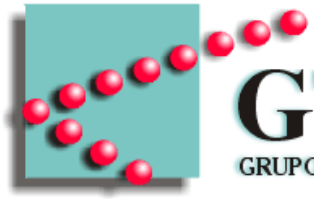


Técnicas de Defesa Contra *Spam*

Danilo Michalczuk Taveira[1], Igor Monteiro Moraes[1],
Marcelo Gonçalves Rubinstein[2] e Otto Carlos Muniz Bandeira Duarte[1]

[1] Universidade Federal do Rio de Janeiro – PEE/COPPE – DEL/Poli
[2] Universidade do Estado do Rio de Janeiro – PEL/DETEL/FEN

Apoiado pelos recursos da CAPES, CNPq, FAPERJ, FINEP, RNP e FUNTTEL

Roteiro

- Introdução
- Mensagens não solicitadas (*spam*)
  - Definição e classificação
  - Motivação e prejuízos
- Técnicas de envio
  - Coleta de dados e formato das mensagens
- Sistemas anti-*spam*
  - Baseados em filtragem simples
  - Com auto-aprendizado
  - Baseados na verificação da origem
- Perspectivas futuras

Introdução

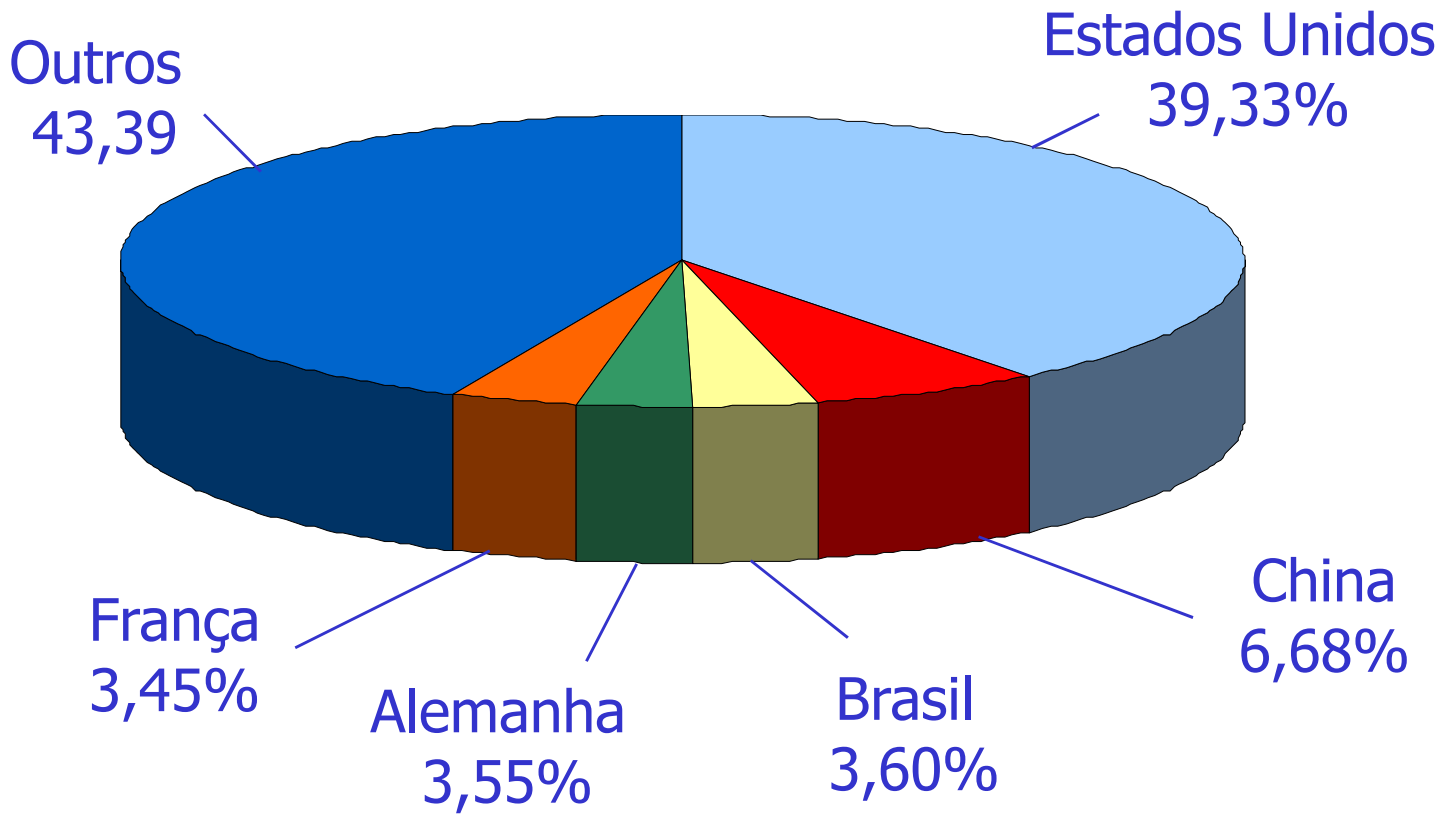
- Mensagem eletrônica não solicitada (*spam*)
  - Toda mensagem não desejada pelo destinatário
- O combate ao envio de *spams* é um grande desafio
  - Difícil definição
    - Opiniões diferentes sobre uma mesma mensagem
  - Constante modificação das mensagens
    - Atualização para burlar os mecanismos anti-*spam*
  - A quantidade aumenta a cada dia
    - Lucratividade
    - Difícil punição
  - Falta de formação técnica dos usuários

Introdução

- Correio eletrônico
  - Aplicação mais afetada
    - Simplicidade do SMTP (*Simple Mail Transfer Protocol*)
  - **Dois terços** do tráfego total de correio eletrônico → *spams*
    - 170 bilhões de mensagens por dia
  - Prejuízo de bilhões de dólares para os provedores
    - Armazenar, processar e encaminhar as mensagens
  - Perda de credibilidade dos usuários
    - Baixar, ler e classificar as mensagens
  - Previsão pessimista
    - 95% do tráfego de correio eletrônico será de *spams*
- Também existem *spams* de voz e vídeo

Introdução

- Países que mais enviam *spams* (08/2006)



Fonte: Commtouch Software Online Labs, http://www.commtouch.com/Site/ResearchLab/statistics.asp

Introdução

- Brasil
  - 5º maior receptor (2005)
    1. China
    2. EUA
    3. Coréia do Sul

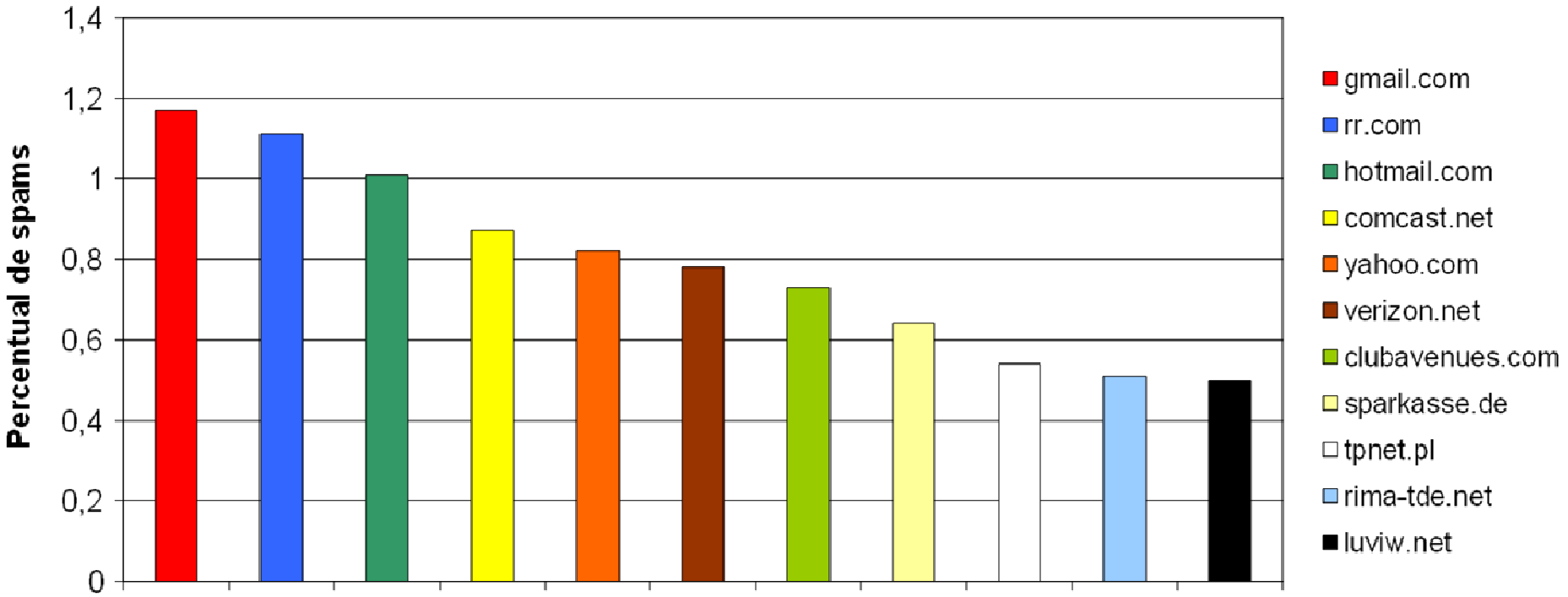
    Fonte: Agência O Globo, 08/2005
  - Um brasileiro está entre os 10 maiores *spammers*
    - Indivíduos responsáveis pelo envio de *spams*

The 10 Worst ROKSO Spammers			As at 06 August 2006
Rank	Photo	Spammer or Spam Gang	Country
6		**Ruslan Ibragimov / send-safe.com** Stealth spamware creator. One of the larger criminal spamming operations around. Runs a CGI mailer on machines in Russia and uses hijacked open proxies and virus infected PCs to flood the world with spam.	**Russia**
7		**Jeffrey Peters - JTel / CPU Solutions** Convicted felon, hard-core spammer host, Peters is also behind a fake Russian "ISP" serving many criminal ROKSO spammers. Forged documents seem to be among his specialties.	**United States** Florida
8		**Tim Govetche / Bulkers.net / Bulkbarn.com** Long time operator of Bulkbarn.com "secret" spammer chat forum. Behind a lot of the criminal method spamming that goes on in the world. Caught in 2005 spamming "pump & dump" stock scams.	**Canada** Nova Scotia
9		**Alexey Panov - ckync.com** Spamming, spammer hosting, spamware peddlers. Author of the DMS spamware that uses hijacked open proxies and virus infected PC to flood the world with spam.	**Russia**
10		**Ivo Ottavio Reali Camargo** A spammer and an "off shore" partner to many other spammers including Alan Ralsky and Michael Lindsay. Provides hosting, domains and spam services from Brazil where enforcement is lax.	**Brazil**

Fonte: The Spamhaus Project, http://www.spamhaus.org/statistics/spammers.lasso

Introdução

- Domínios que mais enviam *spams* (2006)



Fonte: Commtouch Software Online Labs, http://www.commtouch.com/Site/ResearchLab/statistics.asp

Introdução

- Sistemas anti-*spam*
  - Principal contramedida
  - Conjunto de técnicas e procedimentos
    - Prevenção da construção da lista de destinatários
    - Coibição do envio
    - Caracterização e filtragem das mensagens
    - Legislação
- Diferentes tipos
  - Baseados em filtragem simples
  - Com auto-aprendizado
  - Baseados na verificação da origem

Mensagens Eletrônicas Não Solicitadas (*Spams*)

Definição

- É o primeiro desafio
  - Inúmeras
  - Controversas

- Uma mensagem **comercial não solicitada**
  - *Unsolicited Commercial E-mail* - UCE
  - Não inclui mensagens que contêm fraudes e tentativas de golpe
- Uma mensagem **não solicitada enviada em batelada**
  - *Unsolicited Bulk E-mail* - UBE
  - Inúmeras réplicas do mesmo *spam* são enviadas

Definição

- Uma mensagem **não desejada** por um usuário
  - Geral e conflitante
  - Caráter subjetivo
    - Usuários com visões diferentes
- Uma mensagem que **não atende** a um conjunto de **regras**
  - Regras baseadas em características de mensagens classificadas como *spams*
  - Idéia das regulamentações em vigor e propostas de lei

Origem do termo

- Tão controversa quanto as definições
- SPAM
  - Marca registrada pela Hormel Foods LLC
  - Presunto apimentado enlatado
    - Contração de "***SP****iced h****AM***"
- Duas suposições relacionadas ao produto
  - Medalha entregue aos soldados americanos
    - Alimento durante a segunda guerra
    - Sinônimo de algo comum
  - Esquete do grupo Monty Python
    - Repetitivo e sem sentido

Origem do termo

O Primeiro *spam*

- 1978
- Anúncio de uma demonstração de computadores
- Enviado a Arpanet
  - Funcionário do departamento de vendas da DEC
  - Enviado **manualmente**
- "Apenas" 320 destinatários
- Grande debate sobre a utilização do correio eletrônico
- Reprodução da mensagem

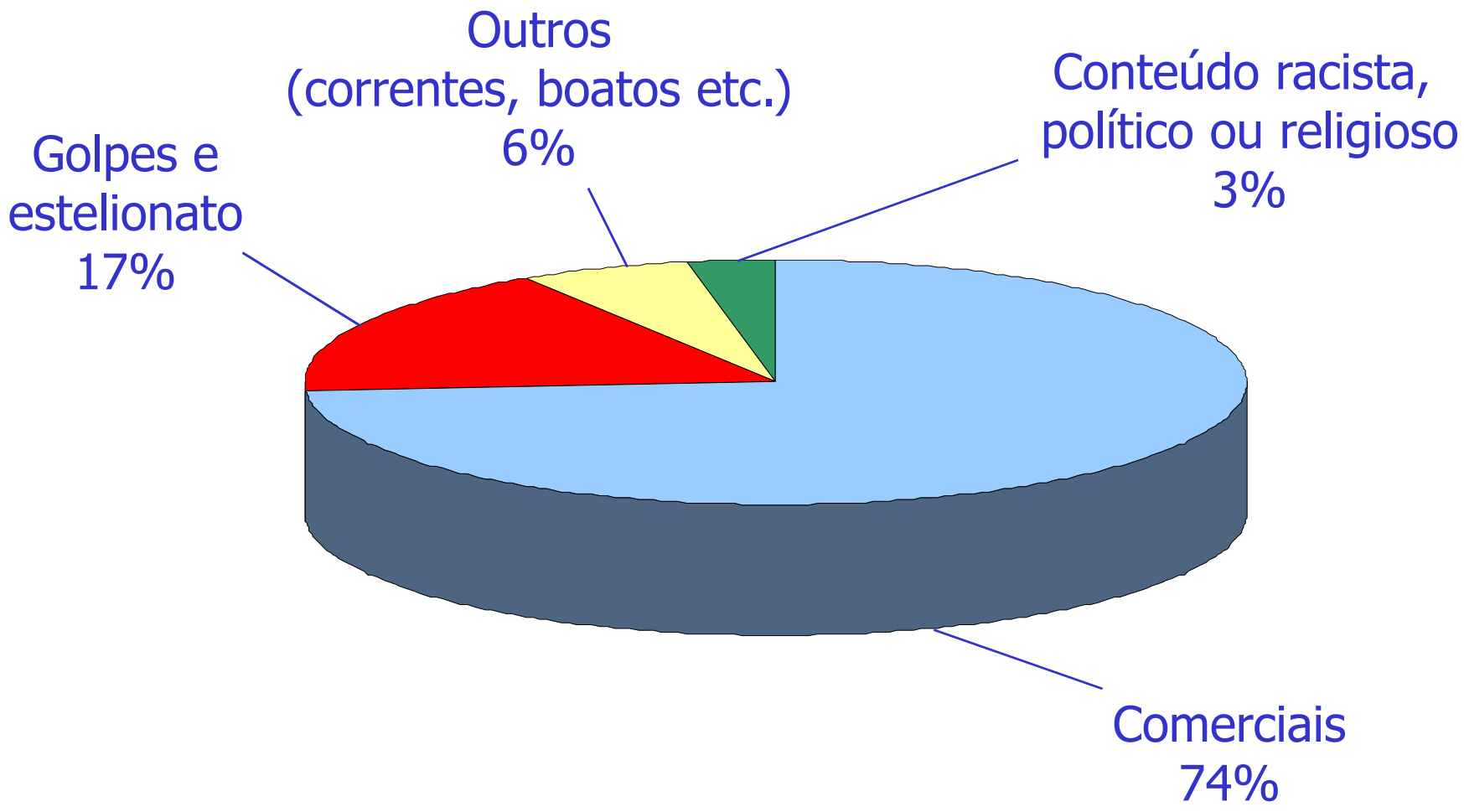
O *spam* do Green Card

- 1994
- Anúncio do prazo de inscrições na loteria de vistos de trabalho americanos
- Envio **automatizado** em um fórum da Internet
  - Casal de advogados (Canter e Siegel)
  - 6 mil grupos de discussão
- Debate: mensagem sem relação com a discussão
- Primeira associação
  - Tipo de mensagem e a palavra *spam*
- Reprodução da mensagem

Classificação dos *spams*

- Comerciais
- Golpes e estelionatos
  - *Scam*
    - *Phishing*
  - Disseminação de programas maliciosos
- Conteúdo político, religioso ou racista
- Outros
  - Correntes
  - Boatos
  - Etc.

Classificação dos *spams*



Fonte: Inside the SPAM Cartel: Trade Secrets from the Dark Side [Spammer X et al., 2004]

Comerciais

- Representam 74% dos *spams*
- Anunciam produtos e serviços diversos
  - Oportunidades financeiras
  - Sítios de conteúdo pornográfico
  - Oferta de remédios sem prescrição médica
  - Tratamentos estéticos
  - Softwares piratas
- Alguns são de empresas conhecidas
  - A maioria tem origem suspeita

Golpes (*scams*)

- Toda mensagem que contém fraude ou tentativa de golpe
- Nocivo aos destinatários
- Correspondem a 7% dos *spams*
- Histórias como pano de fundo
  - Encobrir a ação do *scammer*
  - Produtos com resultados enganosos
  - Negócios miraculosos
  - Prêmio de loteria
  - Etc.
- Suspeita de crime organizado

Golpes (*scams*)

- Um dos mais famosos: príncipe nigeriano
  - Perseguição política
  - Transferência de uma grande quantia de dinheiro para o destinatário
  - Depósito para cobrir as despesas
    - Conta bancária a ser informada
  - Mensagens **419**
    - **Lei nigeriana** sobre a prática de fraudes

Union Bank of Nigeria Plc. - Thunderbird

Arquivo Editar Exibir Ir Mensagem Ferramentas Ajuda

Receber Nova msg Catálogo Responder Re: Todos Encaminhar Excluir Spam Imprimir Parar

Assunto: Union Bank of Nigeria Plc.
De: Prince David Okafor <princedavidokafor@tiscali.no>
Responder a: princedavidokafor@ubbi.com.br
Data: 7/2/2005 06:45
Para: princedavidokafor@tiscali.no

According to Nigerian banking law, after four years, the money will revert to the ownership of the Nigerian Government. If the account owner is certified death and nobody comes forward to claim it. My proposal is that I am looking for a foreigner who will stand in as the beneficiary/next of kin.

This is simple. All you have to do is to immediately send me the details of
a bank account anywhere in the world, your address, your telephone and fax numbers for me to arrange payment of the money in your favour as the relation of late Peter Bush who died in a plane crash and the money will then be
paid to you legally for us to share in the ration of 60% for me 30% for you while the remaining 10% will be set aside for expenses re-imbursement.

I know that a transaction of this magnitude would make anyone apprehensive and worried. But I am assuring you that all will be well at the end of the day. A boldstep taken shall not be regretted, I assure you. Please do be informed that this business transaction is 100% legal and there is no risk involved. I am using my position as the bank manager and connection in the
bank here to arrange the payment in your favour as the relation of Late Mr. Peter Bush. We shall employ the services of an attorney to obtain all necessary documents and letter of administration in your favour for the payment. If you are interested, please replyimmediately through my private email address or call me for discussion as soon as you receive this mail and be rest assured that this transaction could be mostprofitable for both of us. Please do not expose this deal if you are not interested because if you do, it will take me out of seat. Please reply in strict confidence to the contact number above. As soon as i receive your response, i will send you some documents which you will show toour bank when demanded.

Sincerely with warm regards,
Prince David Okafor

Estelionato (*phishing*)

- Também é um tipo de mensagem de golpe
- Corresponde a 10% dos *spams*
- Contém **iscas**
  - Roubar dados pessoas e/ou bancários do destinatário
  - "Pescar" os dados → *fishing* → ***phishing***
  - Tentam se aproximar de uma mensagem legítima
    - Remetente acima de qualquer suspeita
  - Solicitação de recadastramento
    - Bancos, administradoras de cartão de crédito, órgãos públicos etc.

