



# Conceitos de Sistemas Informáticos

## Módulo 3 – Comunicações por Computador

*António Costa <costa@uminho.pt>*

*Grupo de Comunicações por Computador  
Departamento de Informática  
Universidade do Minho*



Horário de Atendimento:

- Quarta-Feira, 15h – 18h
- Terça-Feira, 10h – 13h



## Objectivos

- Abordar de forma genérica alguns conceitos de suporte às Redes de Computadores (*Internet*)
  - Forma de funcionamento
  - Aplicações mais utilizadas
- No decorrer de LESI existirão cadeiras onde serão aprofundados conhecimentos mais técnicos sobre a Área das **Comunicações por Computador**
  - *Fundamentos das Telecomunicações (FT), Comunicações de Dados e Redes (CDR), Comunicações por Computador I (CCI), Comunicações por Computador II (CCII), Sistemas Telemáticos (ST)*



# AVISOS

- Laboratórios do DI a usar:
  - *Terça* (11 – 13h) – **DI 1.09**
  - *Quarta* (14 – 16h) – **DI 0.03**
  - *Quarta* (17 – 19h) – **DI 0.03**
  - *Quinta* (16 – 18h) – **DI 1.04**
  - *Sexta* (14 – 16h) – **DI 1.04**
- Três trabalhos práticos,
  - Como há feriado(s) na quarta-feira, e semana académica, é necessário proceder a alguns ajustes:

1. Os alunos dos turnos de quarta-feira podem ter de frequentar duas aulas práticas na última semana:
  - o seu turno habitual das quartas-feiras, e,
  - um turno adicional à escolha (terça, quinta ou sexta)
2. Os alunos dos turnos de terça, quinta e sexta estão dispensados da última semana de aulas práticas, se já tiverem assistido a três aulas, a menos que tenham trabalhos em atraso...



# AVISOS

segunda-feira	terça-feira	quarta-feira	quinta-feira	sexta-feira
29 de Abril	30	1 de Maio	2	3
5	7	8	9	10
13	14	15	16	17
20	21	22	23	24



## Conteúdo

- Introdução às comunicações. Operação da rede através de camadas de protocolos. Noções sobre encapsulamento, endereçamento e nomeação. Definição de protocolo. A pilha protocolar TCP/IP
- Internet. Acesso à Internet. Requisitos e soluções possíveis. Fornecedores de serviço. Protocolos de acesso PPP e SLIP.
- Categorias de aplicações. Correio Electrónico e listas de distribuição. Protocolos IMAP e POP.
- Segurança em Sistemas Telemáticos. Criptografia simétrica e assimétrica. Procedimentos para operações de assinatura digital e confidencialidade. Software PGP.
- Protocolos e linguagens associadas ao WWW: HTTP e HTML. Proxy e caching e a sua importância. Serviços de pesquisa avançada. Segurança e Extensões.
- Comunicações Audio/Video na Internet. Mecanismo de transmissão: tecnologia Multicast. Mbone e Aplicações.



## Introdução às Comunicações

**Um Modelo**

- A comunicação
- A comunicação de *dados*
- A comunicação de dados *por computador*

GCOM-DI-UM 7 CSI 2002, A.Costa

**Objectivos**

- Partilha de recursos
  - programas, dados, equipamentos
- Partilha de carga
  - trata-se de uma caso particular do anterior...
- Potente meio de comunicação entre pessoas
  - partilha de ideias, edição simultânea de documentos, troca de documentos, reuniões à distância, etc...
- Melhor fiabilidade
  - havendo alternativas, menor probabilidade de falha
- Poupar dinheiro!
  - equipamentos de pequeno porte, investimento gradual...

GCOM-DI-UM 8 CSI 2002, A.Costa



## Aplicações

Algumas, usadas em redes de larga escala, podem ter efeitos na sociedade como um todo

### Três exemplos:

- acesso a “bases de dados” remotas
  - bibliotecas on-line, jornais, papers científicos
  - *homebanking*, reservas para espectáculos e transportes (talvez até compra de bilhetes!)
- comunicações de valor acrescentado
  - correio electrónico multimédia (som e vídeo), videotelefone, videoconferências, foruns de discussão (à escala do planeta!)
- acesso a programas remotos
  - licença de uso do software em vez de compra... Actualizações



## Aplicações

Outras, fazem mais sentido no contexto de uma rede local

### Três exemplos:

- Utilização de uma *impressora de rede*
- Utilização do disco de um *servidor de disco*
- Partilha de monitores gráficos de alta resolução e de processadores mais potentes para cálculos



## Conceitos

### Redes locais e de longa distância

- LAN (Local Area Network) - um edifício ou *campus*
- MAN (Metropolitan Area Network) - uma cidade...
- WAN (Wide Area Network) - um ou mais países...

⇒ Há diferenças na tecnologia (veremos mais tarde!)

### LANs

- são redes privadas, pertença de uma empresa ou instituição
- permitem *débitos elevados*, da ordem dos Mbit/s e *Gbit/s*. (10/100 Mbps)
- Distâncias curtas (alguns km apenas)
- Suporte para muitos sistemas (centenas!)
- Baixas taxas de erros (elevada fiabilidade)
- distinguem-se pela *tecnologia de transmissão* e pela *topologia*



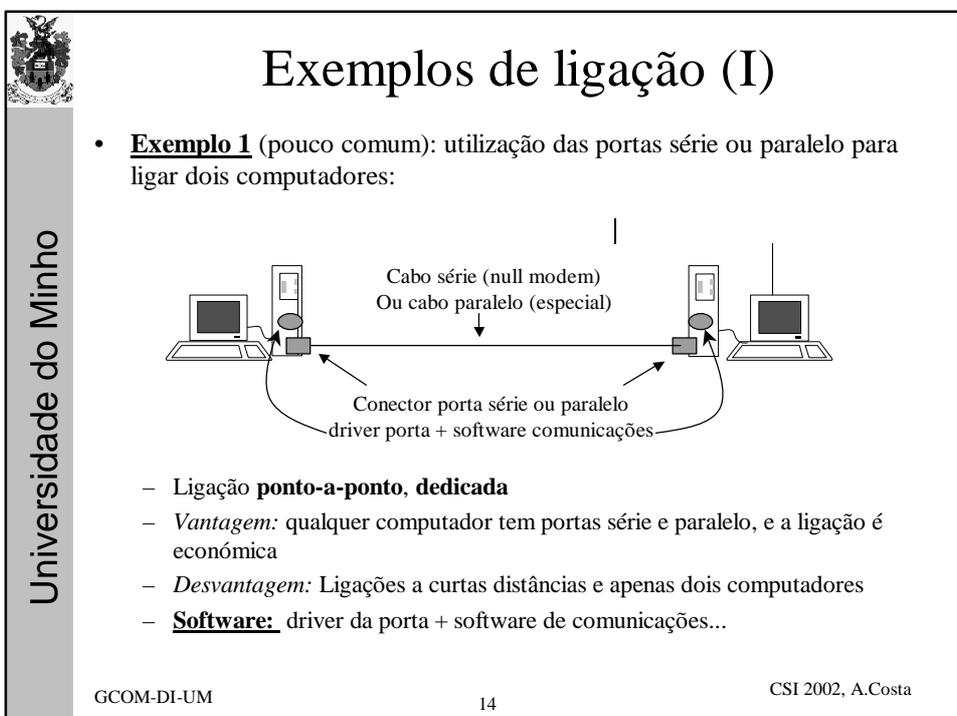
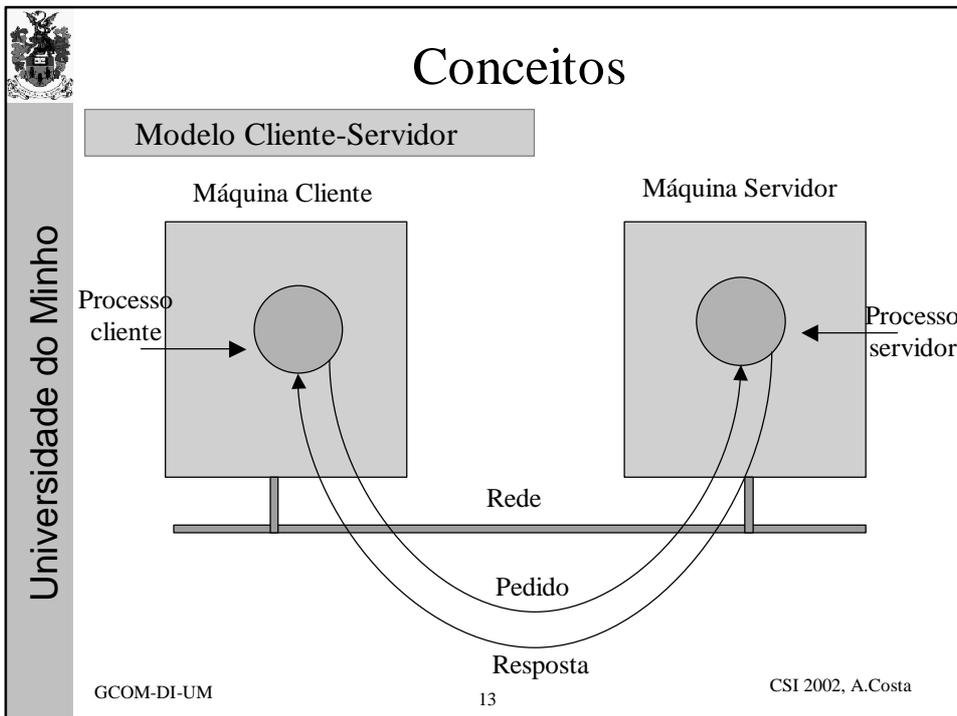
## Conceitos

### MANs

- podem ser privadas ou públicas, a distâncias de alguns kms (uma cidade)
- pode normalmente transportar dados e voz de forma integrada
- tecnologicamente não difere muito das redes locais
- existe uma tecnologia normalizada para o efeito: DQDB

### WANs

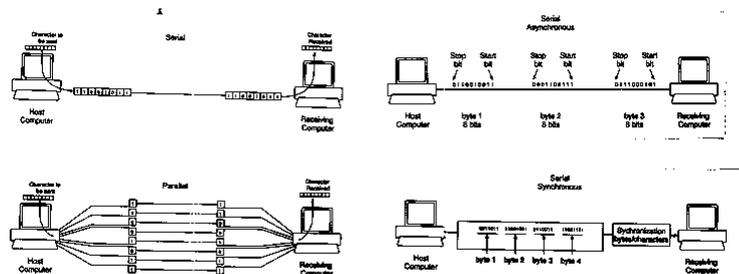
- cobre vastas áreas geográficas, normalmente países ou continentes
- normalmente constituída por um conjunto de *linhas de transmissão* e *routers* (*equipamentos de interligação*)
- dois *routers* podem comunicar entre si, mesmo sem nenhuma *linha* que os interligue directamente, usando outros como intermediários:
  - um *router* intermédio, deve receber e armazenar pacotes de dados vindos de uma linha antes dos reenviar, se for caso disso...
- são por isso designadas de redes *store-and-forward*, redes *ponto-a-ponto* ou simplesmente redes de *comutação de pacotes*





## Modos de transmissão

- Transmissão em série (síncrona e assíncrona) e em paralelo



## Modos de transmissão

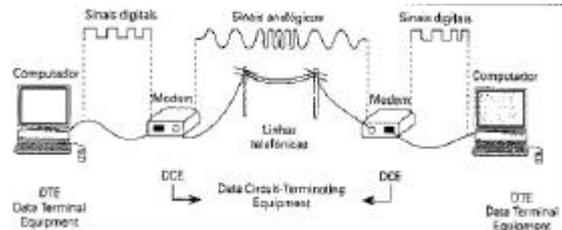
- A transmissão em paralelo só costuma ser usada internamente ao computador (*barramentos* ou *bus*), ou na ligação a periféricos a curtas distâncias (ex: impressoras)
  - a grande quantidade de fios necessários eleva astronomicamente o custo dos cabos!
  - há limitações eléctricas!...
- Os PC's normalmente possuem uma ou mais portas série (COM1, COM2, ..) e portas paralelo (LPT1, LPT2, ...)

