



Redes de Computadores II

O Protocolo HTTP

Camada de Aplicação

Prof. Elias P. Duarte Jr.

Universidade Federal do Paraná (UFPR)

Departamento de Informática

www.inf.ufpr.br/elias/redes

Sumário da Aula de Hoje

- Nesta aula vamos estudar um dos principais protocolos da camada de aplicação da Internet
- A Web & O Protocolo HTTP

A Web: Importância

- Não há como super-estimar a importância da Web
- Sem sombra de dúvidas a aplicação mais importante da Internet
- Base de todo o comércio eletrônico, governo eletrônico, etc. etc. etc.
- Se a Internet representou uma revolução para a humanidade, foi em especial através da Web

A Web: Componentes

- Uma possível definição da Web é que ela consiste de um conjunto de documentos interligados
- Os documentos são as páginas Web e estão acessíveis através da Internet
- O conceito já existia antes da Web: hiper-texto e hiper-mídia
- O cliente da Web é o browser, que comunica com um servidor Web para obter e mostrar uma página Web

A Página Web

- Uma página Web é um documento formatado de acordo com a linguagem HTML
 - *HyperText Markup Language*
- A página consiste de texto entremeado por *tags*, que indicam comandos para a exibição da página
 - `<exemplo-tag> ... </exemplo-tag>`

URL: *Universal Resource Locator*

- O nome, identificador de uma página Web é chamado de URL (*Universal Resource Locator*)
- A URL é definida dentro de um contexto mais amplo, de identificadores universais:
 - URI (*Universal Resource Identifier*)
- A URL começa com uma indicação do protocolo usado para acessar para página, seguido de //
- Os mais comuns: http:// e https://
- O protocolo é também chamado de “esquema”

URL no Esquema HTTP

- O formato de uma URL do esquema HTTP é o seguinte [colchetes indicam opcionais]:

`http: // nome-host [:porta] /path [; parâmetros] [? consulta]`

- A porta padrão HTTP é a porta 80, o opcional é usado apenas para especificar outra porta
- O *path* indica onde o servidor Web encontra a página solicitada, por exemplo a página seguinte está no host www.inf.ufpr.br e diretório *fundresd* do usuário elias:
- <https://www.inf.ufpr.br/elias/fundresd>

URLs: Absoluta e Relativa

- Quando se especifica com o host a URL é dita absoluta
- <https://www.inf.ufpr.br/elias/fundresd>
- Uma URL relativa é especificada depois do host ser conhecido:
- </elias/fundresd>

Formato Geral da Página HTML

- O protocolo HTML versão 2 é definido no RFC1866
- Hoje a versão mais usada é a 5 (>90% da Internet)
- Um exemplo de página com texto e 1 link, que é definido por uma “âncora” - tags `<A>...`

```
<HTML>
```

```
  O professor responsável pela disciplina é
```

```
  <A HREF="http://www.inf.ufpr.br/elias">Elias P.  
Duarte Jr.</A>
```

```
</HTML>
```

O Protocolo HTTP

- HTTP: *HyperText Transfer Protocol*
- O protocolo usado para o browser comunicar com o servidor Web
- Um protocolo sobre TCP, porta 80
- Consiste de sequências de requisições/respostas
- *Stateless*: cada requisição HTTP é auto-contida, o servidor não mantém informações sobre requisições anteriores

O Protocolo HTTP -cont.

- As informações fluem tanto do servidor Web para o browser, como do browser para o servidor Web (exemplo deste caso: formulário preenchido)
- “Negociação” da sessão: o browser e o servidor Web podem entrar em acordo quanto a diversos aspectos, por exemplo codificação de caracteres
- *Caching*: o próprio browser mantém uma cache das páginas obtidas, mas contacta o browser para saber se houve alteração
- Proxy Web: definida como intermediária, funciona como cache

O Comando HTTP GET

- O principal comando definido no PDU do HTTP é aquele usado pelo browser para solicitar uma página ao servidor Web: GET
- Transmitido como texto normal, os campos são:
`GET https://www.inf.ufpr.br/elias/fundresd HTTP/1.1`
 - GET é o nome do comando
 - em seguida vem a URL
 - Depois a versão do protocolo HTTP: história parecida com a do IP → versão 1.1 mais usada, versão 2 pouco adotada

O Comando HTTP GET cont.