Arquivo Editar Exibir Ir Mensagem Ferramentas Ajuda

Receber Nova msg Catálogo Responder Re: Todos Encaminhar Excluir Spam Imprimir Parar

Assunto: **Ministerio da Fazenda**

De: Receita-Federal <receita@fazenda.gov.br>

Data: 14/2/2006 01:40

Para: guaresma@gta.ufrj.br

CC: receita@fazenda.gov.br

Caro contribuinte,

Devido a uma falha no sistema da Receita Federal algumas restituições do **Imposto de Renda de Pessoa Física (IRPF) 2005** foram emitidas de forma equivocada. Se você já teve seu imposto restituído e deseja verificar se possui algum valor residual ainda a ser restituído clique aqui, tendo em mãos seu CPF.

Caso o seu imposto ainda não tenha sido restituído clique aqui, também tendo em mãos seu CPF.

Caso ocorra algum erro, acesse diretamente pelo link abaixo ou ligue para 0300-78-0300.

http://www.receita.fazenda.gov.br/Aplicacoes/ATRJO/ConsRest/Atual/index.asp

Quem não solicitou crédito em conta poderá procurar uma agência do Banco do Brasil ou ligar gratuitamente para 0800-785678 e pedir a transferência dos recursos para qualquer banco em que seja correntista.

Por questões de segurança, informações e dados do contribuinte não são solicitados por e-mail em hipótese alguma.

Atenciosamente,

Coordenação de Integração Fisco-Contribuinte
Secretaria da Receita Federal

http://www.kesaptarim.gov.tr/cache/system/ReceitaFederal.exe

Estelionato (*phishing*)

- Um exemplo: seguro contra fraudes do Banco do Brasil
  - O destinatário deve clicar no atalho para ativá-lo
  - Formulário
    - Campos com dados do usuário
    - Grande semelhança com o verdadeiro
  - Ao clicar no botão de submissão
    - Dados enviados ao *spammer*

Estelionato (*phishing*)

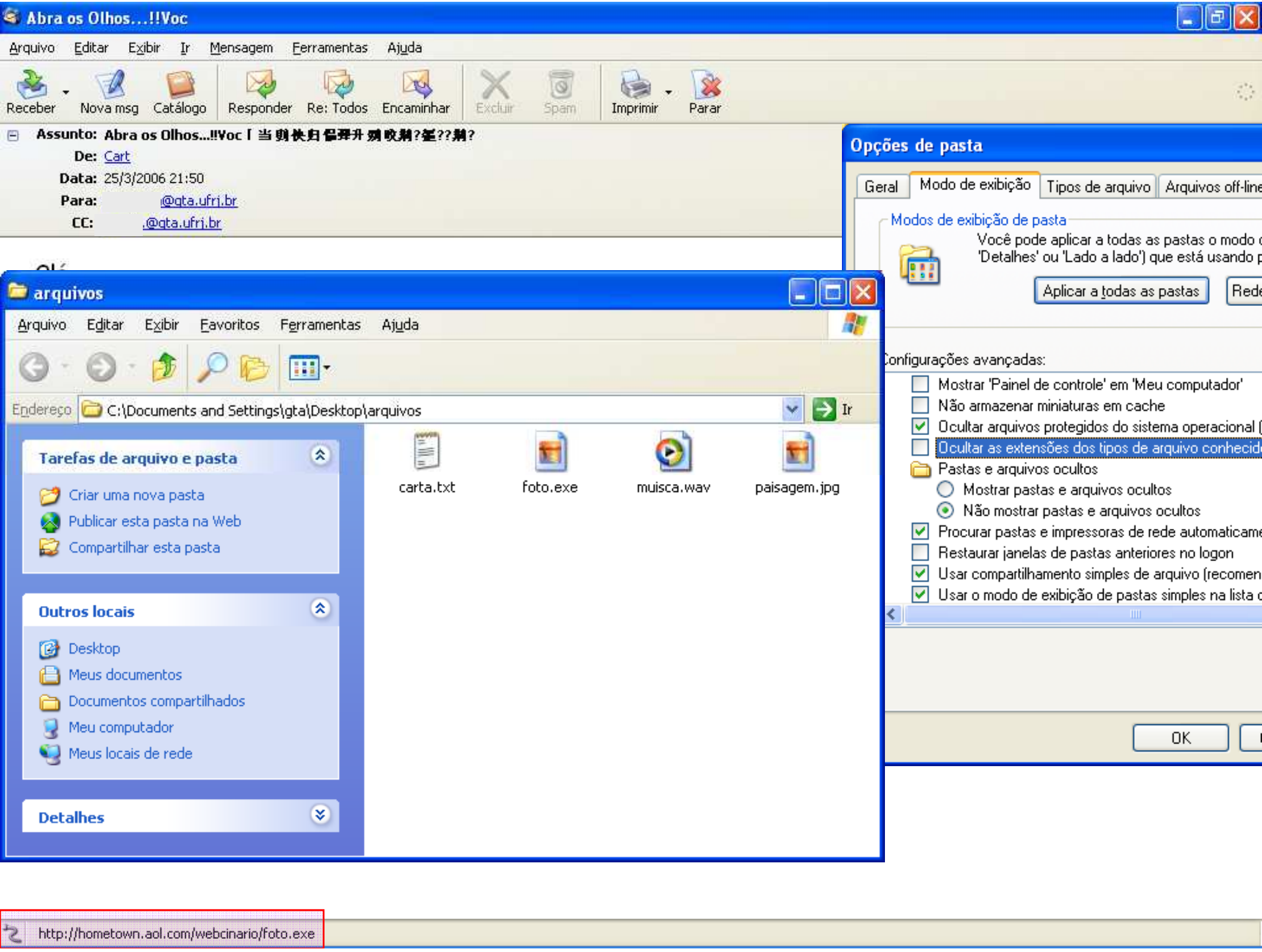
Estelionato (*phishing*)

Disseminação de programas maliciosos

- Efeito nocivo para os destinatários
- Programas maliciosos
  - Vírus
  - Vermes
  - Cavalos de Tróia
  - Etc.
- Iscas para disfarçar o real conteúdo
  - Induzir o destinatário a executar o programa
- Recrutar máquinas zumbis para enviar *spams*
- Capturar dados do destinatário

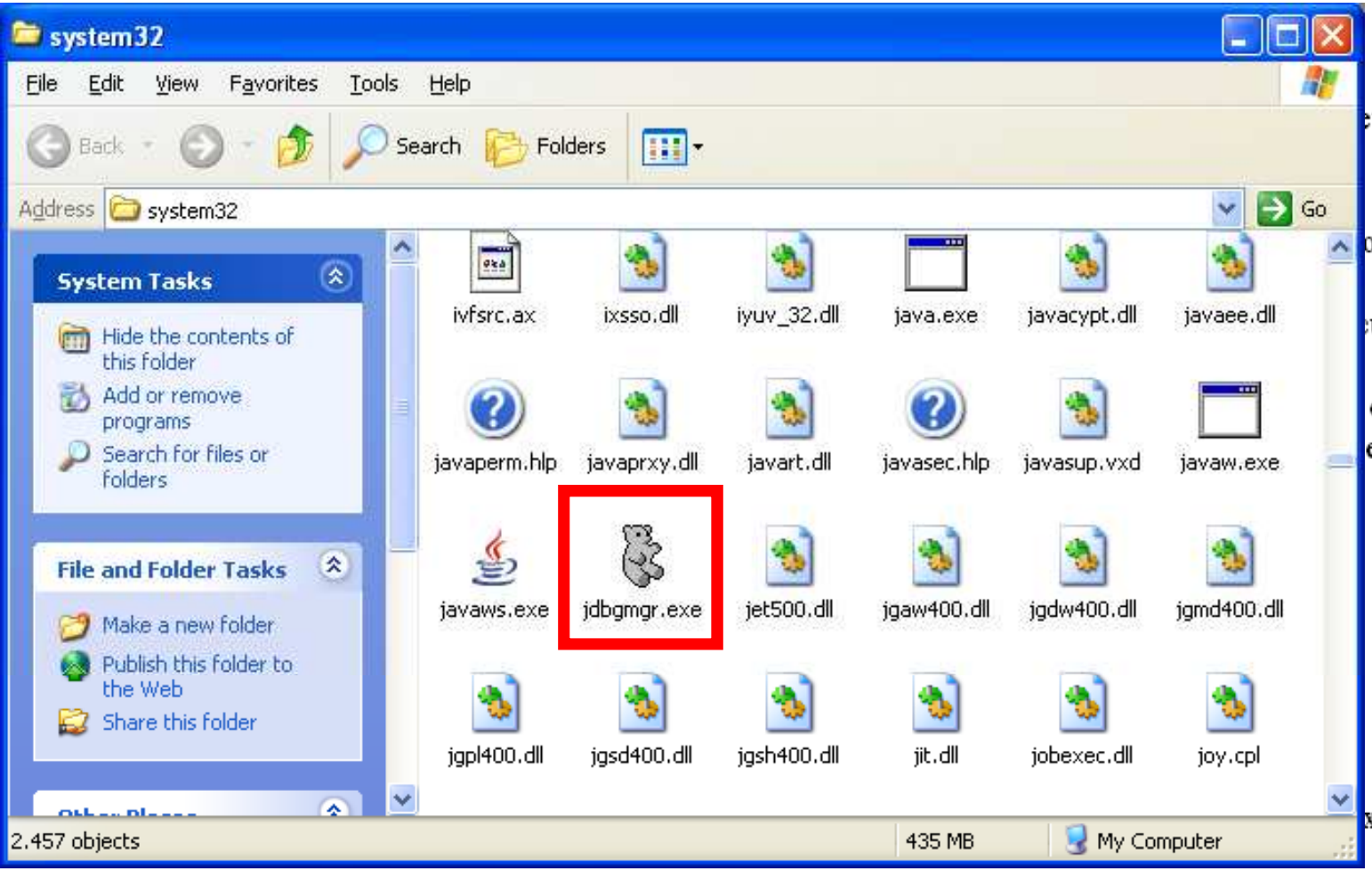
Disseminação de programas maliciosos

- Um exemplo: traição conjugal
- Destinatário é informado de uma suposta traição
  - Comprovada por fotos
- Para ver as fotos
  - Destinatário deve clicar no atalho indicado da mensagem
  - Programa executável do Windows
- Ao clicar
  - Execução de um cavalo de Tróia
  - Destinatário vulnerável à ação do *spammer*


Abra os Olhos...!!Voc
Arquivo Editar Exibir Ir Mensagem Ferramentas Ajuda
Receber Nova msg Catálogo Responder Re: Todos Encaminhar Excluir Spam Imprimir Parar
Assunto: Abra os Olhos...!!Voc
De: Cart
Data: 25/3/2006 21:50
Para: @gta.ufrj.br
CC: .@gta.ufrj.br
Opções de pasta
Geral
Modo de exibição
Tipos de arquivo
Arquivos off-line
Modos de exibição de pasta
Você pode aplicar a todas as pastas o modo
'Detalhes' ou 'Lado a lado') que está usando
Aplicar a todas as pastas
Rede
Configurações avançadas:
Mostrar 'Painel de controle' em 'Meu computador'
Não armazenar miniaturas em cache
Ocultar arquivos protegidos do sistema operacional
Ocultar as extensões dos tipos de arquivo conhecid
Pastas e arquivos ocultos
Mostrar pastas e arquivos ocultos
Não mostrar pastas e arquivos ocultos
Procurar pastas e impressoras de rede automaticame
Restaurar janelas de pastas anteriores no logon
Usar compartilhamento simples de arquivo (recomen
Usar o modo de exibição de pastas simples na lista
OK
arquivos
Arquivo Editar Exibir Favoritos Ferramentas Ajuda
Endereço C:\Documents and Settings\gta\Desktop\arquivos
Ir
Tarefas de arquivo e pasta
Criar uma nova pasta
Publicar esta pasta na Web
Compartilhar esta pasta
Outros locais
Desktop
Meus documentos
Documentos compartilhados
Meu computador
Meus locais de rede
Detalhes
carta.txt
foto.exe
muisca.wav
paisagem.jpg
http://hometown.aol.com/webcinario/foto.exe

Correntes e boatos

- Correntes
  - Prometem benefícios
    - Sorte, riqueza, saúde etc.
  - Encaminhar mensagens em um dado intervalo de tempo
- Boatos
  - Impressionar os destinatários
  - Histórias falsas
    - Crianças desaparecidas, ameaça de vírus, difamação de empresas e pessoas etc.
- Objetivo: atingir rapidamente o maior número de pessoas
  - Criação de **lendas urbanas**



t em BH e em vários grupos de

ontatos.

ulto por 14 dias antes de agir.

o teste a seguir para ver se realmente
e na obrigação de alertá-los, já que

xe;

3) assegure-se de que está procurando-o no drive C;

4) Localizar agora;

5) O virus possui um ícone em forma de ursinho (cinza);

6) Caso você o encontre, não abra de maneira alguma, delete-o imediatamente,

7) Não esqueça de retirá-lo da Lixeira pois senão nada adiantará.

Razões para a proliferação dos s*pams*

1. Facilidade de obter endereços de correio eletrônico
   - Potenciais consumidores
2. Baixo custo para produzir e enviar as mensagens
3. Número de destinatários alcançados com uma mensagem
4. **Efetividade das mensagens**
5. Possibilidade de enriquecimento ilícito

Obtenção de potenciais consumidores

- Grande popularidade do correio eletrônico
  - Facilidade de comunicação entre pessoas
  - Baixíssimo custo para enviar uma mensagem
- Divulgação de endereços
  - Prática comum
    - Empresas
    - Instituições de ensino
    - Sítios pessoais
    - Etc.

Obtenção de potenciais consumidores

- Como os endereços são disponibilizados é um problema
  - Programas robôs
    - Vasculham sítios de forma automatizada
    - Construção de listas de destinatários
    - Milhares de endereços em poucas horas
  - Criação e comércio de listas
    - Atividade lucrativa
      - US$ 100,00 a US$ 1000,00 por lista
    - Nenhum ato ilícito é cometido

de Igor Moraes
assunto **Fwd: Listas de emails do Brasil e de São Paulo**
para Igor Moraes

-------- Mensagem original --------
Assunto: Listas de emails do Brasil e de São Paulo
Data: Tue, 23 Aug 2011 12:02:50 -0300
De: Super Lista <p1efle@ult.edu.cu>
Responder a: supercontatovenda@bol.com.br
Para: igor@existeumlugar.com.br

A Super Lista é a primeira e a melhor fornecedora de emails
do mercado.

Estamos com 2 pacotes de emails atualizados. O primeiro com
listas de todo o Brasil, e o segundo com listas de São Paulo.

Para adquirir nossos novos pacotes de emails, envie-nos os dados
abaixo por email:

Desejo adquirir o(s) pacote(s) marcado(s) abaixo:
( ) R$ 150 - 120 milhões de emails de todo o Brasil
( ) R$ 100 - 5 milhões de emails de São Paulo

Nome:
Cidade:
Estado:
Telefone opcional:

Assim que recebermos seu email, enviaremos a você
os dados para aquisição da Super Lista.

Abraços,

Super Lista - Lista de emais reais

Vantagens para os anunciantes

- Correio eletrônico
  - Menor custo
    - **1 milhão → US$ 250,00 (R$ 550,00)**
  - Maior alcance geográfico
- Anúncios impressos em papel
  - Gastos com
    - Criação
    - Reprodução
    - Distribuição
      - Correio convencional
        - Carta convencional → R$ 0,15
        - **1 milhão → R$ 150 mil reais**
      - Mensageiros

Vantagens para os anunciantes

- Rádio e TV
  - Veículos de massa mais usados por anunciantes
  - Rádio → impacto sonoro
  - TV → impacto audiovisual

  } imbatíveis
  - Grande poder de penetração
  - Limitação geográfica
    - Regiões, estados e países
  - Presença física do espectador durante a exibição
  - Custo elevado
    - Produção
    - Distribuição

Vantagens para os anunciantes

- Internet
  - Novo meio de divulgação
  - Grande alcance
  - Impacto audiovisual cada vez maior
    - Novas tecnologias
- Correio eletrônico
  - Não foi criado para divulgação de propagandas
  - Simplicidade do protocolo SMTP (*Simple Mail Transfer Protocol*)
    - Remetentes e destinatários confiáveis
    - Mensagens relevantes

Vantagens para os anunciantes

- Um *spam* enviado
  - Milhares de destinatários
  - Sem limites geográficos
  - Baixo custo de divulgação
    - Transferido para os provedores de serviço
      - Responsáveis pela entrega das mensagens

- Enviar **1 milhão** de *spams* custa **US$ 250,00**
  - Provedores de acesso
  - Lista de destinatários
  - Construção das mensagens

Efetividade das mensagens

- **Maior estimulante** para o envio de *spams*
- Resultados surpreendentes
  - 39% dos usuários entrevistados já clicaram em um atalho
    - Usuários corporativos e domésticos
  - 13% dos usuários corporativos já compraram produtos
  - 11% dos usuários domésticos já compraram produtos
- Catálogos de produtos enviados via correio eletrônico
  - 12 vezes mais respostas que os impressos
- Volume de mensagens
  - Baixa taxa de resposta já seria lucrativa

Enriquecimento ilícito

- Mensagens com tentativas de golpe
- Atrativo para os *spammers*
  - Muitos destinatários são "fisgados" pelas iscas
  - Iscas cada vez mais sofisticadas
- Golpe: fornecimento de dados pessoais e bancários
  - Compras em cartão de crédito
  - Transferências bancárias
- Crime organizado
  - Ação da polícia

Prejuízos causados

- Usuários de correio eletrônico
  - Conteúdo das mensagens
  - Quantidade das mensagens

- Provedores de serviço
  - Prejuízo de bilhões de dólares
    - US$ 2800,00 para cada milhão de *spams* enviados
  - Aumento dos gastos
    - Infra-estrutura
    - Pessoal

Prejuízos causados - usuários

- Desperdício de tempo
  - Baixar, abrir e identificar um *spam*
- Gastos desnecessários
  - O usuário paga para acessar a Internet
- Não recebimento de mensagens legítimas
  - Caixa postal do usuário cheia
  - Filtros anti-*spam* usados pelos provedores
- Tentativas de golpe
- **Perda de credibilidade**
  - Correio eletrônico e Internet

Prejuízos causados - provedores

- Responsáveis por encaminhar as mensagens
  - Inclusive as não solicitadas
- Oferecem serviços gratuitos de correio eletrônico
  - Usados pelos *spammers*
- Custos desnecessários
  - Banda passante, memória e processamento
  - Receber, armazenar e processar as mensagens
    - Apagadas pelos usuários
- Gastos para combater os *spams*
  - Emprego de sistemas anti-*spam*
  - Equipes de atendimento ao cliente e manutenção

Legislação atual

- Tentativa para coibir o envio de *spams*
- Objetivos
  - Determinar o que é um *spam*
  - Estabelecer punições aos responsáveis
- Em discussão em diversos países
  - Estados Unidos
  - Brasil

Legislação atual - EUA

- Pioneiros na discussão de leis anti-*spam*
  - 35% dos *spams* são originados nos EUA
- Estatuto CAN-SPAM (*Controlling the Assault of Non-Solicited Pornography and Marketing Act*)
  - Em vigor desde 2004
  - Definido pelo FTC (*Federal Trade Commision*)
  - Válido em todo o território americano
  - Baseado em leis estaduais já existentes

O que é *spam*? - FTC

- Qualquer mensagem eletrônica
  - Conteúdo comercial
  - Enviada em batelada
  - Sem a requisição ou consentimento prévio do destinatário

- Base das regras do CAN-SPAM

Regras do CAN-SPAM

- Origem da mensagem
  - Informações verdadeiras
- Assunto da mensagem
  - Relacionado com o próprio conteúdo
- Propósito da mensagem
  - Indicado claramente se for uma propaganda
- Endereço físico do remetente
  - Presente na mensagem
  - Enviar reclamações ou denúncias
- Opção de não receber mais mensagens semelhantes do mesmo remetente

Penas previstas no CAN-SPAM

- Pagamento de multas
  - Violar uma das regras
    - Até US$ 11.000
  - Criar listas de destinatários
    - Coleta de endereços
      - Proibição do uso deve ser explícita
    - Ataques do tipo dicionário
  - Automatizar registros
    - Serviços gratuitos de correio eletrônico
  - Explorar falhas de configuração de servidores

Penas previstas no CAN-SPAM

- Prisão
  - Utilizar computadores de terceiros sem autorização
  - Tentar iludir destinatários sobre a origem da mensagem
  - Falsificar informações no cabeçalho de múltiplas mensagens
  - Iniciar a transmissão de múltiplas mensagens
  - Utilizar endereços IP falsos para enviar *spams*
  - Criar contas de correio eletrônico com informações falsas
- "Múltiplas" mensagens
  - Mais de 100 em 24 horas
  - Mais de 1000 em 30 dias
  - Mais de 10000 em 1 ano