De uma forma geral a transmissão série síncrona utiliza melhor o canal que a transmissão série assíncrona



## Exemplos de ligação (II)

- **Exemplo 2:** utilização de um modem e da rede telefónica



- Ligação **ponto-a-ponto**, **dedicada**, transmissão **série**, normalmente **assíncrona**, a velocidades de transmissão relativamente baixas... (da ordem dos 28, 36 ou 45 Kbps)
- A rede telefónica tem uma cobertura geográfica inigualável
- Rede telefónica foi preparada para transmissão *analógica* de dados *analógicos*, havendo necessidade de *modular e desmodular* - **MODEM**
- Modem **externo** – ligado a porta série, ou modem **interno** – ligado ao barramento do computador: é sempre visto pelo sistema como uma porta série...
- **Software:** *Driver* + software de comunicações...



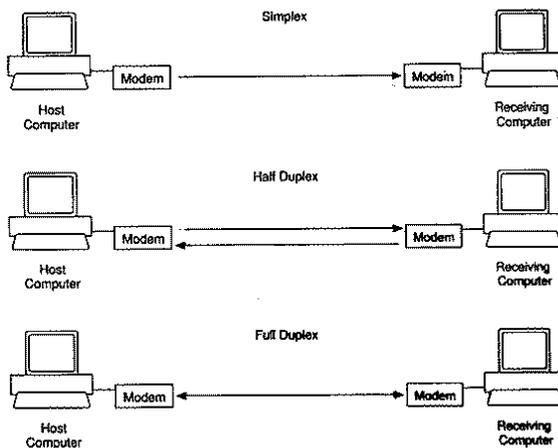
## MODEMs

- Podem ser externos (ligados à *porta série*) ou internos (uma placa a encaixar no *barramento*).
- Em princípio são equivalentes, embora os internos não estejam dependentes do controlador da porta série...
- Nos computadores portáteis ligam-se no interface PCMCIA
- Taxas de transmissão mais frequentes: 1200bps (V.22), 9600bps (V.32), 14400bps (V.32bis), 28800bps (V.34) e 33600bps (V.34Enhanced)



# Modos de transmissão

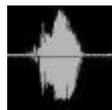
- *Simplex, Half-Duplex, Duplex*



# Conceitos - Sinais analógicos e digitais

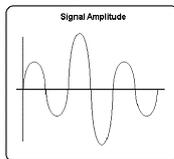
## Sinais analógicos

- variação contínua no tempo...

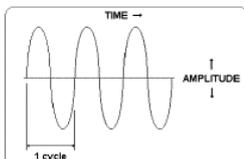


"Hello"

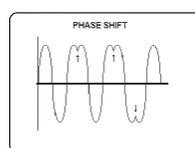
## Amplitude



## Frequência

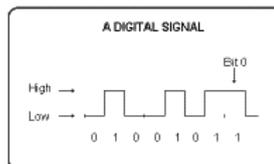


## Fase



## Sinais Digitais

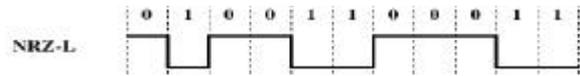
- apenas dois estados:  
ON ou OFF, 0 ou 1, etc.



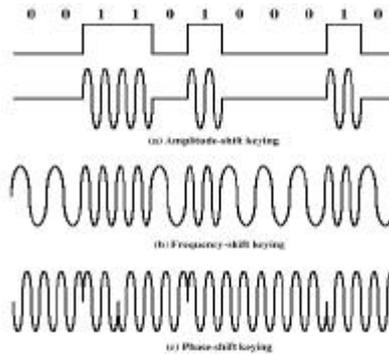


# Sinais analógicos e digitais

- Técnicas de **codificação**



- e de **modulação**



– Modulação em *amplitude, frequência e fase*



# Conceitos

Baud Rate (*baud*)

- Número de variações que ocorrem no sinal em cada segundo... Para sinais digitais, 20 Hz correspondem a 20 baud...

Bits por segundo (*bps*)

- Pode ser o mesmo que baud, se uma variação no sinal representar um bit...

Taxa de transmissão

- Medida da quantidade de informação que pode ser enviada por um canal por segundo... (normalmente em bps)

Largura de Banda

- intervalo de frequências de um canal. A taxa de transmissão máxima depende desta diferença entre a freq. máxima e mínima... (ex canal voz: 3.1Kz)

Universidade do Minho

## Exemplos de ligação (III)

- **Exemplo 3:** ligação a uma rede local do tipo **Ethernet**

- Ligação **multiponto, partilhada**, velocidades de transmissão elevadas... (10 Mbps no exemplo, 100Mbps ou 1Gbps noutros casos)
- Como estão muitos computadores ligados ao mesmo meio físico, é necessário **regras de acesso ao meio físico**, que garantam igualdade de oportunidades para todos!
- Sendo soluções **standard** são também normalmente de baixo custo
- Facilidade de acrescentar e remover sistemas da rede;
- **Hardware:** a placa de rede depende do tipo de rede (**Ethernet, Token Ring, FDDI, etc.**) e do *barramento* interno onde vai ser encaixada (**PCI, ISA, EISA, etc**)
- **Software:** *Driver* da placa de rede (do fabricante) + software de comunicações...

GCOM-DI-UM
23
CSI 2002, A.Costa

Universidade do Minho

## Rede local Ethernet

- Os dados são transmitidos em pacotes ou *frames*:

Preâmbulo	Endereço destino	Endereço origem	Tipo	Dados	Sequência de controlo
8 bytes	6 bytes	6 bytes	2 bytes	de 46 a 1500 bytes	4 bytes

- O preâmbulo permite que o receptor se *sincronize* com o emissor
- Cada sistema tem um endereço único de 48 bits, atribuído pelo fabricante da placa, que em princípio, não é possível alterar
- Cada *frame* contém o endereço do emissor e do receptor;
- O campo *tipo* é uma espécie de etiqueta que indica que dados são transportados em cada *frame*.
- A sequência de controlo permite detectar erros de transmissão!

GCOM-DI-UM
24
CSI 2002, A.Costa

Universidade do Minho

## Redes locais a dois níveis

Lógico

- Pacotes de dados e sua estrutura, endereços, correcção de erros, acesso ao meio físico para recepção e envio...

Físico

- Cabos e tipos de cabos, conectores e suas dimensões, placas de *interface*, sinais a transmitir, codificação ou modulação dos sinais...

Estes dois níveis podem suportar as aplicações dos utilizadores

Serão suficientes estas duas *camadas*?

GCOM-DI-UM
25
CSI 2002, A.Costa

Universidade do Minho

## Exemplos de ligação (IV)

- Exemplo 4:** interligação de redes, utilizando **routers**

Routers, ligados por linha dedicada, série, síncrona ponto - a - ponto

Rede local, multiponto, Ethernet

Rede telefónica

modem

Acesso individual

Outras LANs

GCOM-DI-UM
26
CSI 2002, A.Costa



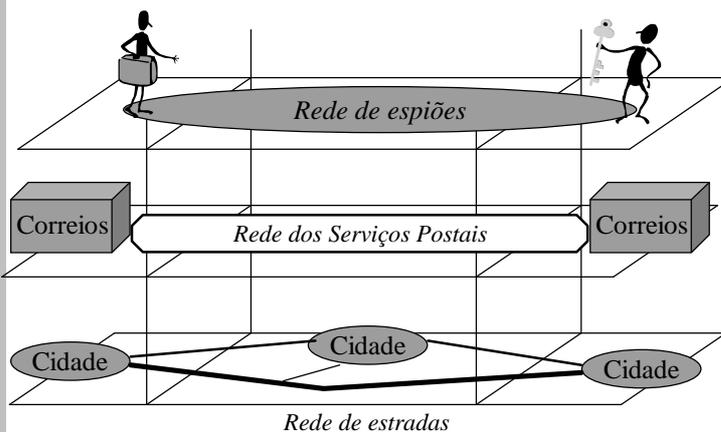
## Redes alargadas

- Todos podem comunicar com todos, embora não hajam ligações directas entre todos os equipamentos como acontece nas redes locais...
- Há nós ***intermediários***, com *várias ligações*...
  - os equipamentos intermediários designam-se por ***routers***...
- Os intermediários necessitam de escolher rotas! (*tomar decisões de encaminhamento*)
- Em caso de falha, é preciso escolher rotas alternativas:
  - só com uma *visão global ou parcial da rede*
- São por isso muito mais complexos e caros!



## Redes sobre outras Redes

- Um exemplo, com três *camadas*:





## Redes *sobre* outras Redes

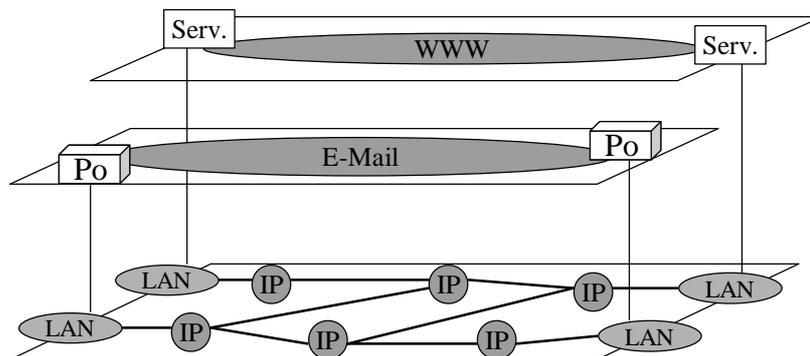
- Reflexões:

- A rede de estradas é constituída por *estradas* (de vários *tipos*) que ligam cruzamentos (ou *nós!*)
- À beira das estradas, ou mesmo dos cruzamentos, moram os *utilizadores* nas suas *casas*
- A rede postal *funciona sobre* a rede viária, e interliga postos que *armazenam e reenviam* correio...
- A rede postal é uma *rede lógica* sobre a *rede física* das estradas (e pode usar outras redes, como a ferroviária)
- Não tem de existir um Posto dos Correios em todos os cruzamentos, nem sequer em todas as cidades!
- Na rede final, no topo, estão os *espiões (utilizadores)*
- Trata-se de uma comunidade com regras próprias, que usa esquemas próprios de *cifragem* da informação
- Comunicam usando várias redes de suporte: a rede postal, ou mesmo a rede viária directamente...