- A execução do GET causa o estabelecimento de uma conexão TCP
- Depois que a conexão está estabelecida, basta informar o *path* na URL:

```
GET /fundresd HTTP/1.1
```

- GET é o nome do comando
- em seguida vem a URL
- Depois a versão do protocolo HTTP: história parecida com a do IP → versão 1.1 mais usada, versão 2 pouco adotada

GET: Mensagem de Erro

- O que acontece se o browser enviar uma solicitação de uma página inexistente ao servidor web?
- Este retorna uma mensagem de erro, em uma página HTML:

```
<HTML>
```

```
<HEAD><TITLE>404 Bad Request</TITLE></HEAD>
```

```
<BODY>
```

```
<H1>Requisição incorreta!</H1>
```

```
A página Web solicitada não foi encontrada.
```

```
</BODY>
```

```
</HTML>
```

Conexões TCP Persistentes

- Originalmente o protocolo HTTP fazia múltiplas conexões TCP para cada página
 - Uma conexão para baixar cada figura, cada elemento, etc.
- Hoje é aberta 1 única conexão, e todos os elementos da conexão vão sendo transmitidos ali mesmo: múltiplas vantagens!
- O HTTP informa o número de bytes de cada elemento antes de transmitir
 - Em páginas dinâmicas é impossível saber! Fecha conexão
`Connection: close`

Header do PDU HTTP Resposta

- O header to PDU HTTP da resposta consiste de uma série de linhas de texto, seguidas de 1 linha em branco e em seguida o item sendo transferido
- O HTTP tem uma série de elementos para definir meta-informações, por exemplo:

`Content-Length: 43`

`Content-Language: pt`

`Content-Encoding: ascii`

`<HTML>Um exemplo simples de conteúdo.</HTML>`

Negociação de Parâmetros

- Diversos parâmetros de naturezas diversas, por exemplo:
 - especificar se a conexão vai ser autenticada ou não
 - formatos permitidos para elementos gráficos (jpg...)
 -
- A negociação pode ser iniciada tanto pelo browser como pelo servidor Web
- Alternativamente, o browser pode pedir as opções disponíveis ao servidor Web
- O browser pode simplesmente informar o que pode receber:
 - Accept-Encoding: ascii**
 - Accept-Language: pt**
 - ...

Requisições Condicionais

- O browser pode especificar requisitos para o servidor Web enviar ou não enviar a página
- Por exemplo considere o seguinte comando enviado junto com um GET:

```
If-modified-since: Thu, 09 May 2024  
07:00:00 GMT
```

- O servidor só envia a página solicitada se ela foi modificada desde 9 de maio de 2024

A Proxy Web

- As proxies Web são muito importantes pois têm duas vantagens:
 - Reduzem o tempo de resposta para o cliente
 - Reduzem a carga no servidor
- As proxies Web não são transparentes: devem ser configuradas explicitamente nos browsers
- A proxy em si deve ser configurada para obter e atualizar o conteúdo desejado
- O protocolo HTTP inclui toda uma série de especificações para o funcionamento de proxies Web (autenticação,...)

Web Caches

- Uma cache na Web tem as mesmas vantagens da Proxy Web: reduz latência e carga
- Além de serem usadas nas proxies, as caches estão presentes nos próprios browsers
- A pergunta chave: quanto tempo manter uma página em cache?
 - se ficar por tempo demais: a página fica “obsoleta” (“stale”)
 - se ficar pouco tempo: causa tráfego desnecessário
- O tempo de expiração é configurável, depende também do espaço da cache e a frequência de uso da Web

Conclusão

- Estudamos hoje um protocolo de aplicação
- A aplicação mais importante da Internet: a Web
- Documentos hipermídia especificados com HTML
- Endereços universais especificados como URL
- Estudamos elementos do HTTP
- Requisição GET, formatos de header exemplo
- Conexões TCP persistentes
- Negociação de parâmetros
- Web Proxy & *Caching*

Obrigado!

Lembrando: a página da disciplina é:
<https://www.inf.ufpr.br/elias/redes>