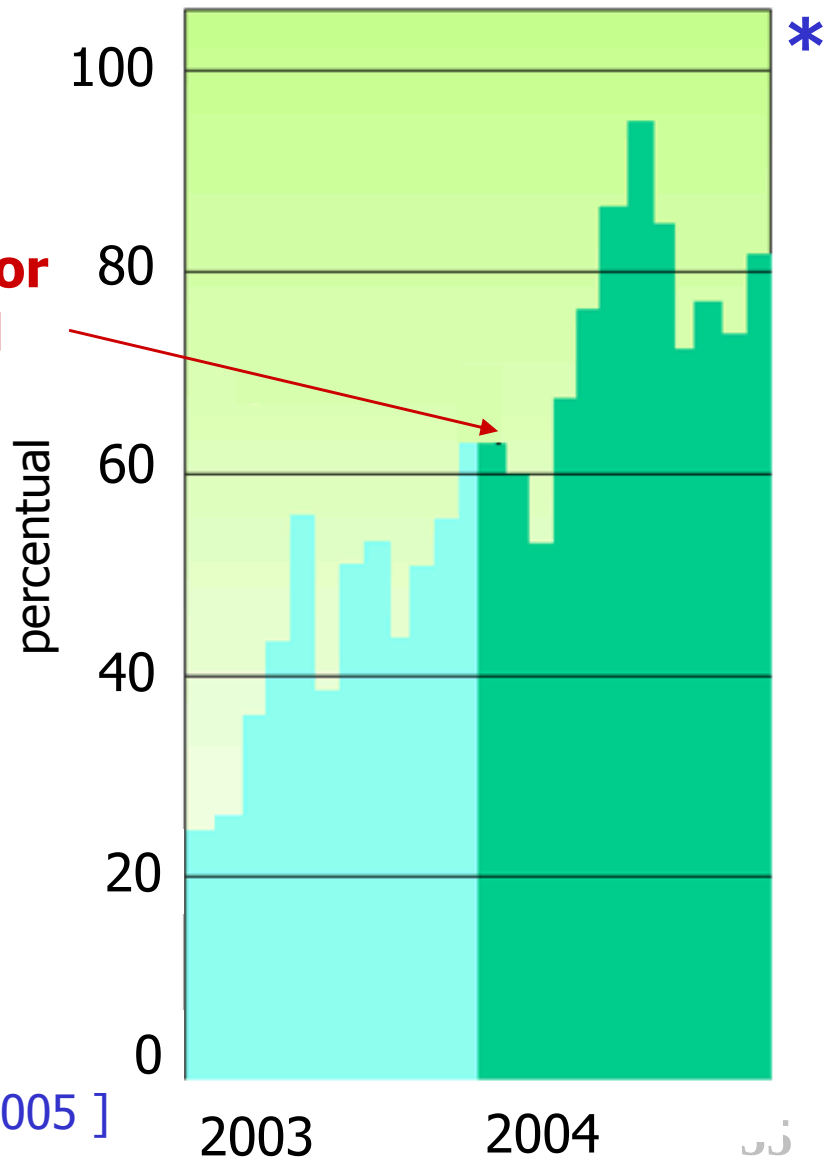
Críticas ao CAN-SPAM

- Mostrar os limites de atuação aos *spammers*
  - Define-se legalmente o que é um *spam*
    - É possível criar um *spam* atendendo o estatuto
- Dificuldade para se identificar o remetente
  - Remetentes falsos
  - Fora do território americano
- Liberdade de expressão x inibição ao envio de *spams*
  - Brechas para os *spammers*
- Medidas legais serão sempre ineficientes
  - Não acompanham a evolução tecnológica

Ineficiência do CAN-SPAM

Percentual de *spams* em relação ao tráfego total de correio eletrônico

Medidas legais são pouco efetivas



* Reproduzido de Pfleeger e Bloom [Pfleeger e Bloom, 2005]

Legislação atual - Brasil

- Não há legislação anti-*spam* em vigor
  - Projeto de lei 021/04
    - Parecer favorável na Comissão de Constituição, Justiça e Cidadania (CCJ) no Senado
  - Grupos de discussão
    - Brasil AntiSPAM
      - Normas de conduta
    - Antispam.br
      - Cartilha de uso do correio eletrônico
- Mensagens em conformidade com a lei brasileira
  - Enganar os destinatários

O que é *spam*? - projeto de lei 021/04

- Conteúdo comercial ou publicitário
- Enviado a mais de 500 destinatários em 96 horas
- Autorizada previamente pelo destinatário
- Objetivo claro
- Origem verdadeira do remetente
- Opção para não receber mensagens semelhantes do mesmo remetente

O que é *spam*? - Grupo Brasil AntiSPAM

- Pelo menos **dois dos itens** devem ser **atendidos**
  - o remetente é inexistente ou possui identidade falsa
  - o destinatário não autorizou previamente o envio da mensagem
  - o destinatário não pode optar em não receber mais a mensagem
  - o assunto não condiz com o conteúdo da mensagem
  - a sigla NS (Não Solicitado) está ausente no assunto de uma mensagem que não foi previamente requisitada
  - o remetente não pode ser identificado
  - uma mensagem semelhante foi recebida anteriormente em menos de dez dias apenas com o remetente ou assunto diferentes

Adaptação de leis existentes

- Legislação anti-*spam* brasileira
  - Tão ineficiente quanto o CAN-SPAM
- Código penal
  - Artigo 161 – Usurpação
    - Detenção de 1 a 6 meses e multa
  - Artigo 163 – Danos
    - Detenção de 6 meses a 3 anos e multa
  - Artigo 171 – Estelionato
    - Reclusão de 1 a 5 anos e multa
- Código de defesa do consumidor
  - Artigos 36 e 37 – Publicidade velada e propaganda abusiva

Técnicas para o Envio de *Spams*

Técnicas para o envio de *spams*

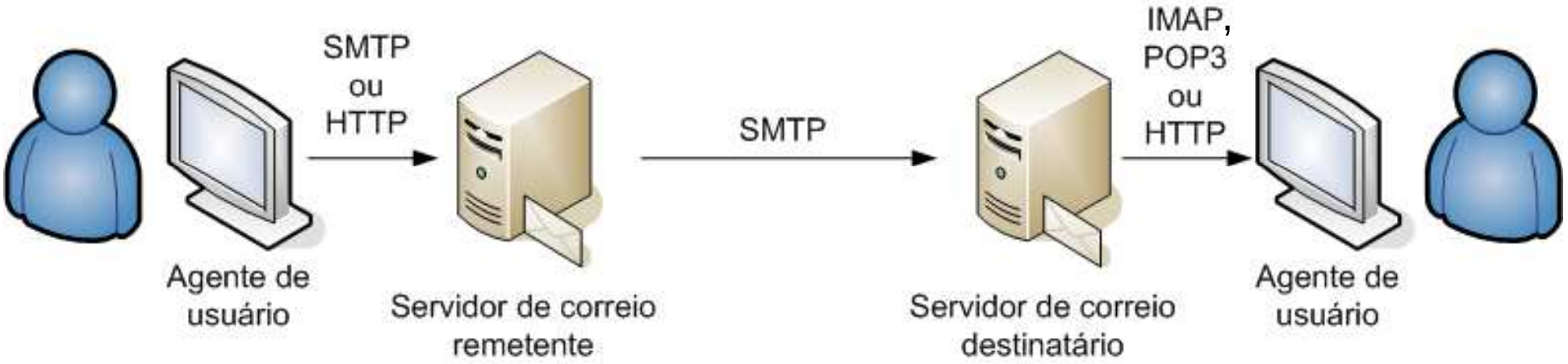
- Três fases principais
  - Coleta de dados
  - Formato (edição) das mensagens
  - Envio das mensagens

Sistema de Correio na Internet

Sistema de correio da Internet

- Composto por:
  - Agentes de usuário
  - Servidores de correio ou agentes de transferência de mensagens
  - Protocolo simples de transferência de correio (*Simple Mail Transfer Protocol* – SMTP)
  - Protocolos de acesso a correio

Sistema de correio da Internet

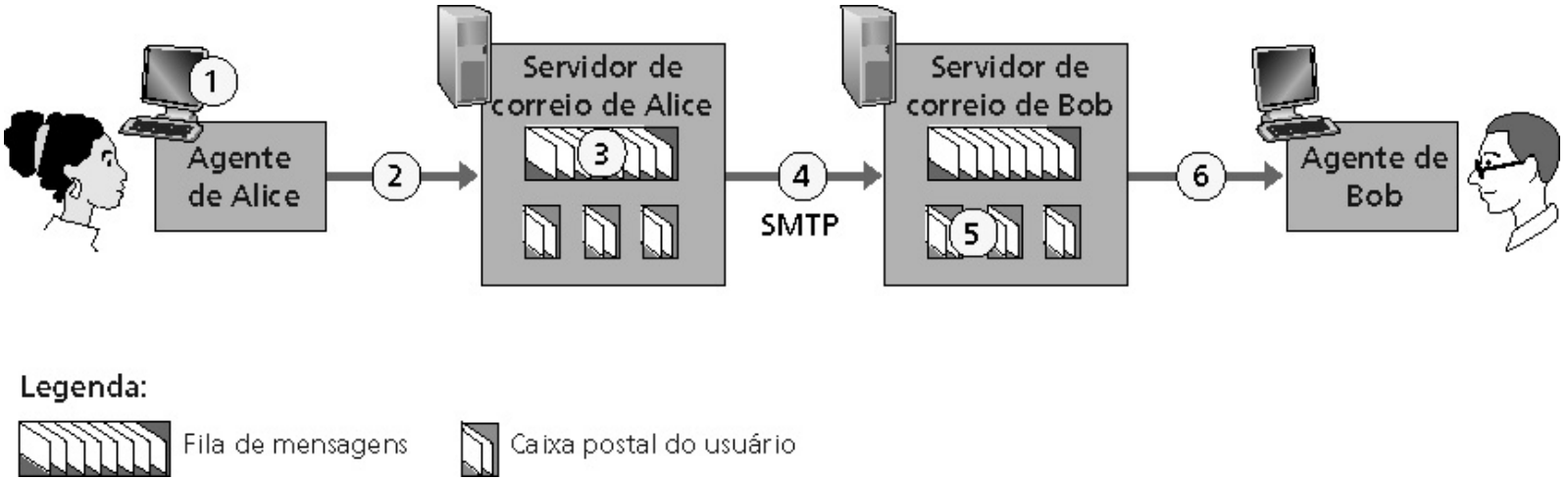


Sistema de correio da Internet

- Agentes de usuário
  - Permitem que usuários leiam, respondam, encaminhem, salvem e editem mensagens
  - Exemplos: Outlook, Eudora, Thunderbird, Mutt
- Servidores de correio
  - Armazenam as mensagens
  - Se comunicam para realizar a transferência das mensagens
- SMTP
  - Transfere mensagens entre servidores de correio
- Protocolos de acesso a correio
  - Transferem mensagens do servidor de correio para o agente de usuário

Cenário: Alice envia uma msg para Bob

1) Alice usa o AU para compor uma mensagem "para" `bob@someschool.edu`
2) O AU de Alice envia a mensagem para o seu servidor de correio; a mensagem é colocada na fila de mensagens
3) O lado cliente do SMTP abre uma conexão TCP com o servidor de correio de Bob
4) O cliente SMTP envia a mensagem de Alice através da conexão TCP
5) O servidor de correio de Bob coloca a mensagem na caixa de entrada de Bob
6) Bob chama o seu AU para ler a mensagem



Sistema de correio da Internet - SMTP

- Descrito na RFC 2821
- Usa o TCP e a porta 25
- Mensagens enviadas
  - Em ASCII (7 bits)
  - Uso de extensão ou de codificação para 8 bits
- Comunicação entre um cliente SMTP (transmissor) e um servidor SMTP (receptor)

Sistema de correio da Internet - SMTP

- Utiliza comandos para fazer a comunicação entre servidores
- Exemplos
  - **HELO**
  - **MAIL FROM**
  - **RCPT TO**
  - **DATA**
  - **QUIT**
  - **VRFY**

Exemplo de interação (telnet servidor.br 25)

```
S: 220 servidor.br
C: HELO cliente.br
S: 250 Hello cliente.br, pleased to meet you
C: MAIL FROM: <usuario@cliente.br>
S: 250 usuario@cliente.br... Sender ok
C: RCPT TO: <usuario@servidor.br>
S: 250 usuario@servidor.br ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: From: usuario@cliente.br
C: To: usuario@servidor.br
C: Subject: Teste
C:
C: Teste de envio de correio.
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 servidor.br closing connection
```

Sistema de correio da Internet

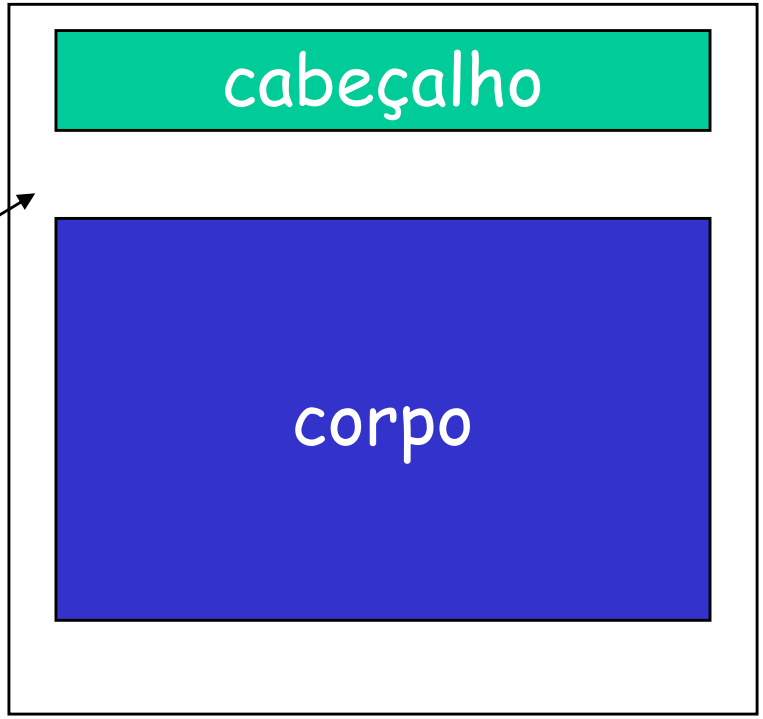
- Correio eletrônico formado por
  - Envelope
    - Encapsula a mensagem
    - Contém as informações necessárias para o transporte da mensagem
  - Mensagem
    - Composta de cabeçalho e corpo

Exemplo de interação (telnet servidor.br 25)

```
S: 220 servidor.br
C: HELO cliente.br
S: 250 Hello cliente.br, pleased to meet you
C: MAIL FROM: <usuario@cliente.br>
S: 250 usuario@cliente.br... Sender ok
C: RCPT TO: <usuario@servidor.br>
S: 250 usuario@servidor.br ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: From: usuario@cliente.br
C: To: usuario@servidor.br
C: Subject: Teste
C:
C: Teste de envio de correio.
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 servidor.br closing connection
```

Sistema de correio da Internet

- Correio eletrônico formado por
  - Mensagem
    - Campos de cabeçalho
      - Exemplos
        - `From:`
        - `To:`
        - `Subject:`
        - `Received:`
    - Corpo
      - Só diz respeito ao destinatário



Exemplo de mensagem

```
Received: from cliente.br by servidor.br; 16 Jul 06 10:30:01
GMT
Received: from maquina.cliente.br by cliente.br; 16 Jul 06
10:29:58 GMT
From: usuario@cliente.br
To: usuario@servidor.br
Subject: Teste

Teste de envio de correio.
```

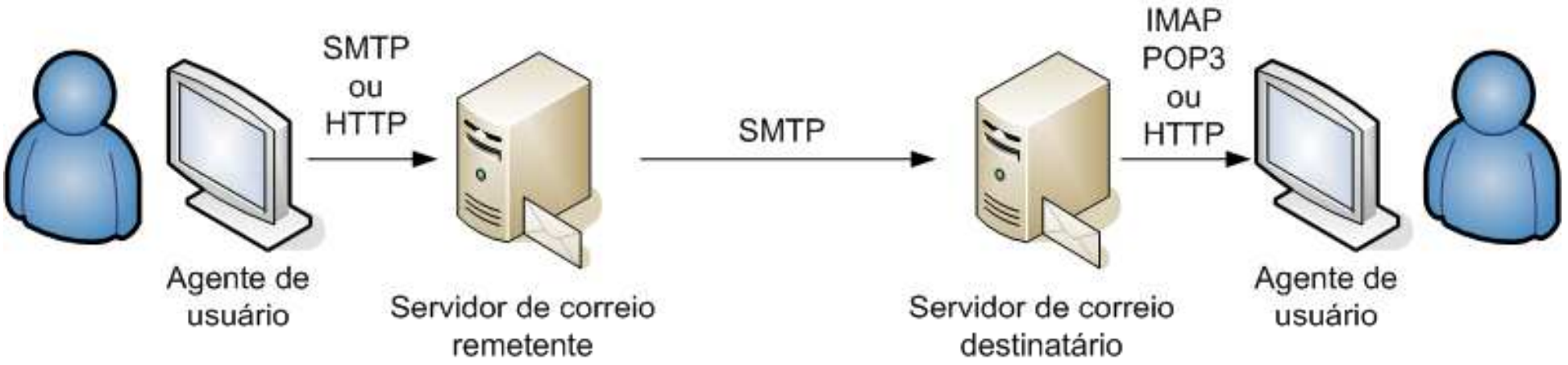
SMTP

- Usa conexões persistentes
- Requer que a mensagem (cabeçalho e corpo) sejam em ASCII de 7-bits
- Servidor SMTP usa `CRLF.CRLF` para reconhecer o final da mensagem

SMTP x HTTP

- HTTP: *pull* (recupera)
- SMTP: *push* (envia)
- Ambos têm interação comando/resposta e códigos de status em ASCII
- HTTP
  - Cada objeto é encapsulado em sua própria mensagem de resposta
- SMTP
  - Múltiplos objetos de mensagem enviados numa mensagem de múltiplas partes (MIME)

Protocolos de Acesso ao Correio



- SMTP
  - Entrega/armazenamento no servidor do receptor
- Protocolo de acesso ao correio
  - Recupera do servidor

Protocolos de Acesso ao Correio

- POP: Post Office Protocol [RFC 1939]
  - Autorização (agente <-->servidor) e **transferência**
- IMAP: Internet Mail Access Protocol [RFC 1730]
  - Mais comandos (mais complexo)
  - Manuseio de mensagens **armazenadas no servidor**
- HTTP
  - Hotmail , Yahoo! Mail, Webmail, etc.

Protocolo POP3

fase de autorização

- comandos do cliente:
  - **`user:`** declara nome
  - **`pass:`** senha
- servidor responde
  - **`+OK`**
  - **`-ERR`**

fase de transação, cliente:

- **`list:`** lista números das msgs
- **`retr:`** recupera msg por número
- **`dele:`** apaga msg
- **`quit`**

```
S: +OK POP3 server ready
C: user ana
S: +OK
C: pass faminta
S: +OK user successfully logged on
```

```
C: list
S: 1 498
S: 2 912
S: .
C: retr 1
S: <message 1 contents>
S: .
C: dele 1
C: retr 2
S: <message 1 contents>
S: .
C: dele 2
C: quit
S: +OK POP3 server signing off
```

POP3 x IMAP

Mais sobre o POP3

- O exemplo anterior usa o modo "*download* e delete".
- Bob não pode reler as mensagens se mudar de cliente
- "*Download*-e-mantenha": copia as mensagens em clientes diferentes
- POP3 não mantém estado entre conexões

IMAP

- **Mantém todas as mensagens num único lugar: o servidor**
  - Mensagens não precisam ser "baixadas" por inteiro
- Permite ao usuário organizar as mensagens em pastas
- O IMAP mantém o estado do usuário entre sessões:
  - nomes das pastas e mapeamentos entre as IDs das mensagens e o nome da pasta

Protocolos de Acesso ao Correio

- POP: Post Office Protocol [RFC 1939]
  - Autorização (agente <-->servidor) e **transferência**
- IMAP: Internet Mail Access Protocol [RFC 1730]
  - Mais comandos (mais complexo)
  - Manuseio de mensagens **armazenadas no servidor**
- HTTP
  - Hotmail , Yahoo! Mail, Webmail, etc.