## Redes *sobre* outras Redes

- Três planos:





# Conceitos

- Isto demonstra alguns conceitos, válidos também nas redes de computadores:

- a maioria das *redes* funcionam umas sobre as outras

... numa estrutura hierárquica *por camadas*

N
...
1

- uma dada *rede* funciona sobre várias outras, embora seja necessário *interligá-las!*

Interligação de LANs e WANs distintas



**Equipamentos:**  
 \* repetidores  
 \* *bridges*  
 \* *routers*



# Conceitos

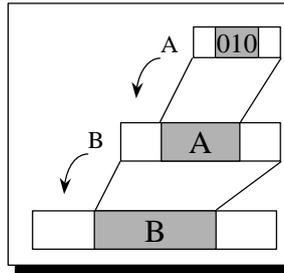
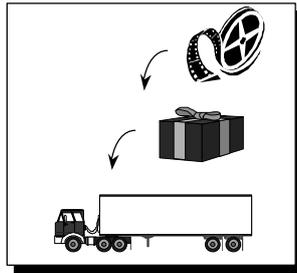
- Interligação de redes





## Conceitos

- cada *camada* da rede adiciona sucessivamente os seus próprios *contentores de informação*:

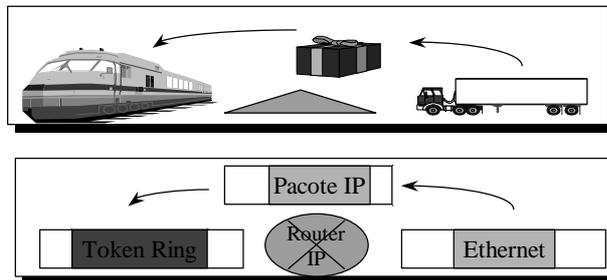


... o que se designa por *Encapsulamento*



## Conceitos

- quando os dados mudam para uma rede distinta, de rede para rede, mudam também de *contentor* para *contentor*:



Necessidade de um denominador comum: o pacote IP  
(*Internet Protocol*)

Universidade do Minho

## Conceitos

– Cada *camada* usa os seus *endereços* próprios, embora baseados nos endereços das camadas de suporte:

- Os endereços, para além do código postal (dos correios), incluem também o nome da cidade e o nome da rua...
- Numa mesma casa podem morar vários indivíduos, pelo que na morada final se tem de incluir o nome do *utilizador*!

**Endereço completo: Av. X + Casa Y + Morador Z**

GCOM-DI-UM 35 CSI 2002, A.Costa

Universidade do Minho

## Conceitos

- *Endereçamento*

- Exemplo de endereço IP: *193.136.16.254*
- Exemplo de nº porta: *80* (porta do servidor WWW!)

**Endereço completo (ip+porta): 193.136.16.254:80**

GCOM-DI-UM 36 CSI 2002, A.Costa



## Arquitectura por camadas

- Modelo de referência OSI (*Open Systems Interconnection*):
  - a comunicação é demasiado complexa para ser monolítica...
  - define 7 (sete!) camadas, independentes:
    - *Aplicação, Apresentação, Sessão, Transporte, Rede, Ligação, Física*
  - cada camada usa serviços da camada inferior e presta serviços à camada superior
  - Vantagens:
    - cada camada pode evoluir separadamente
    - aplicações mais pequenas e mais rápidas



## Modelo de referência OSI

Aplicação	Aplicações e serviços distribuídos
Apresentação	Conversão entre formatos utilizados na representação da informação
Sessão	Estabelecimento de conexões entre aplicações
Transporte	Transferência de informação entre dois pontos de forma fiável
Rede	Comunicação fim-a-fim, endereçamento, independência relativamente às camadas inferiores, encaminhamento
Ligação	Regras para a ligação entre dois pontos, detecção e correcção de erros, transmissão de blocos de bits
Física	Meio de transmissão, conectores, formato dos sinais, técnicas de modulação, transmissão de cadeias de bits

OSI - Open Systems Interconnection  
 ISO - International Standards Organization



# Protocolos

## Protocolo:

Conjunto de regras (*sintáticas, semânticas e temporais*) ou convenções que regulam a comunicação entre duas entidades

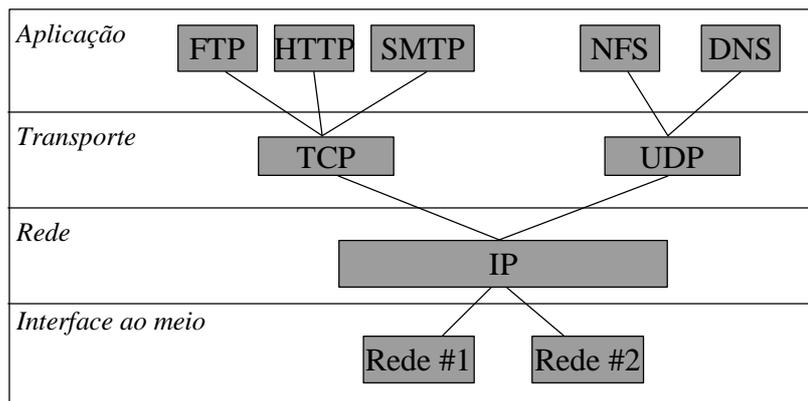
### – Características:

- directos/indirectos
- monolíticos/estruturados
- normalizados (*standard*)/proprietários
- simétricos/assimétricos

Normalmente surgem agrupados em *famílias de protocolos*



# Família de protocolos TCP/IP



- No topo de tudo estão as aplicações!
- O protocolo IP (*Internet Protocol*) é o denominador comum...



## TCP/IP

- O TCP/IP é:
  - “portável”:
    - Funciona em praticamente todos os sistemas operativos
    - Permite endereçamento global (à escala mundial)
  - suportado pela Novell, Microsoft, etc (líderes de mercado)
  - Extensível
  - Totalmente aberto: qualquer vendedor pode escrever a sua própria implementação
- Parte da sua popularidade deve-se ao Unix:
  - Desde o início que foi integrado no Berkeley Unix
  - Usado nas universidades, centros de investigação e agências governamentais (US)



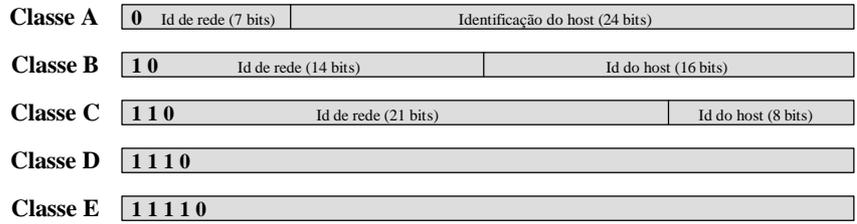
## IP – Internet Protocol

- Principais funções:
  - Unidade básica para transferência de dados: **pacote IP**
  - Endereçamento: **endereços IP**
  - Encaminhamento: nos **routers** com base no **IP destino** contido em cada pacote; O **IP origem** do pacote não é usado no encaminhamento
- Endereçamento IP
  - Endereço IPv4 - 32 bits (IPv6 –128 bits)  
XXXXXXXX.XXXXXXXXXX.XXXXXXXXXX.XXXXXXXXXX
  - parte identifica a rede ou subrede, e a outra parte, a interface do *host* com essa rede
  - numa internet, cada endereço tem de ser único
  - é usada notação decimal (Ex: 193.136.9.240)
  - atribuídos pela *Internet Assigned Number Authority*

*NOTA: host é qualquer equipamento capaz de transmitir e receber pacotes IP*



# IP – Internet Protocol



Classe	Gama de endereços	Número de hosts e de redes
A	0.0.0.0 - 127.255.255.255	126 redes de 16.277.214 hosts
B	128.0.0.0 - 191.255.255.255	16.384 redes de 65.354 hosts
C	192.0.0.0 - 223.255.255.255	2.097.152 redes de 254 hosts
D	224.0.0.0 - 239.255.255.255	Grupos de <i>multicast</i>
E	240.0.0.0 - 247.255.255.255	Reservada para uso futuro



# Acesso à Internet

Universidade do Minho

## Acesso à Internet

- É necessário recorrer aos serviços de empresas:
 

*Fornecedor de Serviço Internet*
- O acesso implica a escolha:
  - do fornecedor de serviço
  - do tipo de ligação até ao fornecedor de serviço

Utilizador
Fornecedor de Serviço

GCOM-DI-UM
45
CSI 2002, A.Costa

Universidade do Minho

## Acesso à Internet

Que tipo de ligação?

