado enquanto estavam sendo executados os passos acima num dos computadores do autor.)

Agora vamos examinar a saída do Ethereal. Você poderá querer antes ler a respeito de autenticação revendo o material de fácil leitura sobre "Arcabouço para Autenticação de Acesso do HTTP" no sítio http://frontier.userland.com/stories/storyReader$2159

Responda às seguintes perguntas:

18. Qual é a resposta do servidor (código de status e frase) em resposta a mensagem HTTP GET a partir do seu servidor?
19. Quando o seu browser envia uma mensagem HTTP GET pela segunda vez, que no campo. Que novo campo está incluído na mensagem GET do HTTP?

O nome de usuário (eth-students) e senha (network) que vocês entram são codificados na cadeia de caracteres (ZXRoLXN0dWRlbnRzOm5ldHdvcmtz) após o cabeçalho `"Authorization: Basic"` na mensagem GET HTTP do cliente. Enquanto pode parecer de que o seu usuário e senha estejam criptografados, eles estão simplesmente codificados num formato conhecido como formato Base64. O nome do usuário e a senha *não* estão criptografados! Para ver isto, vá para o sítio http://www.securitystats.com/tools/base64.php tecle o string codificado em base64 ZXRoLXN0dWRlbnRzOm5ldHdvcmtz e pressione o botão de decodificação. *Voila!* Você traduziu da codificação Base64 para a codificação ASCII, e portanto deveria ver tanto o seu nome de usuário como a senha! Dado que qualquer um pode obter uma ferramente como Ethereal e capturar os pacotes (inclusive os dos outros) que passam pelo seu adaptador de rede, e qualquer um pode traduzir de Base64 para ASCII (você acabou de fazê-lo!), deveria estar claro para você de que simples senhas em sites Web não são seguros a não ser que sejam tomadas medidas adicionais.

Mas, não tema! Como veremos no Capítulo 7, há formas de tornar o acesso Web mais seguro. No entanto, necessitaremos de algo que vai além do arcabouço básico de autenticação do HTTP.