Problema do SMTP

Sistema de correio da Internet - SMTP

- Assume entidades confiáveis
- Problema: falta de um mecanismo de autenticação
  - *Spammers* podem fazer uma mensagem parecer ter sido originada de qualquer endereço
  - "Solução": uso da extensão SMTP-AUTH (RFC 2554)
    - Permite a verdadeira identificação do remetente
    - Porém autentica o usuário remetente
      - Usuário autenticado pode falsificar o remetente do envelope ou o campo **`From:`** da mensagem

Exemplo de interação (telnet servidor.br 25)

```
S: 220 servidor.br ESMTP
C: EHLO cliente.br
S: 250-servidor.br
S: 250-PIPELINING
S: 250-8BITMIME
S: 250-SIZE 255555555
S: 250 AUTH LOGIN PLAIN CRAM-MD5
C: auth login
S: 334 VXNlcm5hbWU6
C: avlsdkfj
S: 334 UGFzc3dvcmQ6
C: lkajsdfvlj
S: 235 ok, go ahead (#2.0.0)
C: MAIL FROM: usuario@cliente.br
S: 250 usuario@cliente.br... Sender ok
                                  ...
```

Exemplo de interação (telnet servidor.br 25)

```
S: 220 servidor.br ESMTP
C: EHLO cliente.br
S: 250-servidor.br
S: 250-PIPELINING
S: 250-8BITMIME
S: 250-SIZE 255555555
S: 250 AUTH LOGIN PLAIN CRAM-MD5
C: auth login
S: 334 VXNlcm5hbWU6
C: avlsdkfj
S: 334 UGFzc3dvcmQ6
C: lkajsdfvlj
S: 235 ok, go ahead (#2.0.0)
C: MAIL FROM:e= usuario@cliente.br AUTH=e+3D usuario@cliente.br
S: 250 usuario@cliente.br... Sender ok
                                    ...
```

Coleta de Dados

Coleta de dados

- Nessa etapa os *spammers* constroem as listas de destinatários
- Quanto maior o número de endereços válidos, melhor
- Principais técnicas de obtenção de endereços
  - Varredura
  - Invasão de sítios
  - Utilização de ataques de força bruta
  - Uso de ofertas ou concursos
  - Compra de listas prontas de endereços

Coleta de dados - varredura

- Varredura de lugares onde endereços são divulgados
  - Grupos de notícias
  - Listas de distribuição
  - Salas de bate-papo
  - *Bulletin boards*
  - Bases de dados livres
  - Páginas *web*
- Forma mais simples de obtenção de endereços
- Surgiu no fim da década de 90
- Conhecida como coleta de endereços (*harvesting*)

Coleta de dados - varredura

- Inicialmente a busca era manual
  - Procura pelo símbolo "@" e extração correspondente
- Robôs automatizados foram introduzidos
  - Para realizar a tarefa com maior rapidez e menor esforço

Coleta de dados - varredura

- Possíveis soluções
  - Troca de "@" por "at"
  - Uso de espaços em branco entre os caracteres
  - Uso de figuras que contêm o endereço (ao invés de texto)

- Páginas *web* com o comando **`mailto:`** facilitam a busca
  - Solução: pode-se usar códigos em JavaScript

Coleta de dados - varredura

- Listas de discussão também são usadas
  - Pode-se obter diretamente endereços em listas abertas
  - Pode-se fazer a inscrição em uma "lista fechada"

- Mensagens com diversos endereços no campo **To:** facilitam a atuação dos *spammers*
  - Fato agravado quando a mensagem é encaminhada
    - Solução: Uso do campo **Bcc:**

- Base de dados *whois*
  - Dados dos responsáveis pelos domínios

Coleta de dados - varredura

- Principal problema da coleta
  - Falta de correlação dos usuários listados com os produtos oferecidos

Coleta de dados - ataques

- Ataques ou invasões de sítios têm sido explorados
- Informações como números de cartões de crédito deixadas de lado
  - Busca de endereços e informações relacionadas
- Invasões procuram explorar falhas de segurança dos SOs e dos aplicativos
- Lojas *on-line* são um dos principais alvos
  - Usam criptografia para comunicação com o servidor
  - Dados dos clientes armazenados em texto simples

Coleta de dados - ataques

- Invasão de servidores específicos pode gerar listas de usuários identificados com os produtos
  - Exemplo: sítio de um clube de futebol

- Programas maliciosos também são usados para obter endereços
  - Acessam os dados, criam as listas e enviam aos *spammers*

Coleta de dados - uso de dicionários

- Ataques de força bruta ou de verificação em massa
  - Adivinhação dos nomes dos usuários
- Usuários geralmente utilizam seus nomes, apelidos, nomes de seus personagens preferidos
  - Exemplo: silva@meudominio.com.br
- Previsibilidade de nomes leva ao uso de dicionários para gerar potenciais endereços

- Técnica de força bruta é a principal responsável por uma pessoa receber *spam* sem ter divulgado o endereço

Coleta de dados - concursos ou ofertas

- Solicitação de dados para participação em concursos ou ofertas
- Na maioria das vezes os anúncios são falsos
- Dados do usuário podem ser repassados
  - Alguns anúncios indicam esse fato em letras miúdas
  - Muitos afirmam que grandes empresas de software repassam os contatos dos usuários

Coleta de dados - compras de listas

- Listas prontas de endereços podem ser negociadas
  - Diretamente entre *spammers*
  - Livremente através de *spams*
- Maioria das listas contém endereços já utilizados de forma abusiva

Coleta de dados - validação de endereços

- É interessante validar os endereços
- Validação dos endereços realizada de várias formas
  - Mensagens são enviadas e verificadas
    - As que não retornarem com erro (usuário inexistente) serão válidas
      - Pode haver um limite de tentativas mal sucedidas de uso do comando **RCPT TO**
  - Uso do comando **VRFY**
    - Alguns administradores desabilitam o seu uso

Coleta de dados - validação de endereços

- Validação dos endereços realizada de várias formas
  - Referência a uma figura armazenada remotamente com um atalho contendo o endereço do destinatário

```
<img
src="http://dominiospammer.com.br/figs/ferias.jpg?emai
l=usuario@meudominio.com.br">
```

Coleta de dados - validação de endereços

- Validação dos endereços realizada de várias formas
  - Referência a uma figura armazenada remotamente com um atalho contendo o endereço do destinatário (cont.)
    - No arquivo de registro (*log*) do *spammer*:

```
10.0.0.1 - - [04/Aug/2006:13:45:34 -0300] "GET
  /figs/ferias.jpg?email=usuario@meudominio.com.br
  HTTP/1.1" 200 5035 "-" "Mozilla/5.0 (X11; U; Linux
  i686; en-US; rv:1.8) Gecko/20060313 Fedora/1.5-6
  Thunderbird/1.5"
```

    - Thunderbird não abre automaticamente uma figura remota
      - Usuário ainda pode clicar na figura
  - Opção de remoção do endereço de uma lista

Exemplo de *spam* com opção de remoção de uma lista

```
<p align=3D"center"><font face=3D"Verdana, Arial, Helvetica,
sans-serif" s=
ize=3D"1">
Caso deseje remover seu nome desta lista, <u><font
color=3D"#0000ff">
<a
href=3D"http://dominiospammer.com.br/out/outlist.asp?email=3
D=
\%25TO_EMAIL">
clique aqui</a></font></u><a
href=3D"http://www.dominiospammer.com.br/sair/out/o=
utlist.asp?email=usuario@meudominio.com.br">.</a></font></p>
```

Coleta de dados - perfil do usuário

- *Spammers* tentam obter também algo que identifique que um usuário é um potencial cliente
  - Perfil do usuário pode ser obtido
    - Listas
    - Invasão de servidores específicos
    - Busca em sítios com palavras chaves relacionadas

Formato das mensagens

- Inicialmente não havia preocupação com o texto
- Atualmente para algumas expressões
  - Evita-se colocá-las nos *spams*
  - São criados artifícios para tentar escondê-las
- Principais formatos
  - Texto simples (*plaintext*)
  - Texto formatado em HTML

Formato das mensagens - texto simples

- Grande parte dos *spams* ainda utiliza o texto simples
- Principalmente porque
  - Maior dificuldade de identificação pelos filtros anti-*spam*
  - Alguns programas de correio não aceitam outros formatos como o HTML
- Não costumam atrair a atenção do usuário
  - Baixa taxa de retorno

Formato das mensagens - HTML

- Preferido pelos *spammers* atualmente
  - Pode-se colocar textos que piscam ou figuras
    - Chamam a atenção dos leitores
- Uso incorreto da linguagem HTML e de marcações "inúteis" permitem uma identificação mais fácil dos *spams*

Formato das mensagens - artifícios

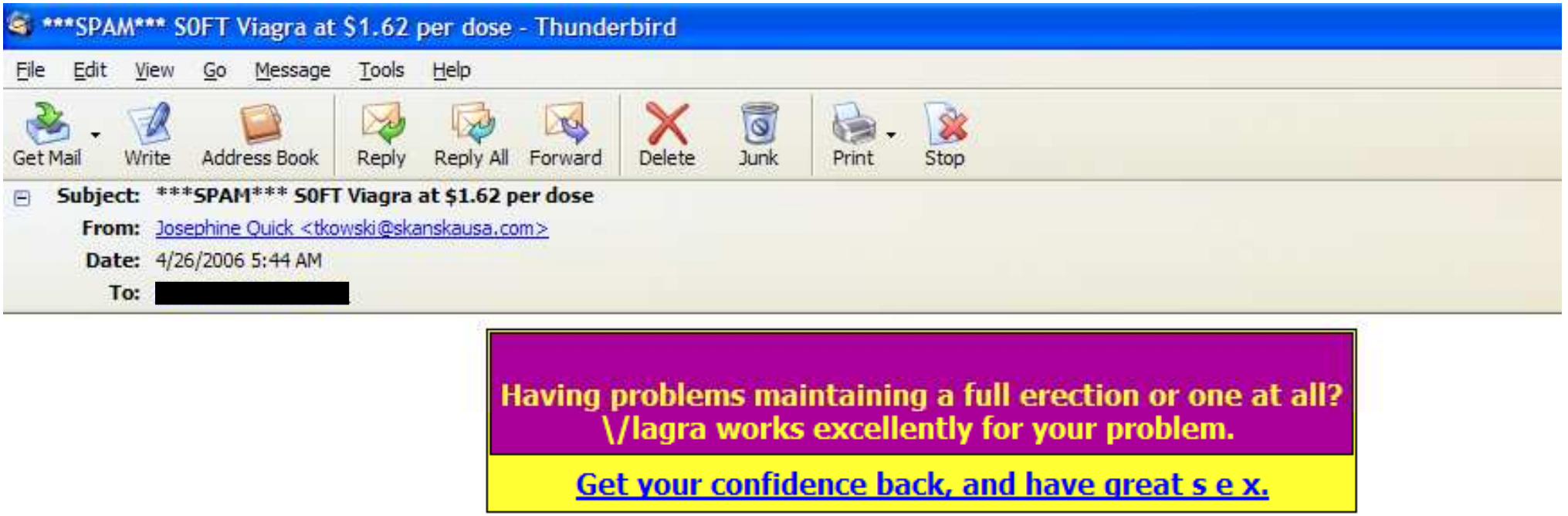
- *Spammers* tentam enganar os programas anti-*spams*
  - Principalmente modificando as palavras mais comuns
    - Exemplos: Viagra, Cialis
  - Através do uso de uma série de outros artifícios

Formato das mensagens - artifícios

- Troca da forma de escrever
  - Exemplo: Vi@gr@, V-i-@-g-r-@, \/iagra
  - Variações tornam-se comuns → filtradas

- Uso do HTML
  - Caracteres invisíveis ao olho nu
    - Figuras de 1x1 pixel entre os caracteres
  - Comentários

Exemplo de mensagem



Exemplo de mensagem

```
Having <font>prob</font><font>lems</font>
maintaining a full erection or=20
one at all?<br>
<font>\/l</font><font>agra</font> works
excellently for your=20
<font>pro</font><font>blem.</font>
```

Formato das mensagens - artifícios

- Uso do HTML (cont.)
  - Codificações
    - Exemplo:
      - mailto: → mailto:

Formato das mensagens - artifícios

- Uso da técnica *snowflaking messages*
  - Introduz texto aleatório
    - Programa anti-*spam* pensa que o correio é pessoal e individual
    - Servidores reportam a listas negras quando recebem uma grande quantidade de mensagens iguais
      - Aleatoriedade torna difícil a identificação das cópias

Exemplo de mensagem com texto aleatório

Hi,

VIAGvRA from 3, 35 $

AMBIvEN

VALIvUM from 1, 20 $

CIALxIS from 3, 75 $

http://www.lazexionvertin.com

at the height of their wealth and skill: straight as a ruler,

smooth-floored and smooth-sided, going with a gentle never-varying slope

direct-to some distant end in the blackness below.

Exemplo de mensagem com texto aleatório

Hi,

VALIvUM from 1, 20 $

AMBIvEN

CIALxIS from 3, 75 $

VIAGvRA from 3, 35 $

http://www.lazexionvertin.com

Fili and Kili and the hobbit went along the shore to the great bridge.

There were guards at the head of it, but they were not keeping very

careful watch, for it was so long since there had been any real need.

Formato das mensagens - artifícios

- Modificação dos campos **`From:`** e **`Received:`**
  - Também dificulta a identificação dos *spammers*

Envio das mensagens

- *Spammers* procuram se manter no anonimato
  - *Spams* podem conter esquemas de fraudes, estelionato, ofertas de produtos inexistentes
- *Spammers* não enviam as mensagens de seus servidores
  - Mesmo falsificando os campos **`From:`** e **`Received:`** poderia-se chegar ao spammer
  - É necessária uma grande banda para enviar as mensagens
- Utilizam-se
  - Servidores alheios
  - Contas em serviços de *webmail* gratuitos

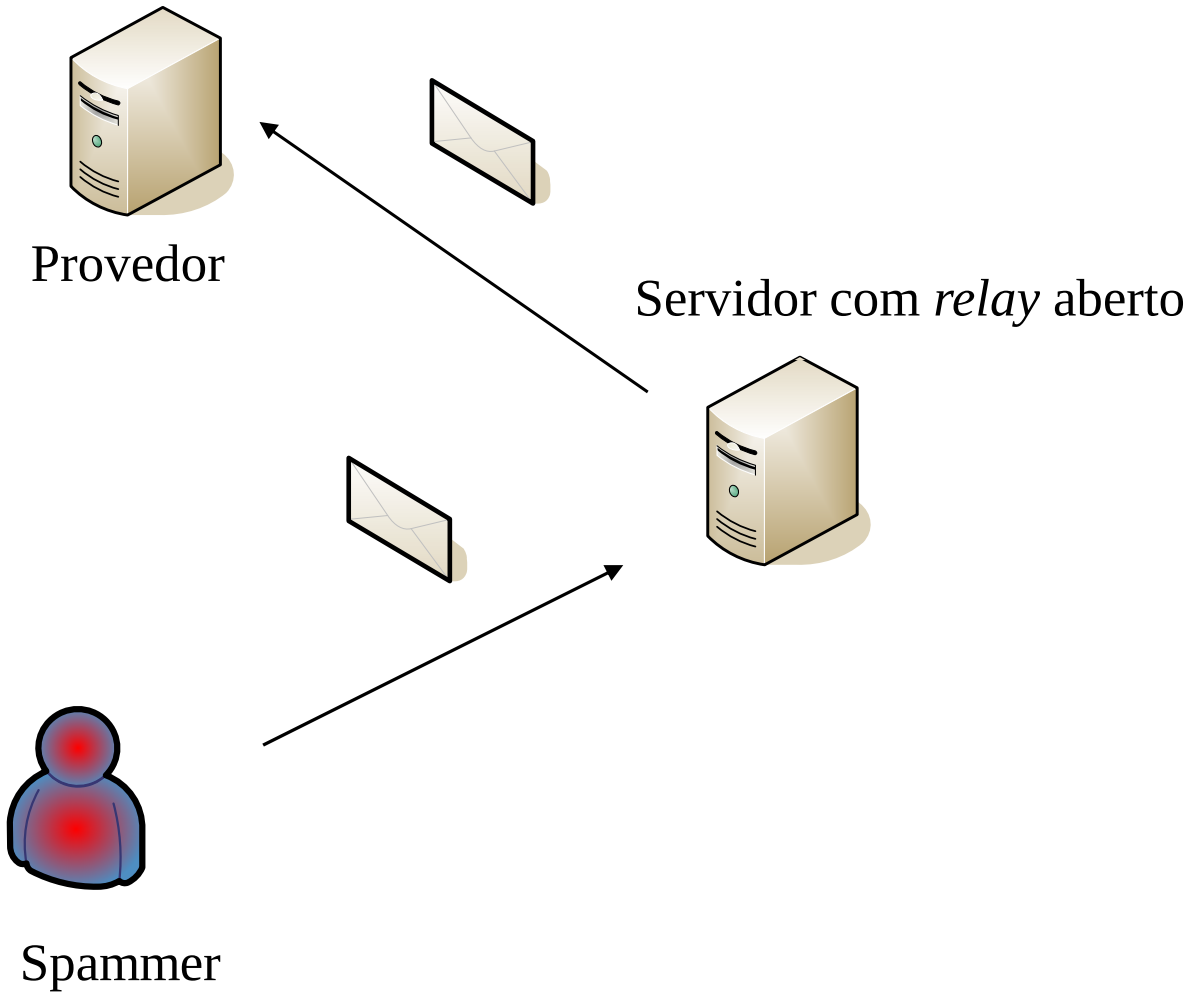
Envio das mensagens - *relay* aberto

- Servidor de correio configurado para encaminhar mensagens enviadas a ele de qualquer lugar, para qualquer receptor
  - Sem verificar o remetente

Exemplo de interação (telnet servidor.br 25)

```
S: 220 servidor.br
C: HELO cliente.br
S: 250 Hello cliente.br, pleased to meet you
C: MAIL FROM: <usuario@cliente.br>
S: 250 usuario@cliente.br... Sender ok
C: RCPT TO: <usuario@outroservidor.br>
S: 250 usuario@outroservidor.br ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: From: usuario@cliente.br
C: To: usuario@outroservidor.br
C: Subject: Teste
C:
C: Teste de envio de correio.
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 servidor.br closing connection
```

Envio das mensagens - *relay* aberto



Envio das mensagens - *relay* aberto

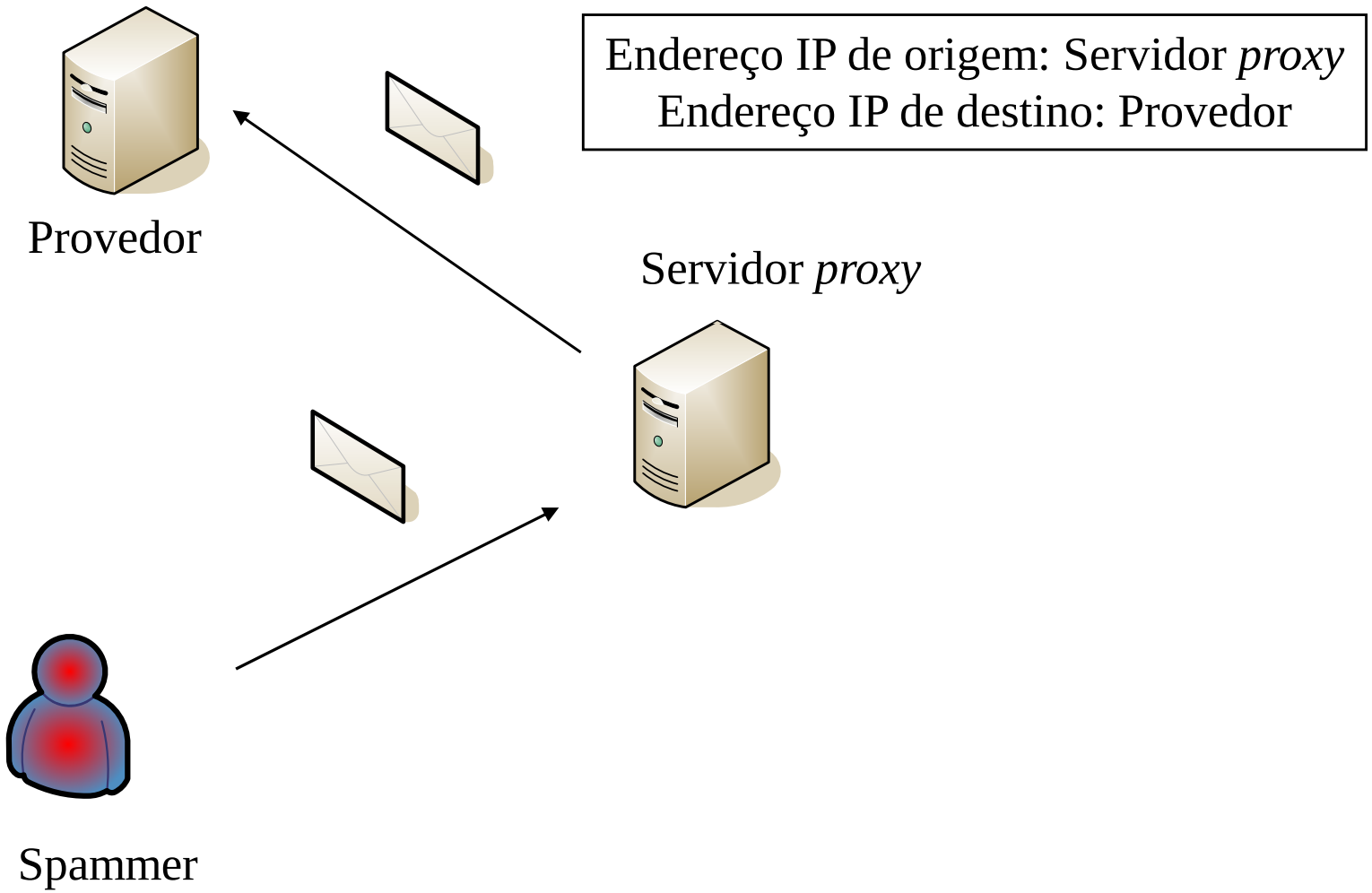
- Origem pode ser identificada ou não
  - Alguns *relays* "apagam" a origem