- A escolha do *tipo de ligação* depende das características da organização ou empresa:
  - ligações permanentes
    - destinatários: grandes empresas ou universidades
    - maior custo, maior capacidade de transmissão
    - acessibilidade 24h/24h
    - endereços IP alocados permanentemente
  - ligações não-permanentes
    - destinatários: PME e utilizadores individuais
    - menor custo, menor capacidade de transmissão
    - acessível apenas quando “ligado”
    - endereços IP atribuídos dinamicamente

GCOM-DI-UM
46
CSI 2002, A.Costa



## Acesso à Internet

### Suporte de ligação?

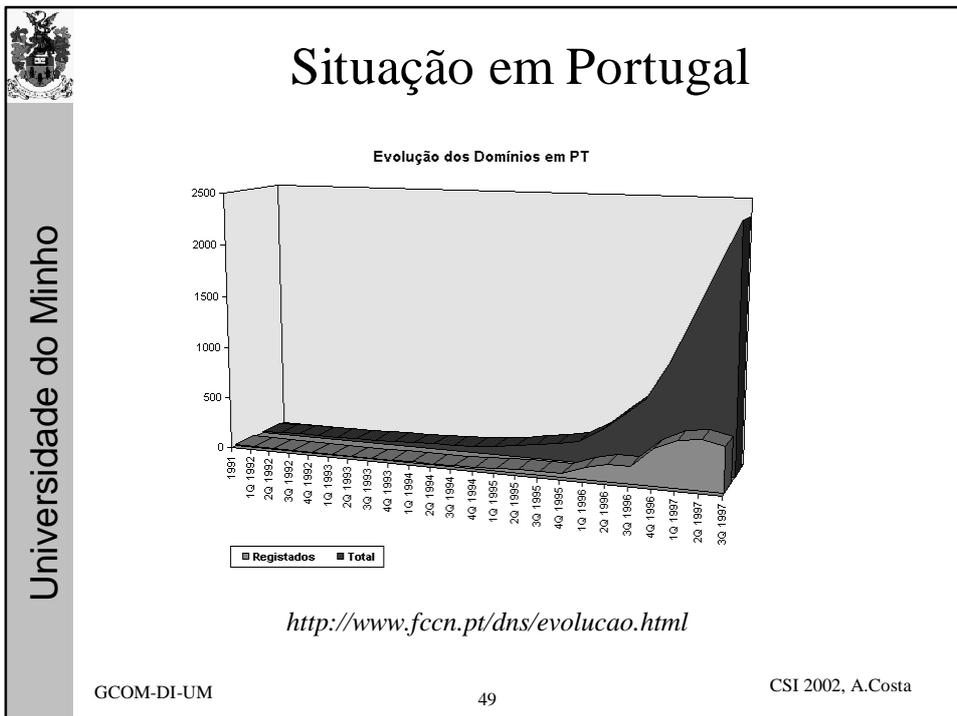
- Rede telefónica (PC + *modem* + linha telefónica)
  - trata-se da solução mais comum...
  - baixas velocidades, baixos custos
- Rede Digital de Integração de Serviços (RDIS)
  - serviço telefónico + transferência de dados
  - linha digital, 2 x 64 *Kbps*!
- Linhas dedicadas ponto-a-ponto
  - locais, inter-urbanas ou internacionais
  - custo em função da distância e da capacidade em *bps*
- Rede da televisão por cabo (TV Cabo, Bragatel)



## Acesso à Internet

### Que fornecedor de Serviço?

- *Internet Service Providers* (ISP)
- A escolha do *tipo de fornecedor de serviço de acesso* deve ter em conta:
  - as facilidades disponibilizadas:
    - Número e distribuição geográfica pelo país dos pontos de presença (PoP – Points of Presence)
    - Qualidade das ligações nacionais e internacionais a outros fornecedores de serviço (ISPs nacionais ou internacionais)
  - os custos envolvidos:
    - custos fixos (mensalidades)
    - custos variáveis (dependentes do tempo de ligação ou do volume de tráfego)



**Resolução de Nomes (DNS)**

Universidade do Minho

GCOM-DI-UM 50 CSI 2002, A.Costa



# Resolução de Nomes (DNS)

- Nomeação

**Problema:** endereços IP não são fáceis de memorizar!

- É mais fácil memorizar características como a *côr*, a *marca*, a *localização*, a *tarefa* que desempenha, etc..

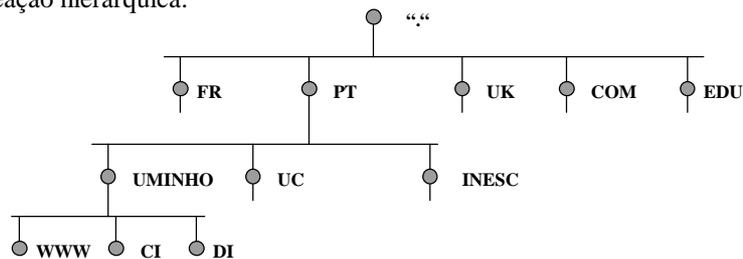
Em vez de endereços, podemos baptizar as máquinas com nomes, como *cinzento*, *LaserHP*, *MacCPII*, *WWW*, etc...

- precisamos de uma espécie de lista telefónica com nomes e endereços: DNS (*Domain Name System*)
- Nomes locais* não servem, precisamos de *nomes globais*!



# Resolução de Nomes (DNS)

- Nomes globais conseguem-se mais facilmente com uma estrutura de nomeação hierárquica:



Domínios de topo:  
.net (redes)  
.com (organizações comerciais)  
.gov (organizações do governo)  
.edu (instituições académicas)  
.mil (agências militares)

Domínios de topo(países):  
.pt (Portugal)  
.ca (Canadá)  
.de (Alemanha)  
.jp (Japão)  
.us (Estados Unidos)



## Resolução de Nomes (DNS)

- Nomes são obtidos por concatenação:

Exemplos: *www.uminho.pt, ci.uminho.pt*

- É impossível saber há partida se se trata de um *nome de sistema*, ou simplesmente de uma *organização* (domínio)
- É necessário uma *Base de Dados* com todos os nomes e endereços do mundo:

Necessariamente ***distribuída!***  
(cada qual gere a sua parte)

- O serviço DNS permite consultar a informação, mas nem sequer damos por ele...



## Resolução de Nomes (DNS)



- Todas as aplicações consultam o DNS!!
  - Enviar uma mensagem de e-mail pode implicar 2 ou três consultas!
  - Aceder a uma página WWW, implica pelo menos 1 consulta!



Tem de ser  **muito eficiente!**

- Funciona sobre UDP:
  - basta um único datagrama (512 bytes) por cada pedido e por cada resposta!
- Existem múltiplos servidores por cada domínio:
  - Um servidor **primário** e um ou mais **secundários**
  - Os servidores **secundários** mantêm, de forma automática, réplicas dos primários
- Os servidores e os clientes armazenam as respostas obtidas durante um certo tempo (TTL) para não andarem sempre a perguntar a mesma coisa...
  - CACHING!!



## Resolução de Nomes (DNS)

- Como se “distribui” a base dados?
  - De acordo com a estrutura hierárquica...
  - Um domínio **delega autoridade** aos sub-domínios...
  - Cada domínio é mantido e gerido por uma entidade
- Com tantos servidores, como se fazem as consultas?
  - Cada sistema precisa conhecer apenas um único servidor de DNS (normalmente o do seu domínio!)
    - É a esse servidor que as aplicações cliente formulam as suas *queries*
  - Qualquer servidor conhece sempre outros servidores!
  - Um servidor responde sempre a uma *query*:
    - Ou dá uma resposta.... Ou indica outro servidor que o possa fazer...
  - Um cliente obtém sempre uma resposta válida:
    - Nem que tenha de interrogar – automaticamente – vários servidores!



## E-Mail (SMTP e MIME)



## E-Mail

### Endereços

- Cada utilizador precisa de ter um *endereço* para poder trocar mensagens... e tem de ser único!
- Os endereços de E-Mail aproveitam os nomes de domínios e de sistemas mantidos no DNS...
- Formato de endereço:  
<nomeutilizador>@<departamento>.<organização>.<país>
- Duas variantes:
  - *user@dominio* ou *user@computador.dominio*
  - A primeira forma é preferível e a mais usada actualmente...
- Exemplos:
  - *costa@uminho.pt*, *macedo@uminho.pt*, *osg@di.uminho.pt*



## E-Mail

### Funcionalidades básicas

- Escrever, Ler, Imprimir e Apagar mensagens
- responder a uma mensagem (*reply*)
  - a mensagem original pode ser parcial ou totalmente incluída, distinguindo-se pela inclusão de um carácter > no início de cada linha
- reenvio (*forward*)
- ordenar o correio em *Pastas* (*folders*)
- imprimir, gravar, remover mensagens
- agenda de endereços
  - adicionar, pesquisar e remover endereços de e-mail



## Formato das Mensagens

- Protocolo **MAIL**, definido pelo RFC 822
  - Bastante simples e permite apenas texto
- Cada mensagem tem duas partes:
  - um cabeçalho e um corpo separados por uma linha em branco
- O corpo é texto plano com comprimento limitado
- O cabeçalho é um conjunto de campos estruturados com informação sobre a mensagem
- Cada linha é um campo diferente a menos que comece com um ` ` , que marca uma linha de continuação



## Formato das Mensagens

- **Formato:**

From: <endereço do originador>  
 To: <endereço do destinatário>  
 Cc: <cópia de cortesia>  
 Bcc: <cópia de cortesia "cega">  
 Subject: <assunto>  
 -----texto da mensagem-----



Cabeçalho

Corpo

- **Exemplo:**

To: laura.smith@umich.edu  
 Cc:  
 Subject: Citation Verification  
 ----

Laura, please get me a copy of the following article for which, unfortunately, I only have the author and year of publication: (Jonsey, 1993). As usual, I need the information yesterday.Thanks.



## Envio das mensagens: SMTP

- Protocolo **SMTP** (Simple Mail Transfer Protocol)
- Protocolo extremamente simples para transferir uma mensagem de uma máquina para outra, mas apenas texto:
  - limites no tamanho máximo de cada linha: 1000 caracteres
  - só permite caracteres ASCII de 7 bits ASCII



Impossível transmitir binários, ou outros *media* estruturados;

- **Soluções:**
  - Codificação dos conteúdos binários usando codificadores *uuencode/uuencode*
  - Extensões ao protocolo SMTP que ultrapassem as limitações
    - » nem todos os servidores as suportam!
    - Exemplos: transporte a 8bit; negociar tamanho linha; etc.
  - **MIME !!**



## MIME

- ***Multipurpose Internet Mail Extensions:***
  - protocolos que permitem a inclusão de objectos dentro de mensagens, mantendo total compatibilidade com os formatos RFC822 mais antigos
- **Modo de funcionamento:**
  - definição de 5 novos campos de cabeçalho:
    - MIME-Version, Content-Type, Content-Transfer-Encoding, Content-Description, Content-ID
  - cada mensagem MIME tem sempre definido o seu tipo de conteúdo:
    - Exemplo:
      - » Content-Type: multipart/mixed
  - definição de 5 formas de codificação do conteúdo para envio a 7 bits:
    - 7-bit, 8-bit, binary, quoted-printable, base64
    - Exemplo:
      - » Content-Transfer-Encoding: base64

Universidade do Minho



## Tipos de Conteúdo

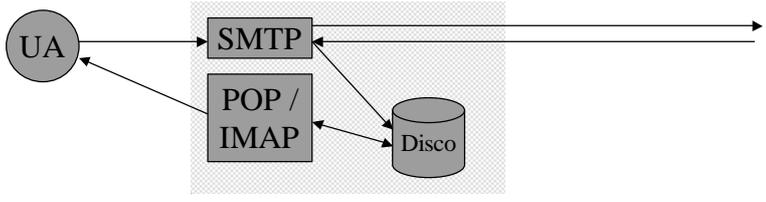
- Exemplos dos principais tipos MIME
  - Texto:
    - *text/plain* (texto não estruturado), *text/html* (página html), etc.
  - Imagem:
    - *image/gif*, *image/jpeg*, etc.
  - Audio:
    - *audio/mp3*, *audio/midi*, etc.
  - Video:
    - *video/mpeg*
  - Aplicação: formatos manipulados por aplicações
    - *application/postscript*, *application/pdf*, etc.
  - Multiparte: incluir várias partes de diferentes tipos
    - *multipart/mixed* (para visualização sequencial)
    - *multipart/parallel* (para visualização em paralelo)
    - *multipart/alternative* (conteúdo repetido em formatos alternativos)

