Spamメール対策の現状と課題

岡山大学 総合情報基盤センター
山井 成良

2004/8/5

SS研究会

Spamメールとは

- SPAM（全て大文字）
  - 米Hormel Foods Corporationの商品&登録商標
    - http://www.spam.com参照
  - “Monty Python’s Flying Circus”の劇に登場
- spam（全て小文字）
  - 一方的かつ大量に送られる電子メール
  - Hormel Foods社も公認
  - UCE (Unsolicited Commercial E-mail)
  - UBE (Unsolicited Bulk E-mail)

Spamメール蔓延の背景(1)

- 安価な宣伝コスト
  - 発信先
    - WWW, ネットニュース等から容易に収集可能
    - 無効なアドレスと有効なアドレスの区別も容易
      - エラーメールの有無で判断
      - 「不要」と返信しても逆効果
    - 業者間での流通
      - 有効なメールアドレスを売買
      - メールアドレスを宣伝するspamメールも存在
      - 最近はspamフィルタを宣伝するspamメールが多い

Spamメール蔓延の背景(2)

- あるspamメール

From: ********* ******* <**********@*******.***>
To: **** *******@******** ***.*** ******* **.**
Subject: Take advantage of the Bulk Email Special today? Broadcasting 500.000
Only $ 59.95
Date: Wed, 30 Oct 2002 12:56:30 -0500

MULTILEVEL MARKETING OPPORTUNITIES

PRODUCT ORDER Disks are in TEXT file format and fully EXPORTABL:

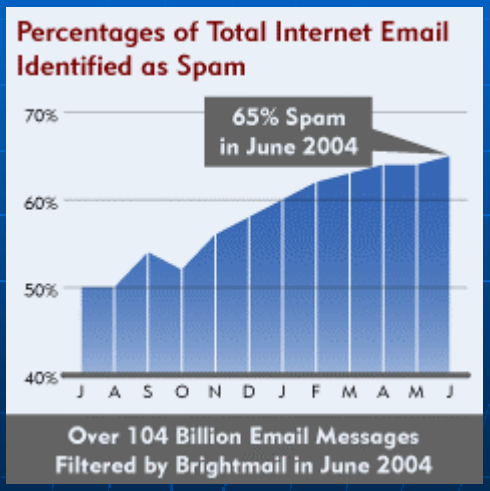
1)[ ] 200 Million email addresses all fresh!!!!! ==Only $69.95==
2)[ ] 100 million email addresses all fresh!!!!! ==Only $49.95==
3)[ ] 1.5 Million USA Business FAX NUMBERS, ==Only $29.95==
4)[ ] 7.5 million Chinese e-mail addresses all fresh!!!! ==Only $49.95==
5)[ ] 100 Thousand Toronto Canada business fax numbers ==Only $49.95==
...
90% DELIVERABLE

Spamメール蔓延の背景(3)

- 安価な宣伝コスト（続き）
  - 発信コスト
    - 常時接続環境であれば，ほぼ無料
    - 不正中継を利用すれば帯域も無関係
- 発信者の匿名性
  - 詐称した発信者アドレスでも配送
  - 返信（特にエラー通知と苦情）は不要
  - 本文中のURLへのアクセスを期待

Spamメールの現状(1)

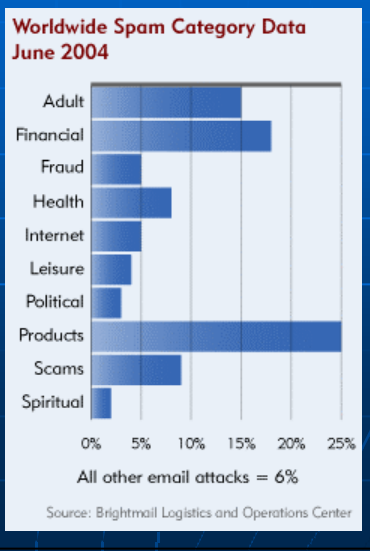
- 受信件数



http://www.brightmail.comより

Spamメールの現状(2)