- SMTP original usava *relay* aberto
- Servidores atuais em geral não permitem o uso do *relay* aberto
  - Falhas de segurança em alguns programas às vezes permitem o seu uso

Envio das mensagens - *proxy* aberto

- Servidor dentro de uma rede usado como um *gateway* para a Internet
  - Protocolo SOCKS transfere dados de um cliente para um servidor
    - Transferência de modo transparente
    - Exemplos de protocolos: HTTP, FTP, SMTP
- Permite que qualquer "cliente" use o servidor para enviar mensagens
  - Por má configuração
  - Por invasão

Envio das mensagens - *proxy* aberto



Envio das mensagens - *proxy* aberto

- Servidor SMTP que detecta um grande número de mensagens vindo de um mesmo *proxy* pode
  - Rejeitar as mensagens
  - Notificar o endereço IP do *proxy* para listas negras
- *Spammers* podem usar uma série de artifícios
  - Grande número de *proxies*
  - Servidores em países com línguas diferentes
    - Dificultam a comunicação entre os responsáveis pelos servidores
- Servidores *proxy* são compartilhados entre vários *spammers*
  - Acelera a detecção e a inclusão em listas negras

Envio das mensagens - invasão

- Invasão de computadores
  - *Spammers* podem formar redes de máquinas zumbis
    - Coordenadas por máquinas mestras
  - *Spammers* sujeitos a severas punições
    - Apropriação de bens de terceiros
  - Maioria dos *spams* é enviada desta forma

- Seqüestro de interfaces CGI
  - Modificação de *scripts* CGIs

Envio das mensagens - invasão

- Injeção de rota BGP ou seqüestro de sistema autônomo através de
  - Seqüestro de faixas de endereços IP válidos
  - Invasão de roteador para ser responsável pela faixa
  - Escolhe um endereço IP da faixa
  - Quando um IP é adicionado a uma lista negra, basta usar outro da faixa
  - Técnica é uma das mais efetivas no envio de *spams*

Envio das mensagens - contas

- Contas em serviços gratuitos de correio
  - Várias contas são criadas por robôs
  - Alguns servidores de correio via *web* usam CAPTCHAs para tentar evitar a ploriferação dessas contas

Envio das mensagens - empresas

- Empresas especializadas
  - Voltadas para *spammers* sem conhecimento técnico
  - Podem usar servidores hospedados em países que não proíbem o envio de *spams*

Técnicas de Combate aos *Spams*

Mecanismos anti-*spam*

- Classificar mensagens como *spam*
  - Filtrar mensagens
- Dificultar o envio de *spams*
- Uso de vários mecanismos
  - Aumento da eficiência

Requisitos

- Adoção de um mecanismo
  - Precisão
  - Amigável para o usuário
  - Eficientes computacionalmente

Características

- Falsos positivos
  - Alto impacto para usuários
    - Atrasos no processo de comunicação
    - Perda de informações
- Falsos negativos
  - Menor impacto
    - Perda de tempo com *spams*
- Facilidade de operação/manutenção
  - Evolução dos *spams*
  - Interação do usuário com o sistema

Classificação

- Divisão em três classes
  - Filtragem simples
  - Auto-aprendizado
  - Verificação da origem

Filtragem simples

- Uso de regras, filtros ou listas
  - Análise baseada em informações já existentes
    - Sem aprendizado
- Manutenção periódica
  - Evolução dos *spams*
- Atualização manual
  - Grande intervenção do usuário
  - Dificuldade de manutenção

Auto-aprendizado

- Aprendizado automático
  - Adaptação à evolução dos *spams*
- Baixa intervenção do usuário

Verificação da origem

- Verificar autenticidade da origem
  - Autenticidade do endereço de correio eletrônico
    - Evitar fraudes
  - Verificação de mecanismos automatizados
    - Limitar o envio de *spams*

Sistemas Baseados em Filtragem Simples

Sistemas

- Listas negras
- Uso de pesos e regras
- Filtros Bayesianos

Listas negras

- Lista negra => Mensagem descartada
  - Máquinas que enviam *spam*
    - Endereço IP
  - Endereços eletrônicos usados por *spammers*
    - Baixa eficiência
      - Falsificação do endereço de origem
- Lista branca => Mensagem encaminhada
  - Endereços reconhecidamente legítimos

Listas negras

- Ações tomadas pelo mecanismo
  - Pertinência da mensagem nas duas listas
    - Lista negra
      - Mensagem descartada
    - Lista branca
      - Mensagem encaminhada
    - Nenhuma das duas
      - Análise por outros mecanismos disponíveis
- Baixo custo computacional
  - Busca em uma lista

Listas negras

- Listas pessoais
- Listas coletivas
  - Entidade responsável pelas listas
- Manutenção das listas
  - Manual
  - Mecanismos automatizados
    - Busca e adição de servidores *relay* aberto
  - Sistemas com auto-aprendizado
    - Aproveitamento do baixo custo computacional das listas

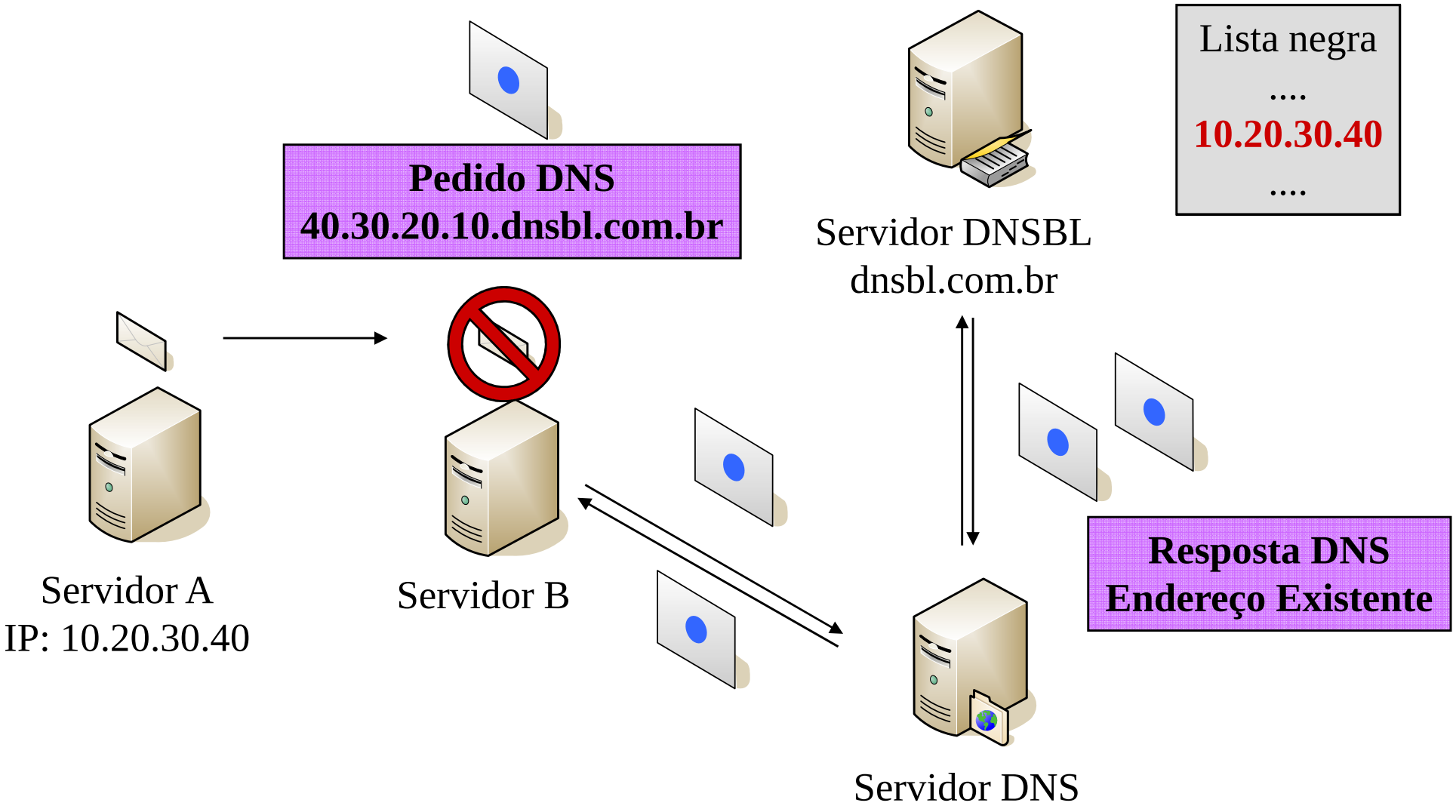
Listas negras

- Listas DNSBL (*Domain Name System Black List*)
  - Listas coletivas
    - Entidade ou grupo responsável pela lista
  - Consulta através do protocolo DNS
    - Facilidade de implementação
    - Utilização de protocolo maduro
    - Aproveitamento de funções já implementadas
      - Cache
      - Distribuição da base de dados

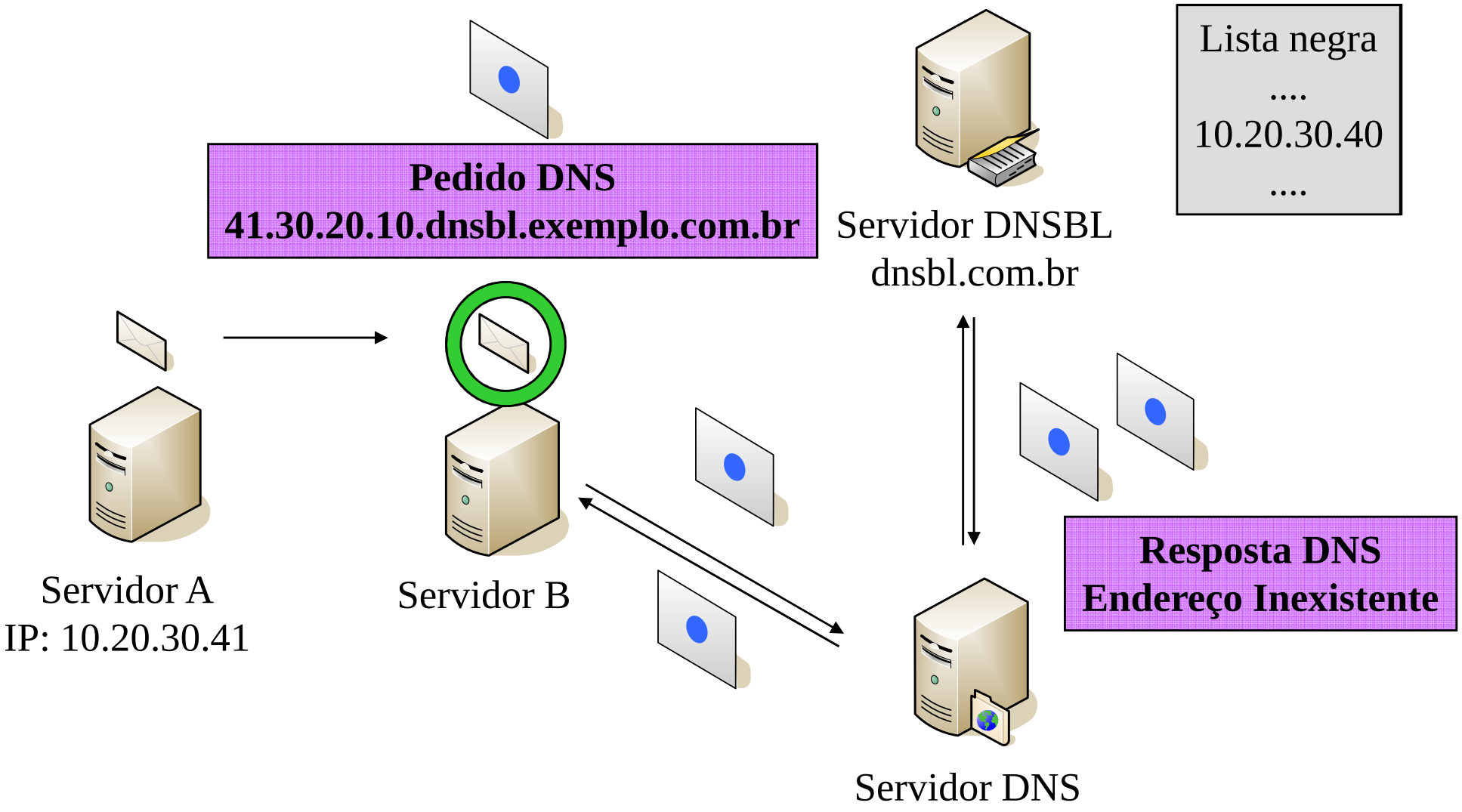
Listas negras

- Consulta à lista DNSBL
  - Verificação do endereço IP do servidor remetente
  - Criação de um nome fictício
    - Inversão do IP do servidor remetente
    - Agregação do domínio do servidor da lista DNSBL
  - Consulta DNS do nome fictício criado
  - Resposta positiva => Endereço na lista negra

Lista negras



Lista negras



Listas negras

- Gerenciamento das listas DNSBL
  - Entidade responsável pela lista
  - Retirada manual da lista
    - Máquinas zumbis
    - Máquinas mal configuradas
  - Falsos positivos
    - Atingem todos os usuários do servidor

Listas negras

- Exemplos de listas DNSBL
- Spamhaus - http://www.spamhaus.org
  - Duas listas negras
    - SBL => Servidores que enviam spam
    - XBL => Servidores que enviam vírus/cavalos de tróia
- SORBS - http://www.us.sorbs.net/
  - Lista endereços dinâmicos

Uso de pesos e regras

- Regras
  - Testes lógicos
    - Características de *spams*
  - Pesos
    - Positivos ou negativos
    - Probabilidade de ser *spam*
  - Resultado positivo ou negativo
- Verificação de todas as regras
  - Regras com resultado positivo
    - Acumulo dos pesos
  - Probabilidade de ser *spam* => Somatório dos pesos

Uso de pesos e regras

- Construção das regras
  - Manualmente
  - Características presentes em *spams*
    - Palavras
    - Frases
    - Presença de figuras
    - *Tags* html
  - Generalidade x Especificidade
    - Compromisso
    - Falsos positivos x Negativos
    - Adaptação a evoluções

Uso de pesos e regras

- Avaliação das regras
  - Base de mensagens spams e legítimas
  - Falsos positivos e negativos
    - Altos => Regra é alterada
  - Taxa de acerto para spams
    - Alta => Peso maior
  - Taxa de acerto para mensagens legítimas
    - Alta => Peso negativo
  - Combinação das taxas => Peso final
  - Constantemente reavaliadas
    - Evolução dos spams
    - Perda de eficiência

Uso de pesos e regras

- Manutenção das regras
  - Manual
  - Individual
    - Regras personalizadas
    - Desconhecimento dos usuários
    - Alta intervenção do usuário
    - Específicas para mensagens do usuário
  - Coletivas
    - Grupo ou entidade responsável
    - Regras gerais
      - Alteração manual de alguma regras
    - *Spammer* conhece as regras
      - Verificação das regras antes do envio

Uso de pesos e regras

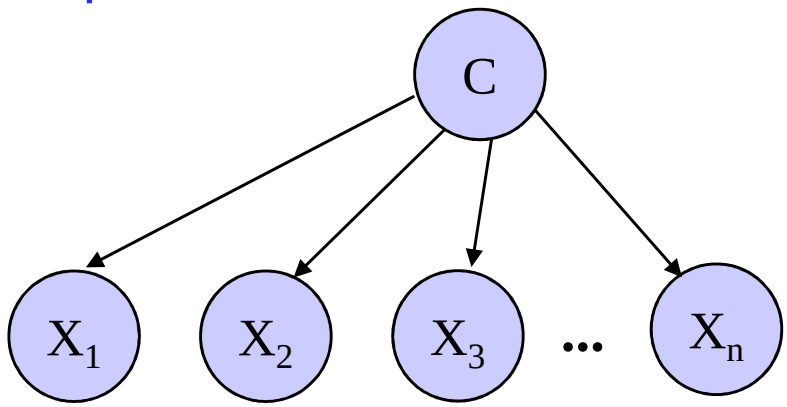
- Falsos positivos e negativos
  - Regras mal adaptadas
- Impacto para o usuário
  - Alteração de regras gerais
  - Criação de regras específicas
  - Alta interação com o sistema

Filtros bayesianos

- Evolução do sistema de pesos e regras
- Ocorrência de frases/construções em *spam*
  - Características específicas
    - Reconhecimento das características
- Determinação estatística de características de *spam*
  - Classificação manual das mensagens
    - Intervenção do usuário
  - Classificadores Bayesianos
    - Redes bayesianas
  - Personalizado
    - Mensagens e *spams* do usuário

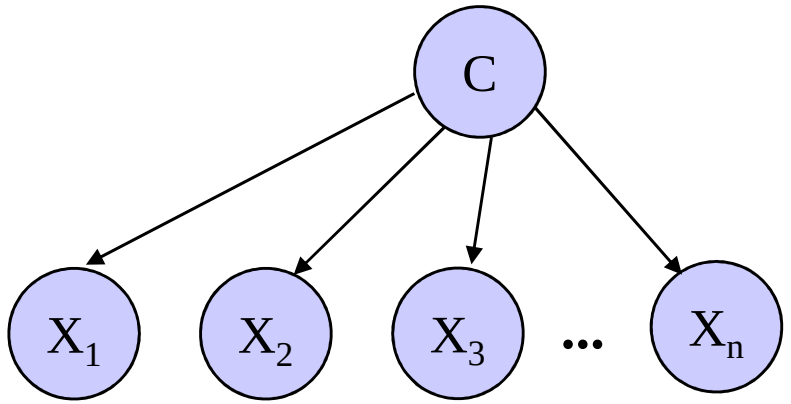
Filtros bayesianos

- Redes bayesianas
  - Grafo acíclico direcionado
  - Representa distribuição de probabilidade
  - Nó
    - Variável aleatória Xi
    - Distribuição de Xi dada através dos pais
  - Arestas
    - Influencia do pai sobre o filho



Filtros bayesianos

- Classificador bayesiano
  - Rede bayesiana
  - Nó Classificador
    - Possíveis classificações
      - Spam, não spam
  - Nós Xi
    - Características testadas



- Probabilidade de classificação em uma classe c_k
  - Spam
  - Não spam

$$P(C = c_k \mid \mathbf{X} = \mathbf{x}) = \frac{P(\mathbf{X} = \mathbf{x} \mid C = c_k) P(C = c_k)}{P(\mathbf{X} = \mathbf{x})}$$

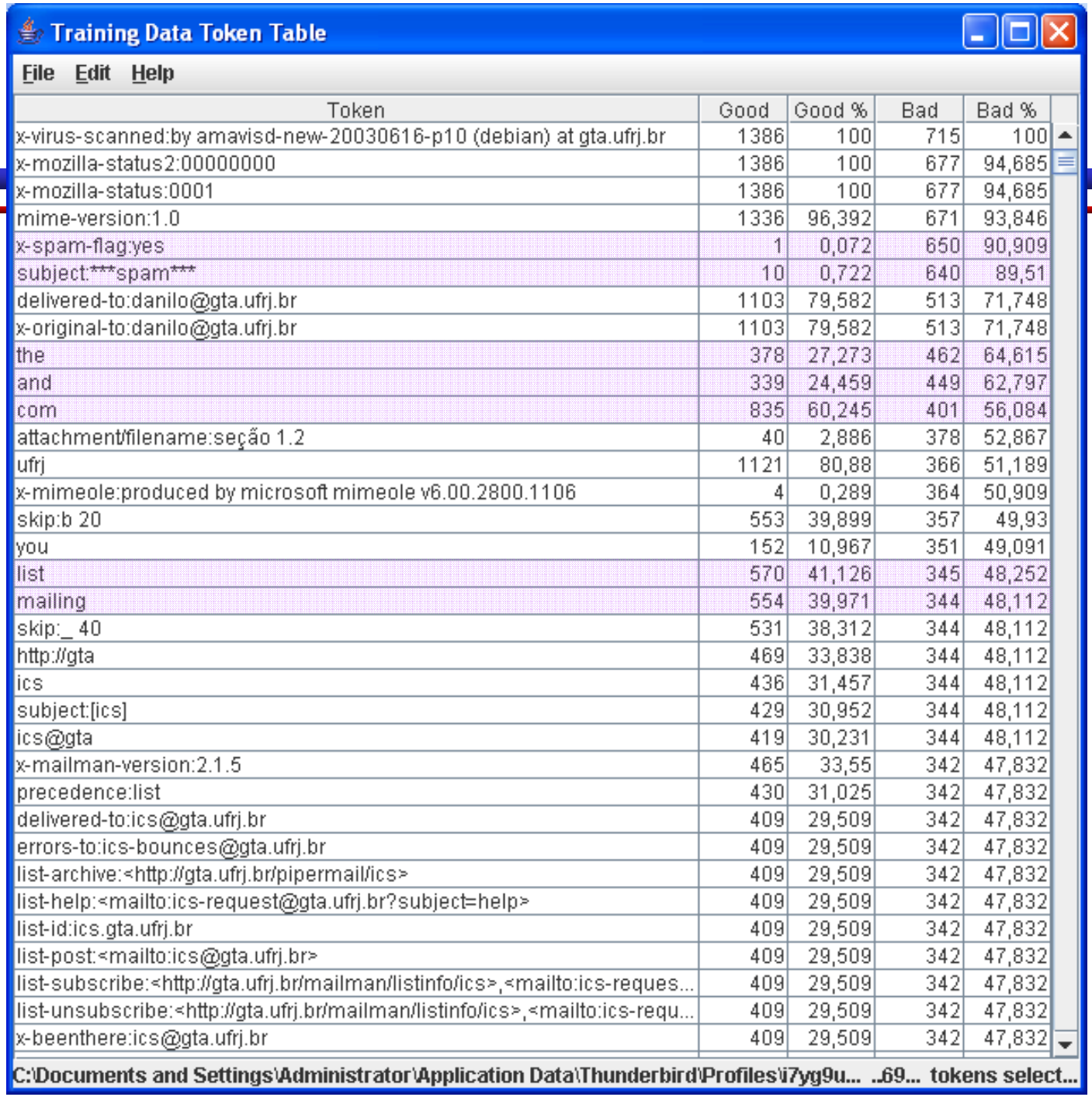
$$P(\mathbf{X} = \mathbf{x} \mid C = c_k) = \prod_i P(X_i = x_i \mid C = c_k)$$