GCOM-DI-UM
63
CSI 2002, A.Costa

Universidade do Minho



## Acesso a caixas de correio remotas



### POP - Post Office Protocol

- funções básicas de manipulação de 1 mailbox
- processamento local do mail
- 3 fases:
  - autenticação (login e password)
  - acesso à mailbox (LIST, RETR)
  - quit (actualiza a mailbox)

### IMAP - Internet Message Access Protocol

- permite manipular múltiplas maiboxes
- mantém mail no servidor - disponibilizando funções gestão remota
- suporta três modelos de *mail*:
  - *offline*
  - *online*
  - *disconnected*

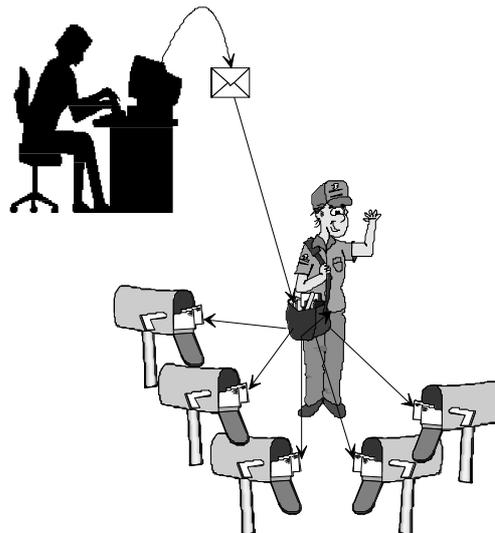
GCOM-DI-UM
64
CSI 2002, A.Costa



## Outros: Listas e News



## Listas distribuição





## Listas distribuição

- Para que servem?
  - Grupos de discussão
  - Projectos
  - Mailing para clientes
  - Jornais electrónicos
- Tipos de listas de distribuição
  - Inscrição
    - fechadas
    - abertas
  - Envio de mensagens
    - moderadas
    - não moderadas



## Listas distribuição

- Inscrição numa lista
  - Enviar uma mensagem para
    - *nome-da-lista-request@endereço*  
com o seguinte conteúdo
      - subscribe <Primeiro Nome> <Último Nome>
    - listmanager@endereço  
com o seguinte conteúdo
      - subscribe *nome-da-lista* <Primeiro Nome> <Último Nome>
- Outros comandos para gestores de listas
  - help [topic]; set <list> [<option> <arg[s]>]; unsubscribe <list>; signoff <list>; recipients <list>
  - OU
  - information <list>; statistics <list>; run <list> [<password> <cmd [args]>]; lists; index [archive | path-to-archive] [/password] [-all]



## USENET NEWS

- Grupos de interesse onde se discute tudo e mais alguma coisa (mais de 5000)
  - milhares de participantes em cada grupo
  - Mensagens similares às de correio electrónico trocadas aos milhares por dia
- Muitas listas de distribuição também são grupos de *news*
- As *News* são mais eficientes que as listas de distribuição quando há milhares de utilizadores, milhões de mensagens e a informação não é confidencial...
  - São também por vezes menos úteis por excesso de participantes...
- Com as *News* não há necessidade de inscrição centralizada
  - Todos os grupos são abertos
  - Tal como nas listas há grupos moderados e não moderados



## NEWS

- Organização dos Grupos de News
  - Estão organizados hierarquicamente com grupos, subgrupos e sub-subgrupos
  - O nome de cada grupo está separado do seu pai e dos filhos por um (.), por exemplo: *soc.culture.portuguese*
  - Hierarquia das News: primeiro nível
    - *comp* Tópicos relacionados com computadores e informática
    - *news* Rede e software de news
    - *rec* Hobbies, actividades recreativas e artes
    - *sci* Investigação científica e aplicações
    - *soc* Aspectos sociais
    - *talk* Debate em assuntos controversos
    - *misc* Algo que não caiba nos anteriores
    - *pt* Tópicos em português



# Segurança



# Segurança

- Computadores podem ser alvo de ataques:
  - ler informação confidencial
  - produzir alterações na informação armazenada
  - destruir o sistema!
- Quando ligados a uma rede:
  - Também a comunicação de dados pode ser objecto de ataques!
  - Já não é necessário acesso físico ao equipamento
  - O número de potenciais atacantes torna-se maior...



## Segurança

- A introdução de mecanismos de segurança
  - visa minimizar as vulnerabilidades do sistema, mas...
  - ... torna os sistemas mais caros e mais difíceis de usar
- É conveniente:
  - Identificar as vulnerabilidades do sistema
  - Identificar os ataques que podem explorar essas vulnerabilidades
  - Estimar o custo de cada ataque, se concretizado
  - Estimar o custo das contra-medidas a adoptar
  - Fazer uma análise custo/benefício para decidir que mecanismos integrar no sistema



## Ameaças mais comuns

- Ameaças activas
  - se concretizadas produzem alteração da informação (armazenada ou transmitida)
- Ameaças passivas
  - não produzem alteração de informação, estado ou operação dos computadores
  - obtenção de informação a usar na concretização de ameaças activas



## Ameaças mais comuns

- **disfarce**
  - um utilizador faz-se passar por outro...
- **intercepção de dados**
  - os dados armazenados ou trocados numa comunicação são observados por utilizadores não autorizados...
- **intercepção de identidade**
  - a identidade de um ou mais utilizadores em comunicação é observada para uso indevido... tipicamente para o atacante se “disfarçar”...
- **repúdio**
  - Utilizador nega ter realizado determinada acção que de facto realizou... por exemplo ter participado numa comunicação



## Ameaças mais comuns

- **manipulação**
  - os dados que fluem na comunicação são manipulados (substituição, inserção, remoção ou alteração da ordem); mesmo informação armazenada, como os programas, podem ser manipulados...
- **negação do serviço**
  - atrasar ou mesmo impossibilitar a operação da rede ou dos sistemas a ele conectados (por exemplo gerar automaticamente falsas conexões com um servidor!)
- **exploração de erros do sistema operativo e do *software*,**
- **etc.**



## Serviços de Segurança

- **Autenticação**
  - oferece protecção em relação ao *disfarce e interceptação de identidade*
  - exemplos:
    - verificação de identidade na ligação de um utilizador a um computador remoto por *telnet*;
    - verificação de identidade do originador (*From:*) de uma mensagem de correio electrónico
- **Controlo de acessos**
  - protege o uso não autorizado de recursos disponíveis (leitura, escrita, execução, etc.)
- **Confidencialidade**
  - serviço que protege contra a ameaça de *intercepção de dados*, tornando os dados inteligíveis por entidades não autorizadas...



## Serviços de segurança

- **Integridade dos dados**
  - serviço que oferece protecção contra a ameaça de *manipulação*, detectando quaisquer alterações...
- **Não repúdio**
  - prova de entrega
    - impossibilitar que o destinatário negue ter recebido
  - não repúdio da origem
    - impossibilitar que quem originou determinada informação venha a negar tal facto...

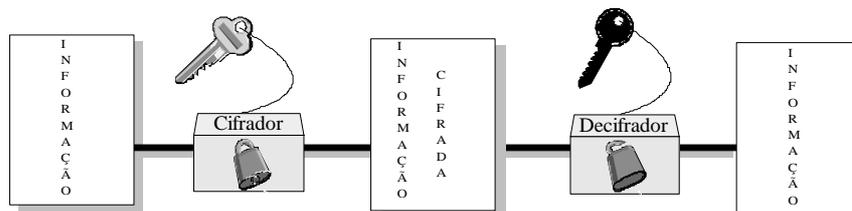


## Criptografia

- O exemplo de Júlio César:
  - mensagem original: “cartago esta no papo”
  - mensagem codificada: “ebtciguxcpqrerq”
  - uso de um algoritmo criptográfico baseado na substituição de letras ( 2 posições no alfabeto )
- Os algoritmos actuais:
  - algoritmos muito mais complexos, que dissimulam qualquer padrão existente no texto
  - número de “chaves” muito grande, que invalidem a procura por tentativas...
  - não há necessidade de manter os algoritmos secretos, mas apenas as chaves...



## Componentes dos Sistemas criptográficos





## Criptografia simétrica e assimétrica

- **Sistemas simétricos:**
  - chave usada na cifragem é igual á chave usada na decifragem...
  - obriga os interlocutores a negociarem uma chave antes de iniciarem uma comunicação: **Problema!**
- **Sistemas assimétricos (ou de chave pública)**
  - cada utilizador possui um par de chaves:
    - o texto cifrado com uma só pode ser decifrado com a outra
    - conhecendo uma delas é impossível descobrir a outra
    - utilizador mantém uma secreta (*chave privada*) e divulga a outra (*chave pública*)
    - não há necessidade de negociar nada á partida
    - Problema: são demasiado lentos
- São muitas vezes usados em conjunto



## Funções de sumariação

- A partir de texto de qualquer tamanho, é produzido um sumário de tamanho fixo.
- A partir de um sumário é impossível determinar o texto que lhe deu origem
- Deve ser impossível encontrar dois textos que produzam o mesmo sumário...
- Os sumários funcionam como se fossem a impressão digital do texto sumariado

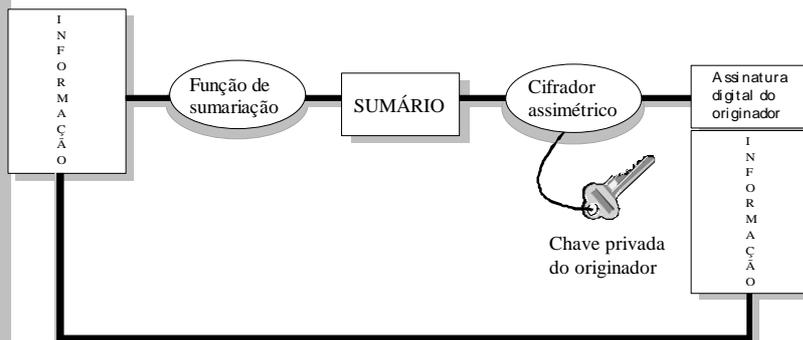


## Mecanismos baseados em criptografia

- Assinaturas digitais
  - tal como uma assinatura vulgar, adiciona-se ao texto para:
    - possibilitar aos destinatários a verificação da origem
    - associar o texto ao seu originador de forma a que este não possa negar esse facto
    - garantir a sua *integridade* ( a mensagem recebida é realmente a mensagem que foi originada )



## Geração de uma assinatura digital



- produzir um sumário
- cifrar o sumário com a sua *chave privada*
- enviar a mensagem com a assinatura anexada