Filtros bayesianos

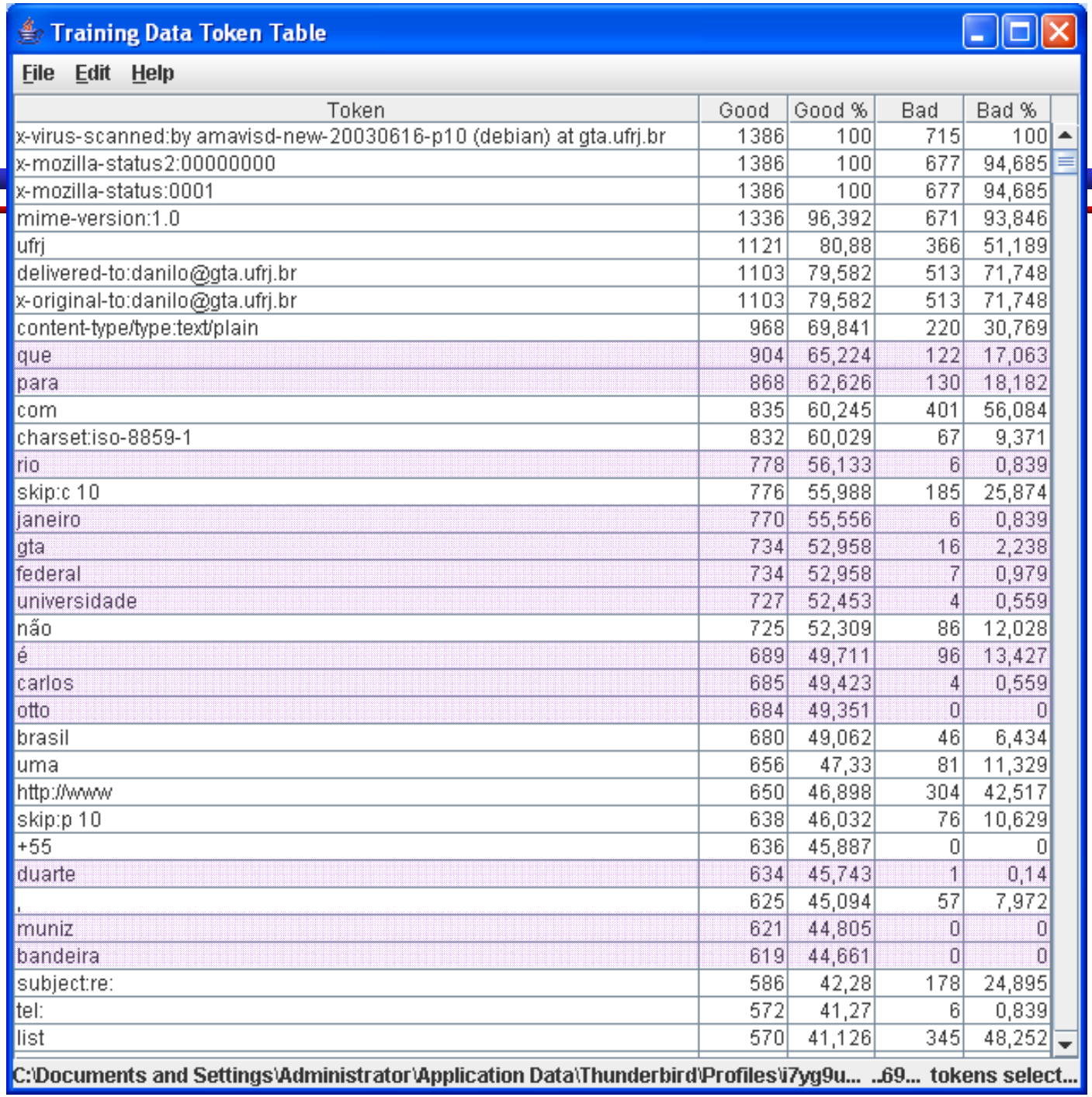
- Representação em vetores de características
  - Separação da mensagem em símbolos
    - Exemplo: Espaços e símbolos de pontuação
- Propriedades específicas
  - Frases com muitas exclamações
  - Domínio do remetente
  - Presença de anexos

Filtros bayesianos

- Grupos separados de spams e mensagens legítimas
  - Atualização do filtro
    - Novas características
    - Alteração de probabilidades
- Análise das características
  - Pertinência em mensagens spams e legítimas
  - Probabilidade de ser característica de *spam*
- Redução do número de características
  - Eliminação de palavras pouco freqüentes
  - Seleção das mais importantes



Token	Good	Good %	Bad	Bad %
x-virus-scanned:by amavisd-new-20030616-p10 (debian) at gta.ufrj.br	1386	100	715	100
x-mozilla-status2:00000000	1386	100	677	94,685
x-mozilla-status:0001	1386	100	677	94,685
mime-version:1.0	1336	96,392	671	93,846
x-spam-flag:yes	1	0,072	650	90,909
subject:***spam***	10	0,722	640	89,51
delivered-to:danilo@gta.ufrj.br	1103	79,582	513	71,748
x-original-to:danilo@gta.ufrj.br	1103	79,582	513	71,748
the	378	27,273	462	64,615
and	339	24,459	449	62,797
com	835	60,245	401	56,084
attachment/filename:seção 1.2	40	2,886	378	52,867
ufrj	1121	80,88	366	51,189
x-mimeole:produced by microsoft mimeole v6.00.2800.1106	4	0,289	364	50,909
skip:b 20	553	39,899	357	49,93
you	152	10,967	351	49,091
list	570	41,126	345	48,252
mailing	554	39,971	344	48,112
skip:_ 40	531	38,312	344	48,112
http://gta	469	33,838	344	48,112
ics	436	31,457	344	48,112
subject:[ics]	429	30,952	344	48,112
ics@gta	419	30,231	344	48,112
x-mailman-version:2.1.5	465	33,55	342	47,832
precedence:list	430	31,025	342	47,832
delivered-to:ics@gta.ufrj.br	409	29,509	342	47,832
errors-to:ics-bounces@gta.ufrj.br	409	29,509	342	47,832
list-archive:<http://gta.ufrj.br/pipermail/ics>	409	29,509	342	47,832
list-help:<mailto:ics-request@gta.ufrj.br?subject=help>	409	29,509	342	47,832
list-id:ics.gta.ufrj.br	409	29,509	342	47,832
list-post:<mailto:ics@gta.ufrj.br>	409	29,509	342	47,832
list-subscribe:<http://gta.ufrj.br/mailman/listinfo/ics>,<mailto:ics-reques...	409	29,509	342	47,832
list-unsubscribe:<http://gta.ufrj.br/mailman/listinfo/ics>,<mailto:ics-requ...	409	29,509	342	47,832
x-beenthere:ics@gta.ufrj.br	409	29,509	342	47,832



Token	Good	Good %	Bad	Bad %
x-virus-scanned:by amavisd-new-20030616-p10 (debian) at gta.ufrj.br	1386	100	715	100
x-mozilla-status2:00000000	1386	100	677	94,685
x-mozilla-status:0001	1386	100	677	94,685
mime-version:1.0	1336	96,392	671	93,846
ufrj	1121	80,88	366	51,189
delivered-to:danilo@gta.ufrj.br	1103	79,582	513	71,748
x-original-to:danilo@gta.ufrj.br	1103	79,582	513	71,748
content-type/type:text/plain	968	69,841	220	30,769
que	904	65,224	122	17,063
para	868	62,626	130	18,182
com	835	60,245	401	56,084
charset:iso-8859-1	832	60,029	67	9,371
rio	778	56,133	6	0,839
skip:c 10	776	55,988	185	25,874
janeiro	770	55,556	6	0,839
gta	734	52,958	16	2,238
federal	734	52,958	7	0,979
universidade	727	52,453	4	0,559
não	725	52,309	86	12,028
é	689	49,711	96	13,427
carlos	685	49,423	4	0,559
otto	684	49,351	0	0
brasil	680	49,062	46	6,434
uma	656	47,33	81	11,329
http://www	650	46,898	304	42,517
skip:p 10	638	46,032	76	10,629
+55	636	45,887	0	0
duarte	634	45,743	1	0,14
,	625	45,094	57	7,972
muniz	621	44,805	0	0
bandeira	619	44,661	0	0
subject:re:	586	42,28	178	24,895
tel:	572	41,27	6	0,839
list	570	41,126	345	48,252

Filtros bayesianos

- Ataques ao mecanismo
  - Induzir símbolos
    - Símbolos que não representam spams
    - *C/A/L/L/ N-O-W - I/T/S F_R_E_E,*
      - Símbolos: C,A,L,L,N,O,W,I,T,S,F,R,E,E
    - Agregação de símbolos
      - Janela de agregação
    - Similaridade entre símbolos
  - Palavras inocentes
    - *Snowflaking messages*
    - Adição de palavras ou textos aleatórios
    - Palavras encontradas em mensagens legítimas
      - Mais características de mensagens legítimas

Sistemas com Auto-Aprendizado

Sistemas

- Caracterização de Tráfego de *Spams*
- Listas Cinzas
- Potes de Mel
- Redes Sociais

Caracterização de tráfego

- Monitorar tráfego de mensagens
  - Limitar usuários enviando muitas mensagens
  - Realizado por cada servidor
    - *Spammer* utilizar servidor próprio
  - Dificuldade de identificação
    - Usuário legítimo pode enviar muitas mensagens

Caracterização de tráfego

- Entender a característica do tráfego de spam
  - Característica e anomalias no tráfego
  - Separação em grupos
    - Spam
    - Não spam
- Grandes diferenças entre os dois grupos
  - Mensagens não spam são parte de uma relação bilateral derivada de uma relação social
  - Envio do maior número de mensagens possíveis

Caracterização de tráfego

- Diferença na distribuição temporal
  - Spam
    - Distribuição constante
  - Mensagens legítimas
    - Picos nos horário comerciais
    - Baixa atividade em finais de semana e madrugadas
- Intervalo entre mensagens
  - Muito pequeno para *spams*

Caracterização de tráfego

- Tamanho das mensagens
  - Mensagens não spam são de 6 a 8 vezes maiores
  - 1% dos spams tem mais do que 60Kb
- Spams geralmente tem mais remetentes
  - Multiplicação da banda para o *spammer*
    - Servidor reenvia para os destinatários

Caracterização de tráfego

- Altas taxas de falsos positivos e negativos
  - Tráfego muito variável
  - Pessoal
    - Aprendizado dos padrões por usuário
  - Utilização em conjunto com outros mecanismos
- Não requer intervenção do usuário para aprender

Listas cinzas

- Lista Cinza
  - Lista intermediária
  - Quarentena dos remetentes
    - Bloqueio temporário no envio de mensagens
- Utilização em conjunto de listas brancas
  - Gerenciadas automaticamente

Listas cinzas

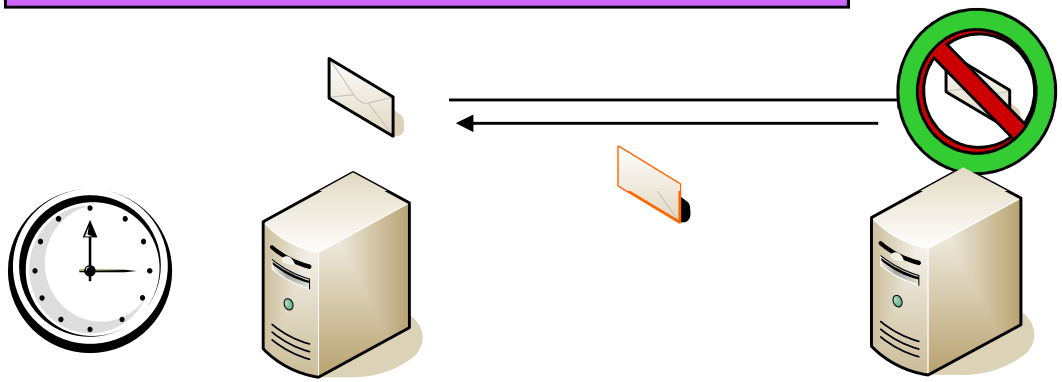
- Princípios explorados
  - Erro temporário no envio da mensagem
    - *Spammers* geralmente não reenviam
      - Aumentar o número de mensagens enviadas
    - Norma define reenvio em caso de erro

Listas cinzas

- Funcionamento
  - Verificação da lista branca
  - Primeira comunicação entre remetente/destinatário
    - Adição do par na lista cinza
      - Tempo de quarentena
      - Tempo de expiração
      - Servidor de origem
    - Mensagem descartada
      - Erro temporário informado
  - Reenvio da mensagem
    - Verificação do tempo de quarentena
    - Adição na lista branca

Listas cinzas

Remetente: joao@exemplo.com.br
Destinatário: maria@exemplo.com.br



Servidor A Servidor B

Lista Branca

Remetente	Destinatário	Servidor
jose	**maria**	**A**
andre	**maria**	**A**
adriana	**maria**	**A**
joao	**maria**	**A**

Lista Cinza

Remetente	Destinatário	Servidor

Listas cinzas

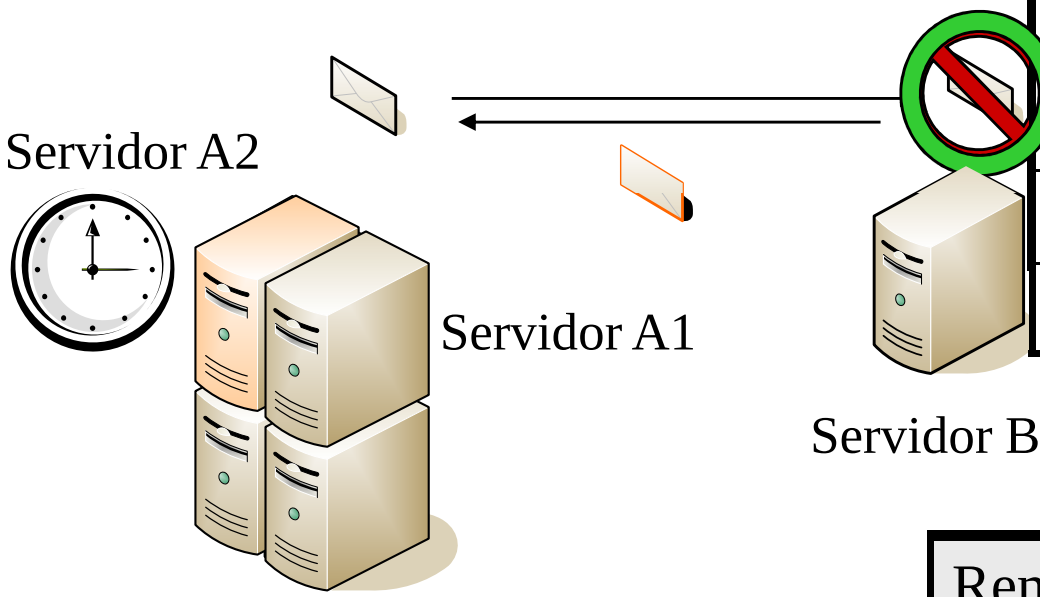
- Falsos positivos
  - Servidores que não reenviam
  - Reenvio antes do período de quarentena
- Falsos negativos
  - *Spammers* podem reenviar mensagens
- Impacto para usuários
  - Atraso no processo de comunicação
    - Confirmações por correio eletrônico

Listas cinzas

- Múltiplos servidores de envio de mensagens
  - Atrasos ainda maiores
  - Lista Cinza => Remetente/Destinatário/Servidor
  - Tentativas por servidores diferentes
    - Nova entrada na lista cinza

Listas cinzas

Remetente: joao@exemplo.com.br
Destinatário: maria@exemplo.com.br



Lista Branca

Remetente	Destinatário	Servidor
jose	**maria**	**A1**
andre	**maria**	**A1**
adriana	**maria**	**A1**
joao	**maria**	**A1**

Lista Cinza

Remetente	Destinatário	Servidor
joao	**maria**	**A2**

Potes de mel

- Não realizam a classificação de mensagens
- Ajudam no processo de aprendizagem
  - Endereços de correio eletrônico especiais
    - Iscas para *spammers*
    - Divulgados propositalmente
    - Não utilizados por usuários legítimos
    - Recebem grande quantidade de *spam*
      - Todas as mensagens são tratadas como *spam*
      - Processo de aprendizado de outros mecanismos

Potes de mel

- Servidores de internet
  - Gera páginas aleatórias
  - Insere endereços de correio eletrônico
    - Recebem *spam*
    - Diminuem a eficiência das listas *dos spammers*
      - Mensagens que não serão lidas
    - Endereços na mesma cor do fundo da página
      - Coleta por mecanismos automatizados

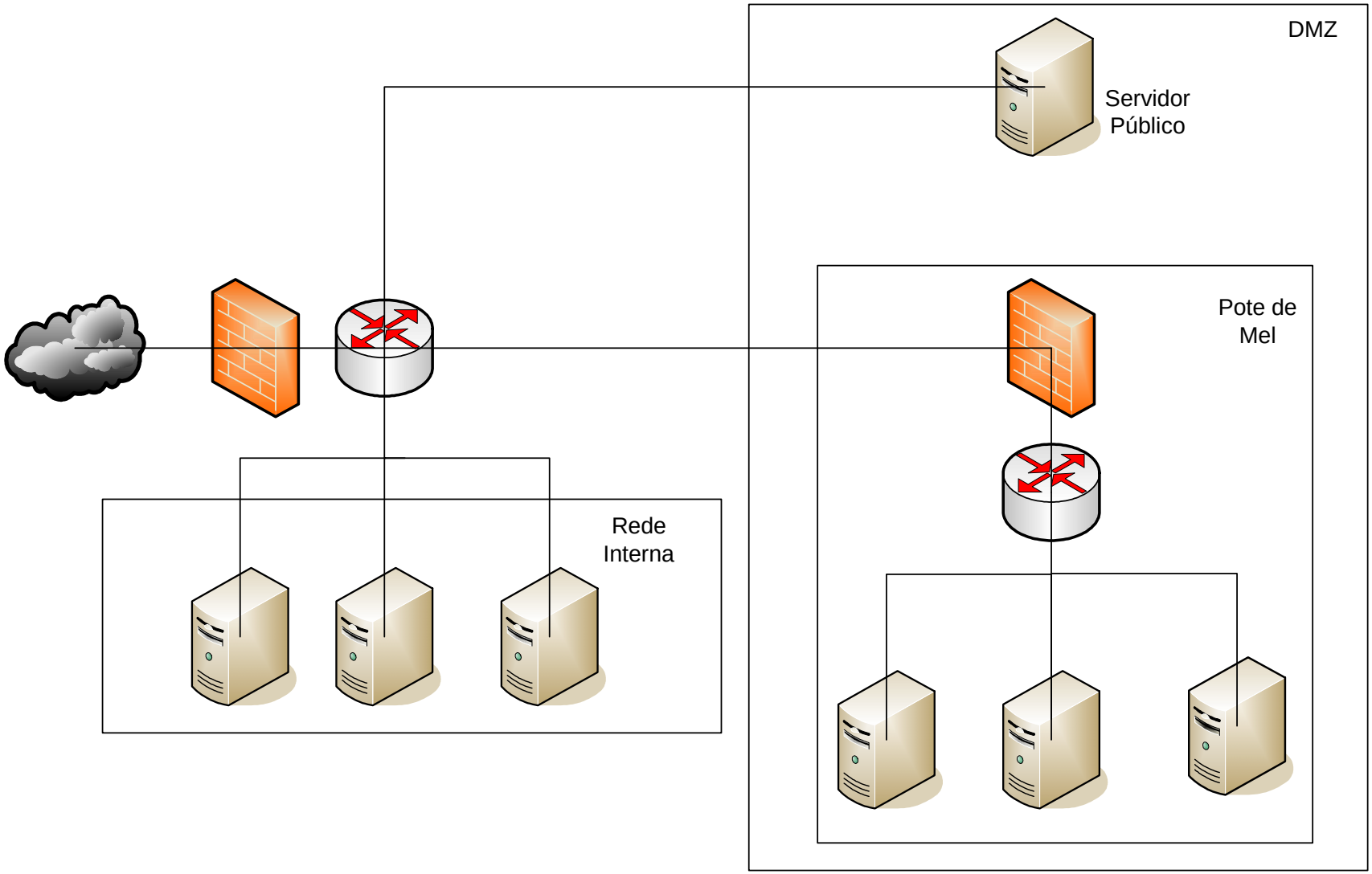
Potes de mel

- Servidores de envio de mensagens
  - Se passam por servidores de *relay* aberto
  - Não enviam as mensagens
    - Não contribuem com os *spammers*
    - Informam que mensagem foi enviada com sucesso
  - Descoberta de *spammers*
  - Diminuem *spams* enviados
- Servidores para receber mensagens
  - Análise dos spams recebidos

Potes de mel

- Arquitetura
  - Um ou mais componentes
  - Máquinas virtuais
    - Simulam vários servidores
  - Isolamento do pote de mel
    - Segurança da rede
    - Não permitir conexões de saída
      - Não enviar *spams*
    - Isolados na Zona Desmilitarizada(*DMZ - Demilitarized Zone*)

Potes de mel

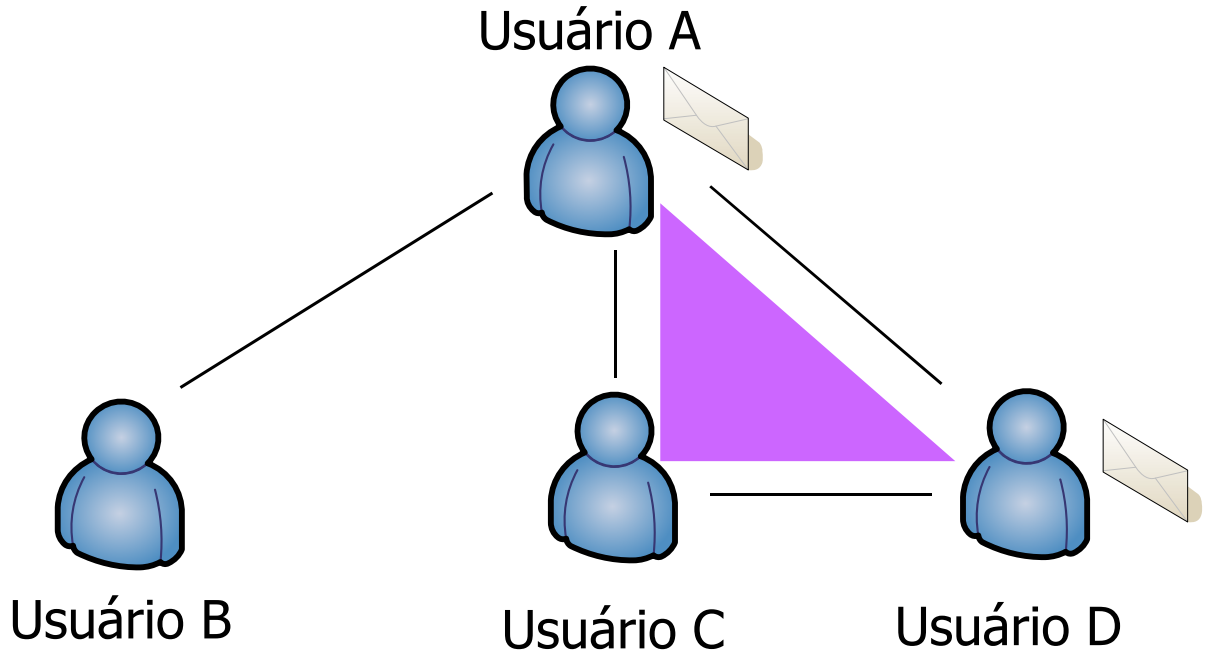


Redes sociais

- Usa informações de redes sociais
- Baseia-se na eficácias dessas redes para tomar decisões
- Relações sociais entre pessoas
  - Representadas por grafos
- Instinto natural de formação de comunidades
  - Pessoas que se conhecem não enviam *spams*
- Criação de grupos de comunidades
  - Marcelo conhece Luis e Carlos
  - Luis provavelmente conhece Carlos

Redes sociais

- Construção das redes
  - Baseada nas mensagens recebidos por uma pessoa
  - Informação de destinatário da mensagem
    - Campos Para e Com Cópia
  - Cada nó representa um endereço
  - Ligações entre nós que se comunicaram através de outro

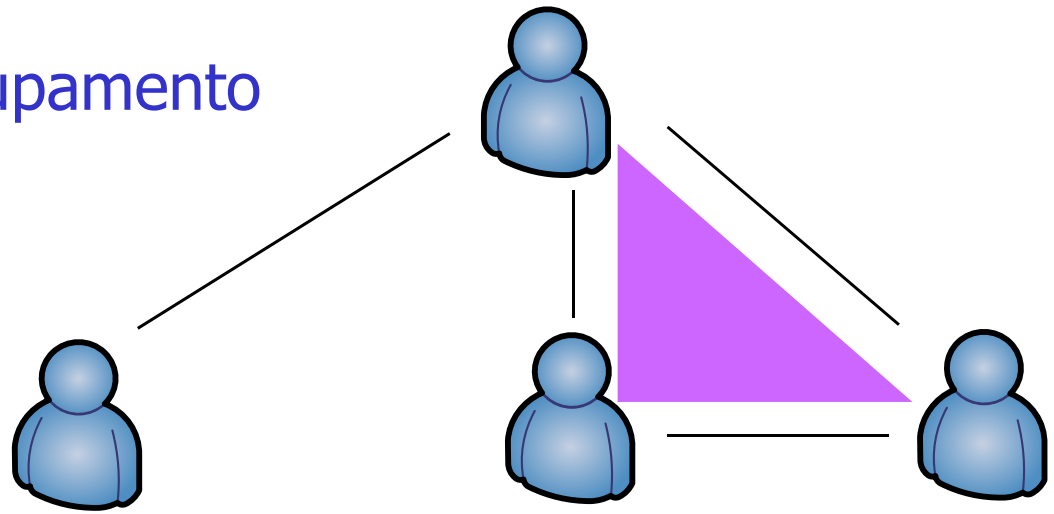


Redes sociais

- Separação do grafo em componentes
  - Subgrafos que não tem vértices entre si
- Tipo de rede
  - Por usuário
    - Nó representa usuário
  - Por domínio
    - Nó representa domínios externos
      - Agrupamento de usuários
    - *Spammers* geralmente forjam o nome mas não o domínio
- Classificações
  - Social
    - Usuários legítimos
  - Anti-social
    - *Spammers*

Redes sociais

- Aspectos quantitativos podem diferenciar redes sociais
  - Tendência a formação de agrupamentos
    - Triângulos no grafo
- Definição qualitativa
  - T=Número de triângulos
  - W=Número de pares de nós ligados diretamente a um terceiro
  - C=Coeficiente de agrupamento
  - C=3T/W



Redes sociais

- Definição quantitativa
  - Fração dos vizinhos do nó que são vizinhos entre si
  - Nó com grau k_i tem k_i vizinhos
  - Possibilidade de $k_i(k_i-1)$ conexões entre vizinhos
  - Número de conexões => E_i
  - C=Coeficiente de agrupamento

$$C = \frac{1}{N_2} \sum_i \frac{2E_i}{k_i(k_i - 1)}$$

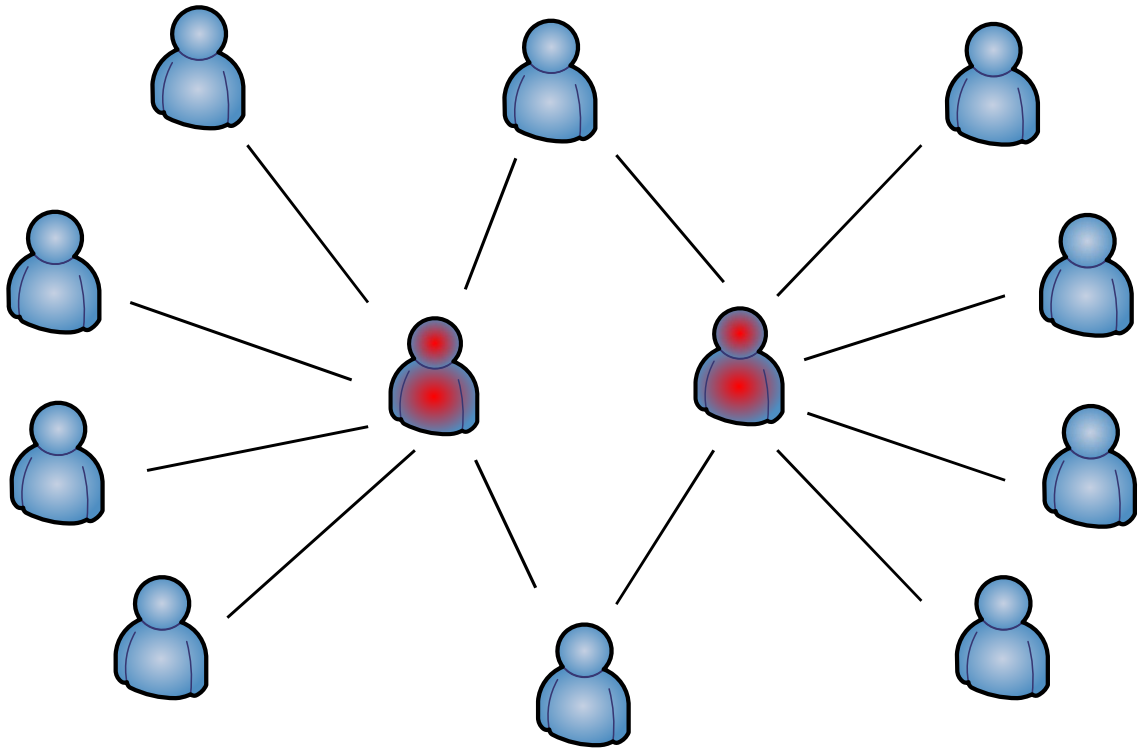
  - Coeficiente de agrupamento cerca de 10 vezes maior que grafos aleatórios

Redes sociais

- Redes sociais tem alto grau de agrupamento
- Análise do grau de agrupamento
  - Alto
    - Rede Social
    - Relacionamento entre pessoas
    - Listas brancas
      - Reduz falsos positivos
  - Baixo
    - Rede Anti-social
    - Não existe relação bilateral entre *spammer* e destinatários
    - Listas negras

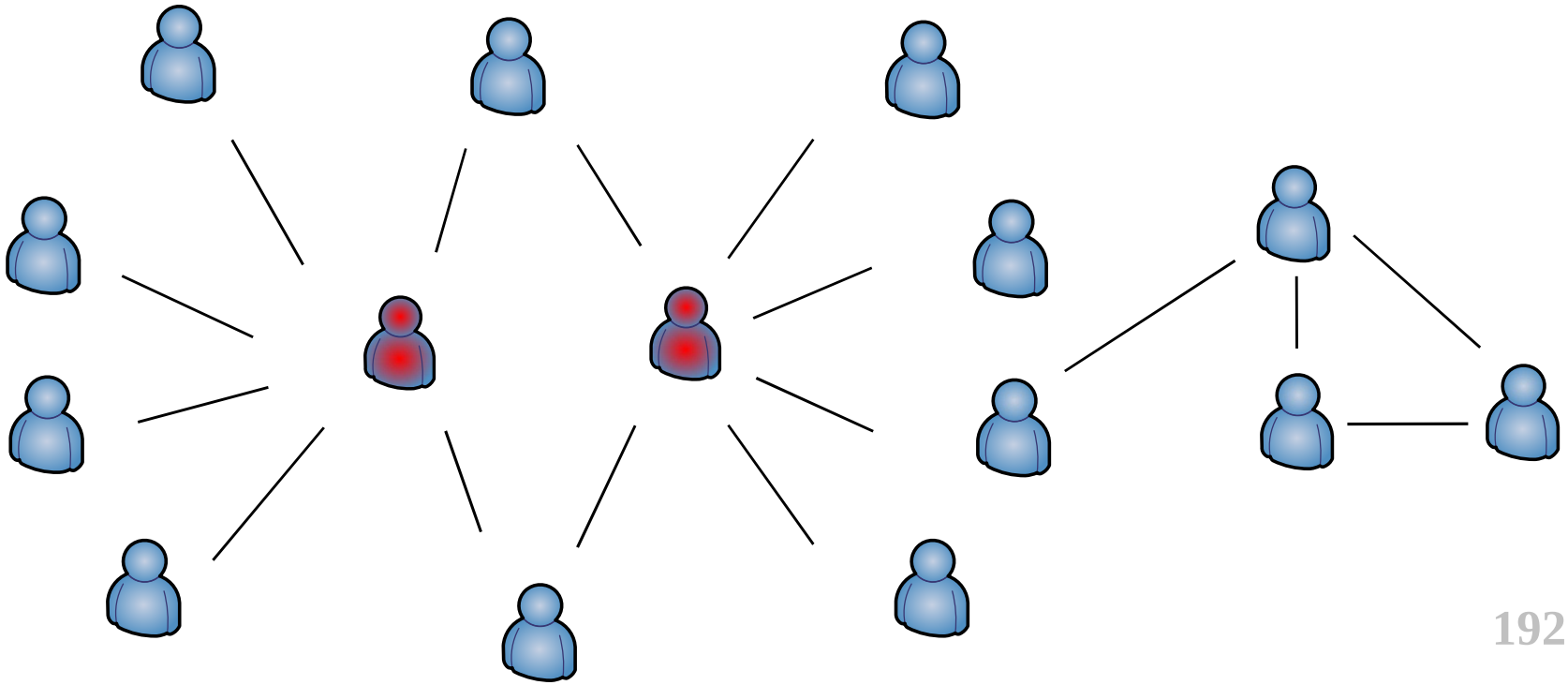
Redes sociais

- Componente Anti-Social
  - Podem ser geradas por ataques de dicionário
  - Compartilhamento de listas
  - Spammers e destinatários do spam



Redes sociais

- União de componentes
  - Spammer pode enviar mensagem para amigos da pessoa
    - Ataques de dicionário
    - União de grandes componentes de spam com componentes legítimas



Redes Sociais

- União de componentes
  - Duas componentes conectadas por poucas ligações e diferentes coeficientes de agregação
    - Aumenta o grau de agregação da nova componente
  - *Edge betweenness*
    - Usado para identificar caminhos que unem duas componentes
    - Separa as duas componentes

Redes sociais

- Falsos positivos
  - Reduzidos através das listas brancas
- Falsos negativos
  - SPAM nunca usar mesmo remetente
    - Formação de componentes isoladas
    - Não gera a lista negra
    - Gera a lista branca => Reduz falsos positivos
  - Formação de redes sociais artificiais
    - *Spammers* imitarem relação social
  - Enviar SPAM como usuário da lista branca
- Não requer intervenção do usuário

Sistemas Baseados na Verificação da Origem

Sistemas

- Verificação do endereço DNS reverso
- *Sender Policy Framework* (SPF)
- Desafio e resposta

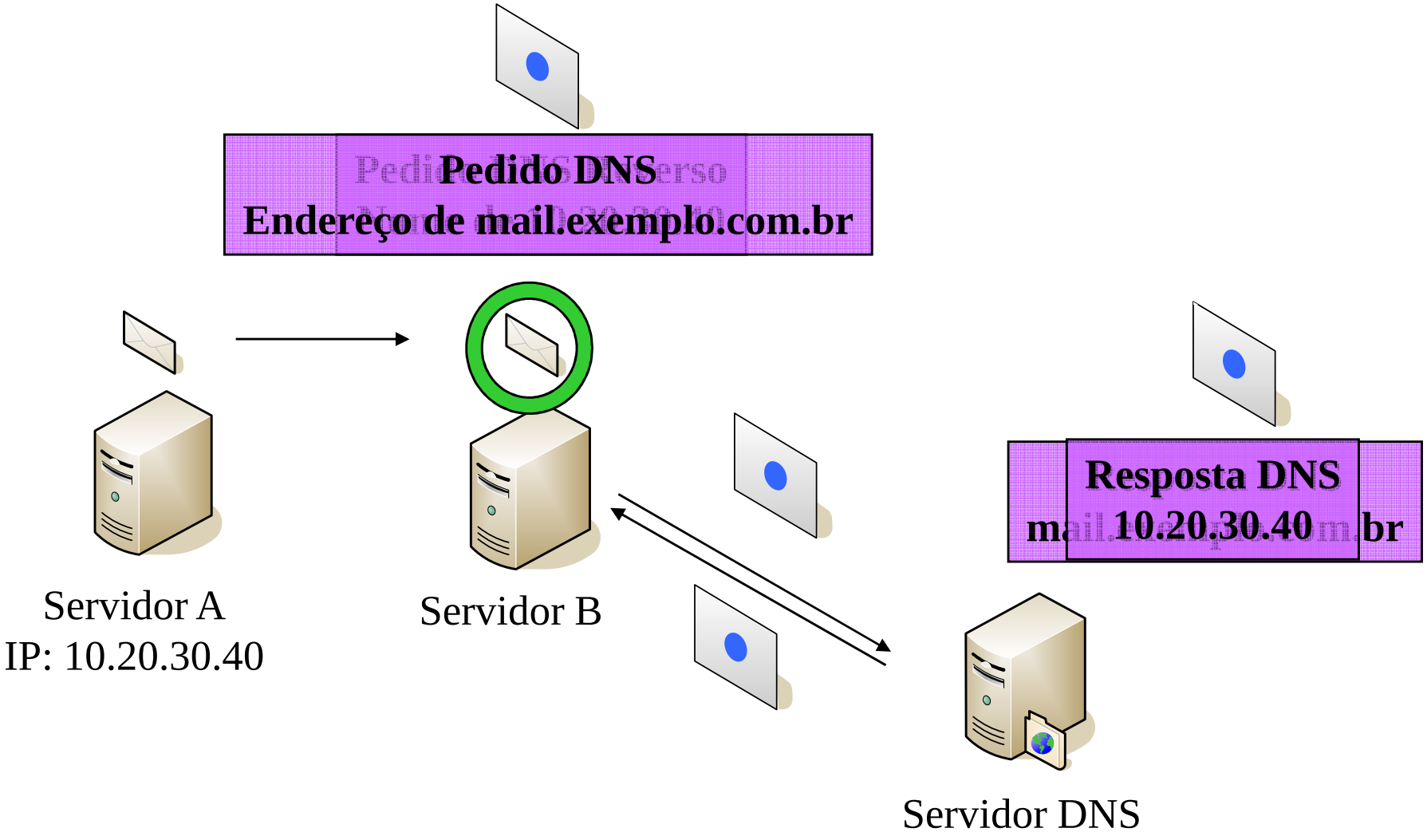
DNS reverso

- Verificação da identidade da origem
  - Dificultar a falsificação do remetente
- Utiliza informações do sistema de DNS
- Consulta do DNS reverso
  - Nome associado ao endereço IP
  - Endereço IP associado ao mesmo nome
  - Nome parte do domínio do remetente