**Verificação duma assinatura digital**

Universidade do Minho

- produzir um sumário a partir do texto original
- decifrar o sumário com a *chave pública* do originador
- comparar os dois sumários obtidos

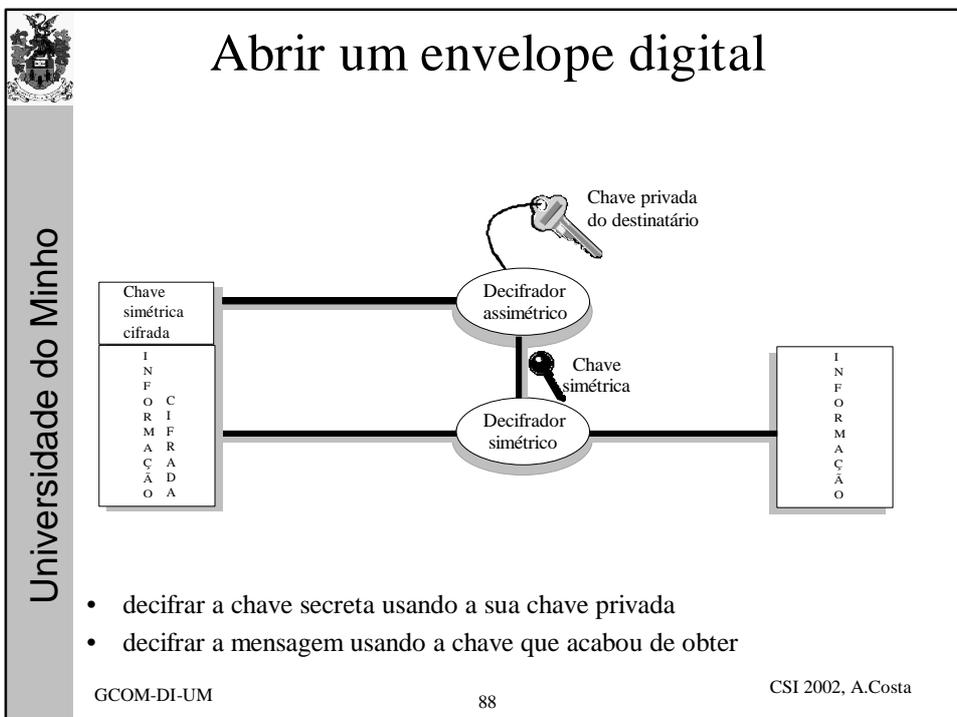
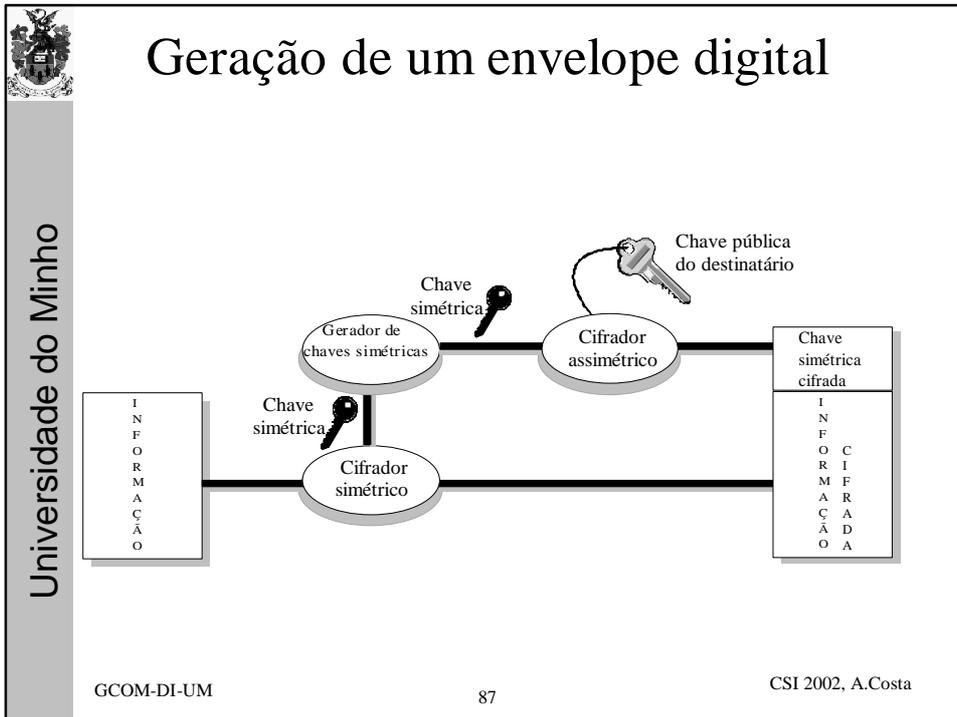
GCOM-DI-UM 85 CSI 2002, A.Costa

**Mecanismos baseados em criptografia**

- Envelope digital
  - tal como os envelopes normais impede os observadores não autorizados de aceder ao conteúdo
  - realiza o serviço de *confidencialidade*
  - é usada criptografia simétrica e assimétrica em conjunto:
    - gera-se aleatoriamente uma chave secreta
    - cifra-se toda a mensagem com essa chave (algoritmo simétrico)
    - cifra-se a chave usada no ponto anterior com a chave pública do destinatário (algoritmo assimétrico)
    - envia-se todo o conjunto para o destinatário...

Universidade do Minho

GCOM-DI-UM 86 CSI 2002, A.Costa





## Mecanismos baseados em criptografia

- Certificação digital
  - Todos os mecanismos baseados em técnicas criptográficas assimétricas dependem da autenticidade das chaves públicas!
  - É preciso garantir que as chaves pertencem mesmo aos seus donos...
  - Como?
    - gerando um certificado com o nome do utilizador, a sua chave pública, a validade da chave, etc e assinando essa chave por uma *autoridade de certificação*
    - é preciso certificar as chaves dessas autoridades de certificação, formando uma cadeia de certificação



## Exemplo de utilização no E-Mail

- Existem duas soluções para correio seguro na Internet: o PEM e o PGP
- O PGP (*Pretty Good Privacy*) é o mais popular e existe para todas as plataformas
- Passos a seguir:
  - obter o software PGP
  - criar um par de chaves para si próprio
  - divulgar a sua chave pública (de todas as formas)
  - assinar mensagens e mandar mensagens secretas!

**Mensagem assinada**

Universidade do Minho



-----BEGIN PGP SIGNED MESSAGE-----

Ola,

A data do exame e' ????

Esta mensagem vai "assinada" por mim, por forma a garantir a integridade da informação e a autenticidade da origem.

Antono Costa

-----BEGIN PGP SIGNATURE-----

Version: 2.6.3ia  
Charset: cp850

iQBVAwUBM3gQUWYb/BZZ2Av1A0FpIq1Aq9X13sLzys22ZCUPB1aF/nQ11o8NFRhF  
vK0coS23UDiVZ0Ye0rNSrqaW37w@buCMT7PIQdoj8d3oRqXV18Qa0==  
=TzQc

-----END PGP SIGNATURE-----

2002, A.Costa

GCOM- 91

**Exemplo de uma chave pública**

Universidade do Minho

```

-----BEGIN PGP PUBLIC KEY BLOCK-----
Version: 2.6.1

mQCNai5Db4AAAEAPBJQcFcTDxvVZ6Z7q9QSWcoF0yYdynGkYaAFot0Lja/bZQ
ax/JBTmWabfTSDUUXb8e3YslQ56wikYDFomvZOJa3nJ532/QPQ7Rv484EvkMhkz
+s4tXx5fJ3obUom8baTW8axiXkX6WGAAzd0Ovw8rJHYAOnWp0iGKhervvsERAAUR
tCZKb3NIE1hbnVlbCBWYXlibmNhlDXqbXZAZGkudW1pbmhvLnB0PokAIQIFEC59
alQhioXK777BEQEBxz8EALScpOZQThwVlczCEprGtMF45JF3SF/al3jgrPEP0zQo
6xpOM7Zt4b5oDL9feV1yyTyxrY3+EBugVfgDvurVgPHGXTZ/7Fv9A3yKKyCPXBHg
/mGKL5QYKF5gbQk+27tr6owCegDYPrPGW0lzW+zLvwHbDnvn2y8ikIXirYZWMJ3
iQCVAgUQLna21y7DGmRvWfpAQGUWAQAvFYXkvuZ8YLlp12sd2OSjB1EajEzxDV7
7FVtKzcz6NVBdZfwloBrTogpdsPDLix1b+LhkBkXi8jF1UMFeTFOiXa23ZwIXSV
PpXqzbZU+PEa3VLXd84/C/koCJEpf+BIXcLGlwAiyt+N3OsjYNjilmAqBfdJi2hD
nB+0tSVC9Ac=
=NxO5
-----END PGP PUBLIC KEY BLOCK-----

```

GCOM-DI-UM 92 CSI 2002, A.Costa



# World Wide Web



## Hipertexto e Hipermedia

- Um conceito para:
  - *a apresentação* (forma como o documento é exibido)
  - *o acesso* (forma de aceder e obter os docs.)
  - *a estrutura* (forma com está estruturado)
  - *o armazenamento* (diferentes formas de armazenamento)

**Hipertexto** é texto com ligações a outros textos. Os documentos hipertexto não são estritamente sequenciais, pois podem conter *referências* a outras partes do documento ou mesmo a outros documentos.

**Hipermedia** é média com ligação a outros média. Trata-se da aplicação do conceito de hipertexto, a documentos multimédia.

Universidade do Minho

## Hipertexto e Hipermedia

- Modelo teórico: *nós* com *ligações* entre si

Âncora (origem)  
 Âncora (destino)  
 Ligação

**Âncoras** são fragmentos de informação (palavras, frases, etc.) dentro de um documento, aos quais se podem associar *ligações*.

**Ligações** são referências ou apontadores, de uma âncora para outra. Devem identificar o documento destino (*Qual o nome?*), sua localização (*Onde está?*) e forma de acesso (*Como obter?*)

GCOM-DI-UM 95 CSI 2002, A.Costa

Universidade do Minho

## Hipertexto e Hipermedia

- Trata-se de uma ideia muito antiga:
  - Notas de rodapé e referências internas* (ver cap. Y)
  - Índices e tabelas de conteúdos*
  - bibliografias*

O que há de novo é apenas a facilidade com que se seguem os *links*, quer estes sejam *internos* quer *externos*: Basta um *click*!

Os documentos hipertexto contêm palavras seleccionadas também designadas por âncoras

As âncoras podem ser expandidas a qualquer altura por quem lê, para obter informações adicionais.