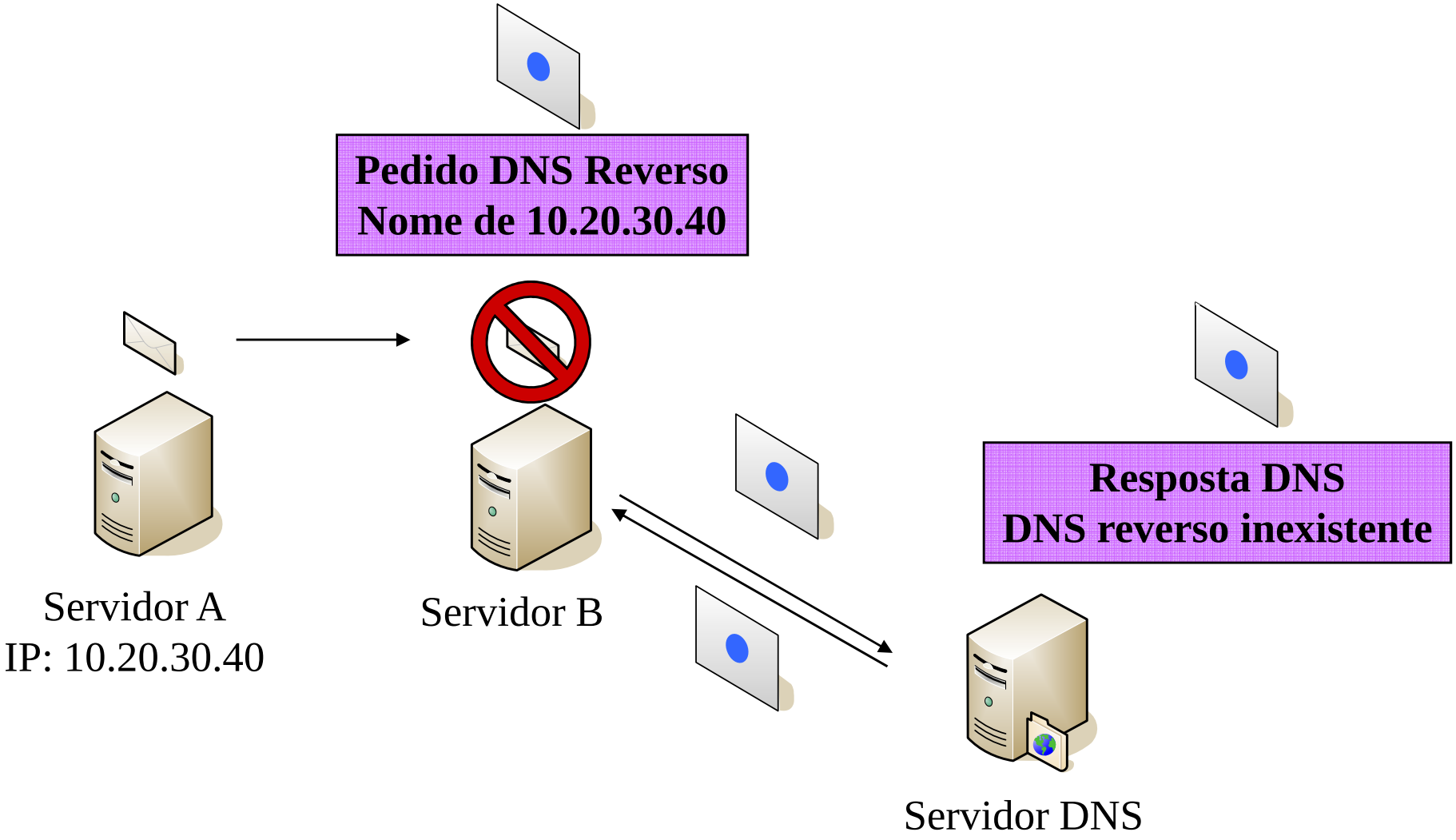
DNS reverso

- Verificação do DNS reverso
  - DNS reverso permite rastrear máquinas
    - Provedor associado à maquina
  - *Spammers* não configuram DNS reverso

DNS reverso


Pedido DNS
Endereço de mail.exemplo.com.br
Resposta DNS
10.20.30.40
Servidor A
IP: 10.20.30.40
Servidor B
Servidor DNS

DNS reverso


Pedido DNS Reverso
Nome de 10.20.30.40
Resposta DNS
DNS reverso inexistente
Servidor A
IP: 10.20.30.40
Servidor B
Servidor DNS

DNS reverso

- Falsos positivos
  - Servidores mau configurados
  - Mensagens do servidor descartadas
    - Alta taxa de falsos positivos
- Falsos negativos
  - Servidores legítimos invadidos

SPF

- *Sender Policy Framework* (SPF)
  - Dificultar falsificação da origem
  - Define políticas para o envio de mensagens
    - Domínio determina políticas
    - Verificadas no recebimento de mensagens
    - Registro SPF
      - Testes realizados
      - Publicado através do DNS
  - Autentica o domínio da origem
  - Não autentica usuário do domínio
    - Domínio autentica seus clientes

SPF

- Registro SPF
  - Registro DNS do tipo TXT
  - Versão do registro
    - Versão 1 => v=spf1
  - Mecanismos
    - Testes a serem realizados
    - Modificadores
      - Ação caso teste seja positivo
    - Avaliados da esquerda para direita

SPF

- Mecanismos
  - all
    - Retorna sempre resultado positivo
    - Definir ação padrão
  - include
    - Inclui registros SPF externos
      - Envio de mensagens através de outros domínios
    - include:<nome do domnio>
      - Exemplo: include:exemplo.com.br
  - a
    - IP corresponde ao endereço do nome do domínio

SPF

- Mecanismos
  - mx
    - Servidor corresponde a servidores de mensagens do domínio
  - ptr
    - Verifica DNS reverso
  - ip4
    - Faixa de endereço IPv4 permitida
    - ip4:<endereço>/<máscara>
    - ip4:10.0.0.0/24

SPF

- Mecanismos
  - ip6
    - Faixa de endereço IPv6 permitida
    - ip6:<endereço>/<máscara>
    - ip6:1080::8:800:200C:417A/64
  - exists
    - Verifica se nome possui entrada NDS
      - Utilização de macros
      - Verificação em listas DNSBL
        - exists:%{ir}.dnsbl.com.br
        - %{ir} => Expandido para endereço IP no formato reverso

SPF

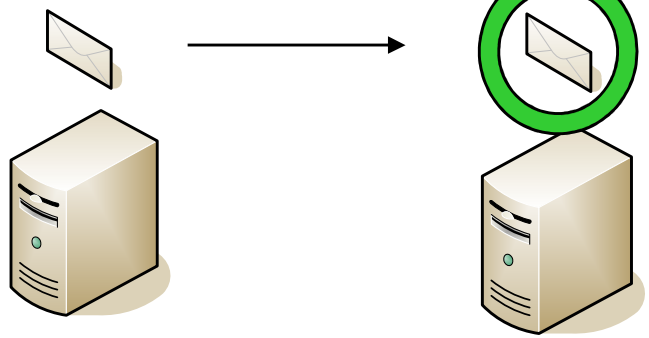
- Modificadores
  - Ações tomadas
  - +
    - Conformidade com a política
    - Pode ser omitido
  - -
    - Servidor não autorizado
  - ?
    - Não define se está autorizado ou não
  - ~
    - Resultado intermediário entre + e –

SPF

v=spf1 a mx -all

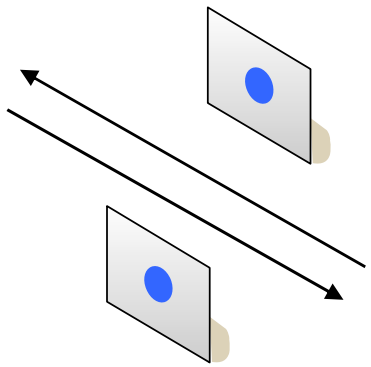
Remetente: marcelo@dominio-a.com.br
Destinatário: carlos@dominio-b.com.br

Consulta DNS tipo MX do domínio: dominio-a.com.br



Servidor A
mail.dominio-a.com.br
IP: 10.20.30.41

Servidor B



Registros MX:
mail.dominio-a.com.br

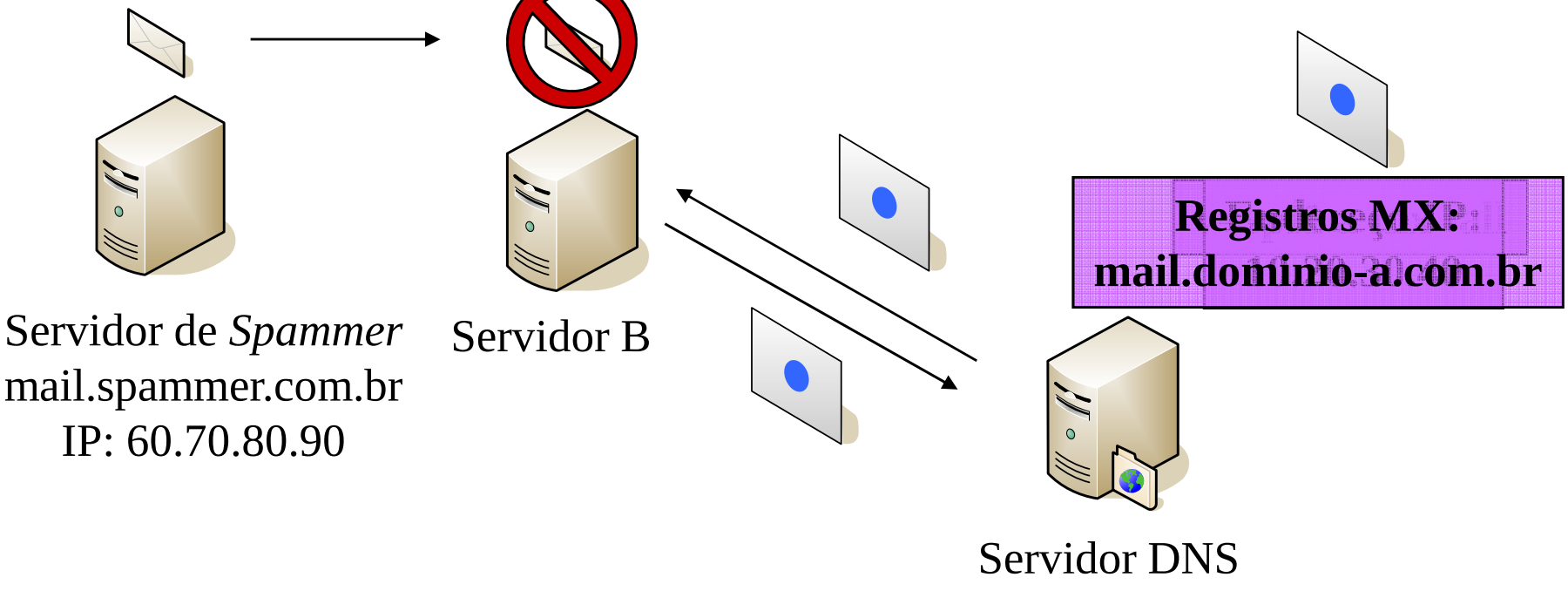
Servidor DNS

SPF

v=spf1 a mx -all

Remetente: marcelo@dominio-a.com.br
Destinatário: carlos@dominio-b.com.br

Consulta DNS tipo MX do domínio: dominio-a.com.br



Servidor de *Spammer*
mail.spammer.com.br
IP: 60.70.80.90

Servidor B

Registros MX: mail.dominio-a.com.br

Servidor DNS

SPF

- Falsos positivos
  - Políticas muito restritivas
- Falsos negativos
  - Falha na autenticação dos usuários
    - *Spammer* envia através de servidor autorizado
  - Política irrestrita para domínios
    - *Spammers* enviam como usuário do domínio
    - Verificação positiva do SPF

Desafio e resposta

- Limitar envio de spam
  - Custo maior que lucro para spammers
    - Lucro da ordem de 0,01 centavos/mensagem
  - Custo baixo para usuários legítimos
  - Punir mecanismos automatizados de envio de mensagens
    - Realização de desafios

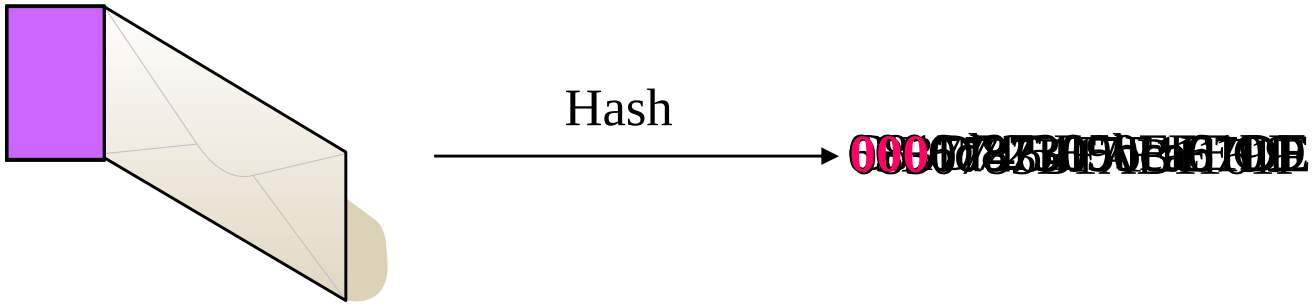
Desafio e resposta

- Realização de desafio
  - Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA)
    - Projetados para humanos resolverem
    - Digitar letras de uma imagem
      - Deficientes visuais

  - Desafios computacionais
    - Projetados para computadores resolverem
    - Complexos computacionalmente
      - Tempo grande para resolução
      - Fácil verificação
  - Pagamentos financeiros

Desafio e resposta

- Exemplo de desafios computacionais
  1. Sortear prefixo aleatório
  2. Concatenar prefixo à mensagem
  3. Calcular hash do conjunto
  4. k primeiros bits do hash em zero?
     - Sim => Fim
     - Não => Volta para 1



Desafio e resposta

- Custos financeiros indiretos
  - CAPTCHA
    - Média de 480/h
    - Salário de $10/h
    - C=2 centavos
  - Computacional
    - Resolver desafios => Tempo t
    - $1000/ano para manter uma máquina
    - t=1min => C=0,02 centavos
    - t=1h => C=11 centavos

Desafio e resposta

- Custo para criação de contas
  - Notificação de envio de *spams*
    - p – Probabilidade de notificação de spam
      - Dados da MSN TV
      - p em torno de 1/800 a 1/900
    - L dias após receber *spam*
      - Dados do Hotmail
      - L em média 2,375 dias
  - Limite de mensagens por dia (D)
    - Dados do Hotmail
    - D=100

Desafio e resposta

- Cobrança inicial ao criar conta
- Limite mínimo de LD mensagens
- Após L dias:
  - Probabilidade de reclamação: $q=1-(1-p)^D$
  - Para D pequeno => $q \approx pD$
  - Mensagens enviadas até o término da conta
    - $LD+D/q \approx LD+D/(pD) \approx LD+1/p$
  - Para L e D pequenos
    - D não influi
    - Criação de mais contas
    - $D<<1/p$ aumenta custos para os provedores
    - Custo/mensagem => Cp

Desafio e resposta

- Baixa eficiência para apenas uma cobrança
- Cobranças adicionais por mensagem
  - Impor uma cobrança para cada mensagem
    - Custo/mensagem=C
  - Impor cobranças a cada n mensagens
    - Custo/mensagem=C/n
  - Impacto negativo nos usuários legítimos

Desafio e resposta

- Custos para cada kn mensagens
  - Custo C para cada n mensagens
  - Cobrado um máximo de k vezes
  - Limite diário de D mensagens
  - Custo diário => DC/n

Desafio e resposta

Dia	Probabilidade de Término da Conta	Custo por dia
1	0	DC/n
2	0	DC/n
...		
L	q	DC/n
L+1	q	DC/n
...		
nk/D	q	DC/n
nk/D+1	q	0
nk/D+2	q	0
...		

Desafio e resposta

- Custo por mensagem

$$C/n - \frac{\frac{((1-q)^{1+nk/D-L}}{q} C/n}{L + (1-q)/q}$$

- Se 1+nk/D-L é grande
  - Custo/mensagem ≈ C/n
  - Mesmo custo para cobrança a cada n mensagens

Desafio e resposta

- Custo razoável e menos intrusivo
  - Para n=100, k=10, D=100, L=2, p=1/1000 e C=2
  - Custo por mensagem=0,012 centavos
  - k=20 => 0,017 centavos
  - K=30 => 0,019 centavos
  - K=inf => 0,020 centavos
- Usuário legítimo
  - Não terá a conta terminada
  - Para k=10 e 10 mil mensagens enviadas
    - Custo por mensagem => 0,002 centavos

Desafio e resposta

- Desafios a cada nova comunicação remetente/destinatário
  - Similar às listas cinzas
    - Quarentena de remetentes
    - Remetente sai da quarentena após resolver desafio
      - Mensagem original encaminhada
      - Servidor não necessita reenviar mensagem
  - Listas de mensagens

Desafio e resposta

Subject: RE: ola
From: AntiSpam UOL
Date: Thu, 10 Aug 2006 04:24:22 -0300 (BRT)
To: marcelo@gmail.com>

ANTISPAM UOL » TIRA-TEIMA

Olá,

Você enviou uma mensagem para **marcelo@uol.com.br**
Para que sua mensagem seja encaminhada, por favor, **clique aqui**

Esta confirmação é necessária porque **marcelo@uol.com.br** usa o Antispam UOL, um programa que elimina mensagens enviadas por robôs, como pornografia, propaganda e correntes.

As próximas mensagens enviadas para marcelo@uol.com.br não precisarão ser confirmadas*.
*Caso você receba outro pedido de confirmação, por favor, peça para marcelo@uol.com.br incluí-lo em sua lista de autorizados.

Atenção! Se você não conseguir clicar no atalho acima, acesse este endereço:
http://tira-teima.as.uol.com.br/challengeSender.html?data=k5zDy0WBRB2XrHnK406

Hi,

You´ve just sent a message to **marcelo@uol.com.br**
In order to confirm the sent message, please **click here**

This confirmation is necessary because **marcelo@uol.com.br** uses Antispam UOL, a service that avoids unwanted messages like advertising, pornography, viruses, and spams.

Other messages sent to marcelo@uol.com.br won't need to be confirmed*.
*If you receive another confirmation request, please ask marcelo@uol.com.br to include you in his/her authorized e-mail list.

Warning! If the link doesn´t work, please copy the address below and paste it on your browser:
http://tira-teima.as.uol.com.br/challengeSender.html?data=k5zDy0WBRB2XrHnK406

Use o **AntiSpam UOL** e proteja sua caixa postal

Desafio e resposta

Desafio e resposta

- Falsos positivos
  - Listas de mensagens
    - Mecanismo automatizado
    - Mensagens legítimas
- Falsos negativos
  - Máquinas zumbis
    - Resolvem desafios computacionais
- Impacto para o usuário
  - Podem ser altos ou imperceptíveis

Perspectivas Futuras

Perspectivas Futuras

- Constante evolução
  - *Spammers* se adaptam
  - Mecanismos se adaptam
  - Ciclo de evolução
- Novas formas de *spam*
  - *Spams* de voz
  - *Spams* de vídeo
  - Manipulação de resultados de busca

Perspectivas Futuras

- *Spams* de voz
  - SPIT (*Spam Over Internet Telephony*)
  - Aumento da utilização de Voz sobre IP
    - Nova área para *spammers*
  - Menores custos
    - Internet => Comutação de pacotes
      - Várias ligações simultâneas
    - Rede telefônica => Comutação de circuitos
      - Uma ligação por linha
      - Custos mais altos
  - Mais intrusivo
    - Interrompem atividades com chamadas

Perspectivas Futuras

- *Spams* de voz
  - Novos mecanismos anti-*spam*
  - Classificação por conteúdo
    - Inviável
    - Reconhecimento de voz
      - Complexo
      - Baixa eficiência
    - Análise após atender ligação
      - Usuário já foi incomodado

Perspectivas Futuras

- *Spams* de vídeo
  - Vídeos em páginas
    - Exibidos de forma não solicitada
  - Classificação por conteúdo
    - Inviável

Perspectivas Futuras

- Manipulação de resultados de busca
  - Conluio entre sítios para aumentar popularidade
    - Referências entre os sítios
    - Palavras chave específicas
      - Palavras relacionadas
      - Palavras comuns
  - Aumento da relevância do sítio
    - Primeiros resultados
    - Engana usuários

Perspectivas Futuras

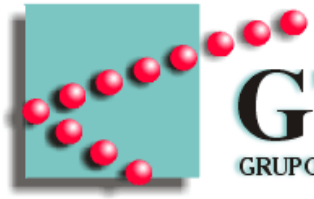
- Usuários com capacidade limitada
  - Sistemas independentes da intervenção
- Sistemas atuais
  - Auto-aprendizado
- Sistemas Autônomos
  - Novo paradigma
  - Auto-aprendizado
  - Auto-gerenciamento
  - Auto-manutenção
  - Auto-configuração
  - Auto-recuperação

Perspectivas Futuras

- Sistemas bio-inspirados
  - Baseados em características biológicas
    - Evolução durante milhares de anos
    - Eficientes na solução de problemas
  - Sistema imunológico humano
    - Classifica e combate antígenos
    - Similar ao problema do *spam*

Perspectivas Futuras

- Atividade financeira atrativa
  - Propaganda a baixo custo
  - Fraudes
- Batalha interminável
- Destruição do modelo de negócios
  - Conscientização dos usuários



Técnicas de Defesa Contra *Spam*

Danilo Michalczuk Taveira[1], Igor Monteiro Moraes[1],
Marcelo Gonçalves Rubinstein[2] e Otto Carlos Muniz Bandeira Duarte[1]

[1] Universidade Federal do Rio de Janeiro – PEE/COPPE – DEL/Poli
[2] Universidade do Estado do Rio de Janeiro – PEL/DETEL/FEN

Apoiado pelos recursos da CAPES, CNPq, FAPERJ, FINEP, RNP e FUNTTEL