**Hipertexto é texto com ligações!**

A forma como a expansão é provocada, depende do interface:

- clicar na âncora (interfaces gráficas!)
- digitar um número de referência (interfaces não-gráficas, só de texto)

GCOM-DI-UM 96 CSI 2002, A.Costa



# Hipertexto e Hipermedia

- Podemos ter *cadeias de ligações* com dois objectivos diferentes e complementares:

## Focagem

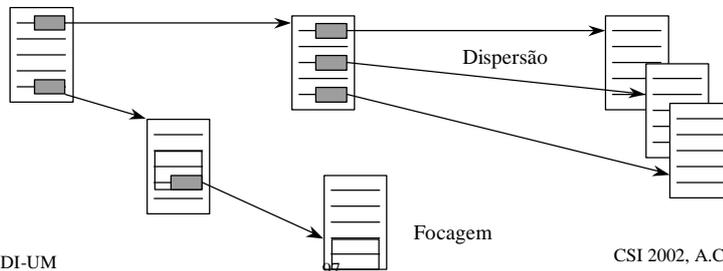


O utilizador progride de documento em documento, restringindo a cada salto o domínio de informação e convergindo para um tópico específico.

## Dispersão



O utilizador não tem uma ideia precisa daquilo em que está interessado, e o sistema abre-lhe novas possibilidades com colecções de referências.



GCOM-DI-UM

97

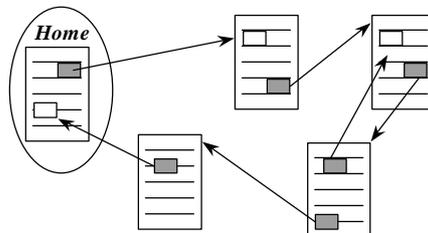
CSI 2002, A.Costa



# Hipertexto e Hipermedia

- As cadeias de ligações não tem de ser necessariamente hierárquicas:
  - A *estrutura topológica* é normalmente uma *teia*, podendo haver *ciclos* (caminhos que conduzem ao ponto de partida)

Como evitar que os utilizadores se percam nestes labirintos, enquanto *navegam*?



1. Documento inicial (*Home*)
2. Caminho percorrido é caminho de retorno

GCOM-DI-UM

98

CSI 2002, A.Costa



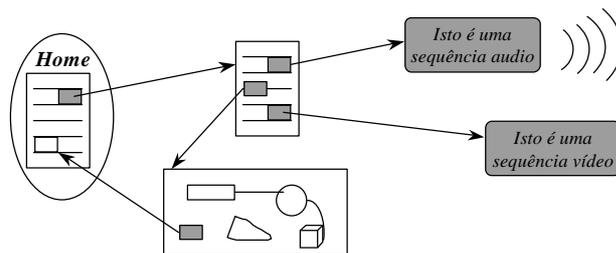
## Hipertexto e Hipermedia

- Outras soluções a que os *autores* podem recorrer:
  - *Mapas* com a estrutura, ou *diagramas gerais* do *site*
  - *Visitas Guiadas* previamente preparadas
  - *Barras de navegação* com sugestões de direcção em cada instante
  - Utilização de *Frames*, fixando numa parte da janela informação de localização (menus, pág. inicial, etc.)
  - Possibilidade de *pesquisa* pelo conteúdo em certos pontos da navegação
- Conceitos válidos, independentemente da localização:
  - CD-ROM local, no disco rígido, em *servidores* distintos, dispersos geograficamente!



## Hipertexto e Hipermedia

- Particularidades dos sistemas hipermedia em *rede*:



- relacionamento *temporal* entre os objectos!...
- é necessário minimizar *tempos* de transferência:
  - evitar imagens grandes, sons e vídeo em contínuo...
  - objectos de grandes dimensões só a pedido do utilizador!



# World Wide Web

Universidade do Minho

**O que é?**

- Nasce em 1989, CERN, Suíça...
- A designação é a do projecto que lhe deu origem...
- Concretiza o conceito de hipermédia na Internet

Definição

→

Sistema de informação *hipermédia*,  
*cooperativo, distribuído e heterógeneo*

**Outros significados:**

- designa o conjunto dos *protocolos* desenvolvidos pelo projecto
- designa o *hiperespaço* de informação disponível na Internet e suportado por *servidores de informação* interligados...

GCOM-DI-UM
101
CSI 2002, A.Costa



# World Wide Web

Universidade do Minho

- Ideias base (revolucionárias?) do WWW:
  1. Todos *podem criar* documentos e inseri-los na *teia!*
    - qualquer um pode ser autor, incluir referências para qualquer outro documento, e ser referenciado livremente...
    - não há nenhuma autoridade administrativa, centralizante, ou restritiva de qualquer outra forma...
  2. Uma forma *uniforme* de *localizar* documentos em *todo o mundo!*
    - foram introduzidos métodos para definir com exactidão *onde* e *como* encontrar documentos na Internet...
    - estas *moradas*, designam-se por Uniform Resource Locators (URLs)
    - não são apenas *endereços*, porque incluem *métodos de acesso*

GCOM-DI-UM
102
CSI 2002, A.Costa

**World Wide Web**

The diagram illustrates the architecture of the World Wide Web. On the left, a local network (LAN) is shown, containing a 'Serv. WWW local' (local WWW server) and a 'Ficheiro local' (local file). A computer is also connected to this LAN. The LAN is connected to the 'Internet'. On the right, a 'Servidor WWW remoto' (remote WWW server) is connected to the Internet. The Internet is represented by a central oval.

**3. Um interface com o utilizador, único e uniforme!**

- similaridade de operação entre clientes distintos...
- transparência da localização (não interessa onde está!)
- esconder incompatibilidades entre sistemas na representação dos dados...

GCOM-DI-UM 103 CSI 2002, A.Costa

**World Wide Web**

**4. Acesso a qualquer "base de dados" de informação**

- é preciso garantir acesso a dados que não tenho sido produzidos especificamente para o WWW...
  - mesmo que forma mais limitada, e sem *links*
- são necessárias aplicações que sirvam de intermediários no acesso:
  - designam-se por *gateways de aplicação*

**5. Um suporte para realizar transacções!**

- o utilizador pode introduzir dados em formulários e enviá-los...
- ...o que permite a realização de vários tipos de transacções.
- Uma das utilizações, entre muitas, é o comércio electrónico!

GCOM-DI-UM 104 CSI 2002, A.Costa



Universidade do Minho

## WWW - Os protocolos

- *Uniform Resource Locators (URLs)*

**Formato geral:**

`<protocolo-acesso>://<computador.organização.país>/pasta/subpasta/ficheiro.ext`

①
②
③
④

- ① – protocolo de acesso aos dados (HTTP, FTP, etc..)
- ② – nome completo da máquina (ou endereço IP!)
- ③ – pastas dentro das quais se encontra o documento
- ④ – nome do documento e respectiva extensão!

**Exemplo:**

✦ <http://med-amsa.bu.edu/Gutenberg/Welcome.html>

GCOM-DI-UM
105
CSI 2002, A.Costa



Universidade do Minho

## WWW - Os protocolos

- *Uniform Resource Locators (URLs)*

**Outros exemplos:**

- <ftp://ftp.ci.uminho.pt/pub/README> (arquivo ftp)
- <http://marte.uminho.pt/MCS/index.html> (servidor www)
- <telnet://orpheu.ci.uminho.pt> (telnet)
- <gopher://gopher.uminho.pt/> (gopher)
- <news:pt.internet> (grupo de *news*)
- <mailto:csxxxxx@ci.uminho.pt> (envio de e-mail)

GCOM-DI-UM
106
CSI 2002, A.Costa

Universidade do Minho

## WWW - Os protocolos

- *HyperText Transfer Protocol (HTTP)*
  - otimizado para hipertexto interactivo
  - Rápido: pedidos satisfeitos numa única interacção (ida e volta)

- Não orientado ao estado (*stateless*)
  - Não se estabelece nenhuma sessão entre servidor e cliente
  - O URL é suficiente para aceder ao documento
- Extensível: suporta transferência de qualquer tipo de dados
- Não é indispensável: podem ser usados outros protocolos...

GCOM-DI-UM
107
CSI 2002, A.Costa

Universidade do Minho

## WWW - Os protocolos

- *HyperText Transfer Protocol (HTTP)*

```

HTTP/1.0 200 OK
Server: Netscape-FastTrack/2.01
Date: Mon, 11 Feb 2002 19:49:32 GMT
Accept-ranges: bytes
Last-modified: Mon, 08 Feb 1999 15:04:21 GMT
Content-length: 969
Content-type: text/html

<HTML>
..... Página HTML pedida ...
</HTML>
  
```

GCOM-DI-UM
108
CSI 2002, A.Costa



## WWW - Os protocolos

- *HyperText Markup Language* (HTML)
  - um tipo simples derivado da linguagem *SGML*
  - suporta algumas estruturas lógicas simples...
    - cabeçalhos, parágrafos, listas numeradas e não numeradas, tabelas, *frames*, inserção de imagens, etc.
    - formulários, com elementos de selecção do tipo escolha múltipla, botões on-off, etc. para permitir transacções...
  - e também naturalmente a inserção de hiperligações :
    - âncoras e *links*
  - é o único tipo que os *browsers* WWW têm que conhecer
  - os browsers WWW *convertem* os códigos HTML em instruções de visualização apropriadas...



## WWW - Os protocolos

- *HyperText Markup Language* (HTML)

```
<HTML>
<HEADER>
<TITLE> Título do Documento </TITLE>
</HEADER>
<BODY>
<H1> Cabeçalho principal </H1>
O texto é em formato livre e deve ser
estruturado em parágrafos. <P>
Novo parágrafo com texto.
<PRE> ... Texto pré-formatado... </PRE>
Clique
<A HREF="ficheiro2.html"> aqui </A>
para ver outra página!
</BODY>
</HTML>
```

Texto original em HTML

```

Título do Documento

Cabeçalho principal

O texto é em formato livre e deve ser
estruturado em parágrafos.
Novo parágrafo com texto.

... Texto pré-formatado ...

Clique agui para ver outra página!
```

O que um cliente WWW exibiria



## Outros formatos

- Linguagens de definição de páginas comuns no WWW:
  - **PDF** *Portable Document Format* ← Mais popular
    - formato universal da Adobe
    - requiere um “Adobe Acrobat Reader” que é gratuito!...
  - **DigitalPaper**
    - requiere também um *PlugIn* gratuito...
  - *PostScript*
    - formato mais adequado para impressão...

Para obter um melhor *layout* e melhores formatos, muitos utilizadores preferem disponibilizar os documentos em formatos de página... adequados tanto para consultas *on-line* como para impressão... Muito usual em *papers* na comunidade académica.



## Servidores Web

- O mercado tem vindo a criar as seguintes categorias de servidores Web (HTTP):
  - *Basic Web Servers*
    - servem apenas documentos (páginas HTML, Imagens, e outros)...
    - suportam **extensões** para serviços adicionais...
  - *Web commerce ou merchant servers*
    - orientados para o comércio electrónico: compras, vendas, transações financeiras, etc.
    - segurança (recorrendo a técnicas criptográficas) é fundamental!
  - *Web exchange servers*
    - integram funcionalidades associadas a troca de mail, fax, news, comunicação em grupo em diferido ou em tempo real, etc...
  - *Web-oriented database servers*
    - são servidores de bases de dados, preparados para o Web

Universidade do Minho

## World Wide Web

- Problemas com o excesso de informação:
  - Existe uma enorme quantidade de informação...
  - ... mas:
    - *muitas páginas não são actualizadas...*
    - *muitas tem incorrecções...*
      - Produzidas à pressa! Afinal é tão fácil publicar...
    - *informação não documentada...*
    - *muito difícil “pesquisar”, apesar dos “engenhos de pesquisa”*
    - *os “engenhos” devolvem muitas páginas indesejadas!*
      - Por vezes devido a estratégias dos seus próprios autores!...

GCOM-DI-UM
113
CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- Problema dos *links* obsoletos:

**Solução:**

Encontrar formas de identificar os documentos,  
 que sejam independentes da localização e do nº de cópias!  
*Uniform Resource Name (URN)* (parecido com ISBN dos livros)

GCOM-DI-UM
114
CSI 2002, A.Costa



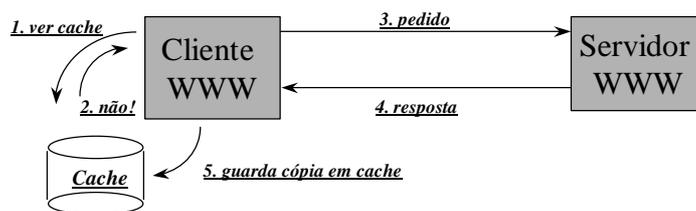
## World Wide Web

- Problema da sobrecarga da rede
  - Assume-se uma rede com cobertura mundial e capacidades infinitas de transferência que não existem!
  - Os utilizadores nem sempre são avisados do tamanho do documento antes da transferência
  - Os utilizadores e autores não estão conscientes das limitações da rede
  - O mesmo documento pode ser transferido mais do que uma vez pelo mesmo cliente:
    - *Proxy, Caching...e Mirroring*
  - Livre acesso a conhecimentos *versus* negócio!



## World Wide Web

- *Caching* no Cliente WWW



- o utilizador passa várias vezes pelos mesmos documentos...
- ...se os documentos mais recentes forem guardados em *cache*,
- podem posteriormente (dias) ser prontamente acedidos em disco
- contribui-se para diminuir a carga na rede e o tempo de acesso

Universidade do Minho

## World Wide Web

- Será possível ter uma *cache* partilhada por todos os clientes de uma organização?

*Caching* num servidor próprio - **Proxy** - usado por todos

GCOM-DI-UM
117
CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- Para que o *caching* seja possível ao nível da organização é necessário que os pedidos de todos os clientes passem por um mesmo servidor - o servidor *proxy*
- Ao receber um pedido, o *proxy* encaminha-o ao servidor original e remete a resposta obtida ao cliente, mantendo uma cópia na *cache*
- Quando é pedido um documento repetido devolve ao cliente a cópia em *cache*
- Ao configurar o seu *browser* só tem vantagens em indicar qual é o *proxy* da organização
- O *proxy* também pode permitir a *browsers* colocados em redes sem conectividade Internet aceder a documentos sem quaisquer restrições

GCOM-DI-UM
118
CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- *Proxy e Caching*

GCOM-DI-UM 119 CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- *Mirroring*

– Objectivos:

- Disponibilizar *réplicas* mais próximo dos utilizadores, optimizando a utilização dos recursos da rede
- Distribuição de “carga” (nº pedidos) entre servidores...
- Tolerância a faltas: se um não estiver disponível, há substitutos

GCOM-DI-UM 120 CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- *Robots* ou *Spiders*

Estes clientes especiais, surgem para tentar realizar tarefas que com o crescimento do WWW deixaram de poder ser feitas manualmente...

São **Agentes de Software** que exploram autonomamente alguma porção da Web, seguindo todas as ligações (*links*) existentes!

→

- Os percursos cíclicos são evitados...
- A porção a ser explorada pode ser demarcada por:
  - *domínios* ou *sites*
  - nº de *níveis*
  - combinações dos dois anteriores

GCOM-DI-UM 121 CSI 2002, A.Costa

Universidade do Minho

## World Wide Web

- *Robots* - Utilização

Todos os *Robots* realizam as seguintes tarefas, continuamente:

- Obter página...
- Procurar e recolher todos os URLs contidos nessa página
- Fazer algum tipo de processamento
- Próxima página ainda não visitada...

É no processamento que fazem que os *Robots* se distinguem entre si

- Ex: **Robot de replicação** (*mirror*) cuja tarefa é gravar todas as páginas de um *site* no disco local, corrigindo todas as referências para continuarem válidas localmente!...

GCOM-DI-UM 122 CSI 2002, A.Costa



## World Wide Web

- Além da **replicação**, as tarefas mais comuns são:

### Estatísticas WWW

- Recolher dados estatísticos: nº de *sites*, documentos por *site*, tamanho médio dos documentos, nº de imagens por pág., etc.

### Manutenção

- Testar a integridade das ligações internas e externas (apenas os primeiros links para fora) e alertar para grandes alterações realizadas em documentos. Auxiliam a gestão de um *site*...

### Construção de Índices

- Criar enormes bases de dados que indexam os documentos e ficheiros existentes no WWW. Alguns catalogam todo o texto, outros apenas títulos ou resumos... São os mais populares!



## World Wide Web

Os chamados **Agentes Inteligentes**, são os novos robots pessoais que podem ser “treinados” pelo seu dono (assuntos de interesse) e depois “buscam” informações semelhantes na rede. Os mais sofisticados usam mesmo técnicas de Inteligência Artificial...

### Problemas com os *Robots*:

- “*inundam*” os servidores com pedidos
  - disparam pedidos muito rapidamente e por vezes em paralelo!
- “*entopem*” a rede
  - consomem largura de banda, prejudicando todos (sites visitados e também a rede local!)

### • Soluções:

- ficheiro *robots.txt* contendo o nome dos indesejados!...
- detectar e impedir acessos em massa da mesma origem!...



## WWW - Pesquisa avançada

“É muito fácil pesquisar... o difícil é pesquisar bem!...”

- A importância crescente do “***Information Retrieval***”:
  - Grande número de documentos multimedia acessíveis na Internet através do WWW, FTP, etc...
  - Bibliotecas digitais
  - Necessidade de mecanismos eficazes e eficientes para descobrir informação relevante
  - Devido a limitações na tecnologia actual:
    - texto apenas, em vez de multimedia



## WWW - Pesquisa avançada

Ferramentas de pesquisa são programas que fornecem informação...

- Classificação das ferramentas de pesquisa:

Catálogos	Motores de Busca	Mistos	Multi-Engenho
Pesquisa hierárquica “por assunto”	Pesquisa “por palavras chave”...	Combinação dos anteriores	Ferramenta que interroga várias outras em paralelo

- Ou, do ponto de vista da indexação:

Indexação manual	Indexação total ( <i>full</i> )	Indexação parcial
Feita por utilizadores, que classificam em categorias!	Indexação automática de todas as palavras!	Indexação de apenas algumas palavras chave!



## WWW - Pesquisa avançada

- **Modelo Booleano**
  - Base de dados consiste em listas de ocorrências de palavras:
    - lista {K1} = {D1, D2, D3, D4};
    - lista {K2} = {D1, D2};
    - lista {K3} = {D1, D2, D3};
    - lista {K4} = {D1}
  - A interrogação do utilizador é um conjunto de palavras chave ligadas por operadores booleanos:
    - Interrogação: (K1 AND K2) OR (K3 AND (NOT K4))
  - O sistema devolve todos os documentos (Di) para os quais a interrogação é verdadeira:
    - Resposta: D1, D2, D3
- Utilizado por um grande número de sistemas operacionais, mas muito difícil de usar por utilizadores ocasionais...



## WWW - Pesquisa avançada

- **Modelo Vectorial**
  - cada documento é representado por um vector de termos cujos pesos são o número de ocorrências no documento...
    - *num espaço de dimensão T (número de termos distintos na colecção)*
  - também a interrogação é vista como um vector...
  - Os vectores que representam o documento e a interrogação são confrontados e é calculado um valor *Retrieval Status Value (RSV)*
  - Para cada interrogação, os documentos são mostrados ao utilizador por ordem decrescente desse valor, podendo ser inspeccionados...
  - a relevância dos documentos ou partes inspeccionadas pelo utilizador pode servir para realimentar o sistema com novas interrogações (*realimentação de relevância*)



## WWW - Pesquisa avançada

- Dicas e truques para uma boa pesquisa
  - Adopte uma estratégia baseada naquilo que sabe:
    - Use os catálogos, se sabe em que tópicos gerais se inclui aquilo que procura...
    - Use os motores de busca (c/ indexação total) se conhece coisas específicas como nomes ou palavras chave...
    - Características adicionais, como autor, localização, língua, etc, podem ser relevantes para auxiliar a pesquisa
  - Utilizar palavras pouco comuns e se possível diferenciadas...
  - Utilize vários motores de busca e procure sempre entender os resultados, distinguindo os “falsos acertos”



## WWW - Pesquisa avançada

- Utilização de alguns operadores para construir perguntas mais sofisticadas:
  - Operadores lógicos (AND, OR, NOT) e de proximidade (NEAR)
  - Sinais mais (+) e menos (-) para incluir ou excluir termos
  - Entre aspas (“...”) para definir uma frase exacta ou palavras muito relacionadas que devem ser procuradas em conjunto...
  - Parentesis rectos [..., ..., ] para agrupar palavras a serem pesquisadas quer isoladamente quer em conjunto
  - Operadores “mais como este” ou “more like this” (pesquisas por exemplos)
  - Letras maiúsculas, usadas para definir nomes próprios ou acrónimos...
- Nem todos os operadores são suportados por todos os serviços de pesquisa!



## WWW - Pesquisa avançada

- **Exemplo 1:** financiamento de investigações sobre o cancro
  - utilizando operadores lógicos para definir sinónimos:  
*(cancer OR oncol\*) AND research AND (grant\* OR fund\*)*
  - utilizando parentesis, ou aspas:  
*[cancer, oncol, research, grant, fund]*  
*"cancer research grant"*
  - operadores de proximidade  
*cancer NEAR research NEAR grant*
- **Exemplo 2:** informação sobre Warner-Lambert ou sobre o seu grupo de investigação farmacêutica Parke-Davis
  - Letras maiúsculas para nomes próprios e aspas para agrupar:  
*"Parke Davis" OR "Warner Lambert"*  
*"Warner-Lambert", "Parker-Davis"*