- 種類と割合
  http://www.brightmail.comより



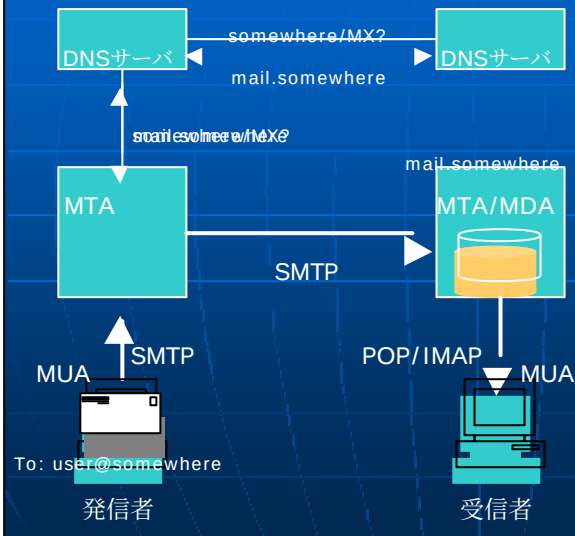
Spamメールによる被害(1)

- CPU ・ディスク・ネットワーク資源の浪費
  - メール全体の70％以上
  - 特に低速ダイアルアップ回線利用者には深刻
    - メールの受信に時間（＝通信費用）がかかる
    - メールボックスがすぐに一杯
- メールの分類・削除
  - メール受信後も時間がかかる
  - 重要なメールの見落としも問題
    - spamフィルタを用いても起こりえる

Spamメールによる被害(2)

- 発信者詐称による間接的な被害
  - spamメール発信者との誤解
    - 苦情メールへの対処
    - 信頼性の低下
      - 通常メールの受信拒否も
- エラーメールの集中
  - 発生頻度小（自組織アドレスに詐称された場合）
  - 被害は甚大
    - 事実上の分散型サービス不能(DDoS)攻撃

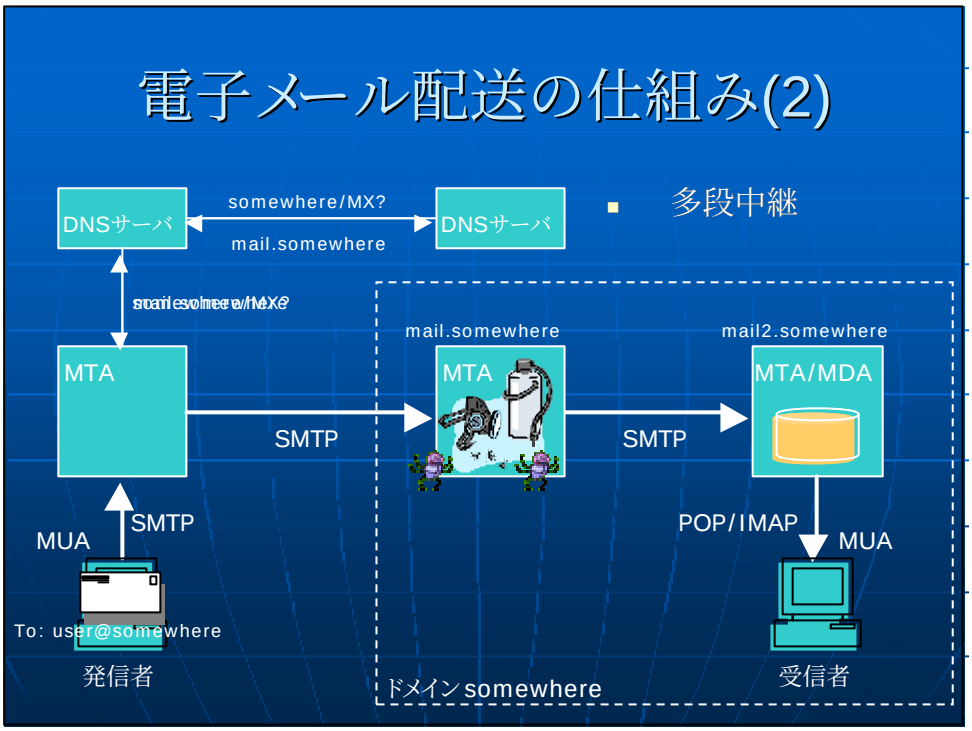
電子メール配送の仕組み(1)



- 単純な配送
  1. メッセージ発信
  2. 配送先問合せ
  3. 配送
  4. メールボックス格納
  5. 受信

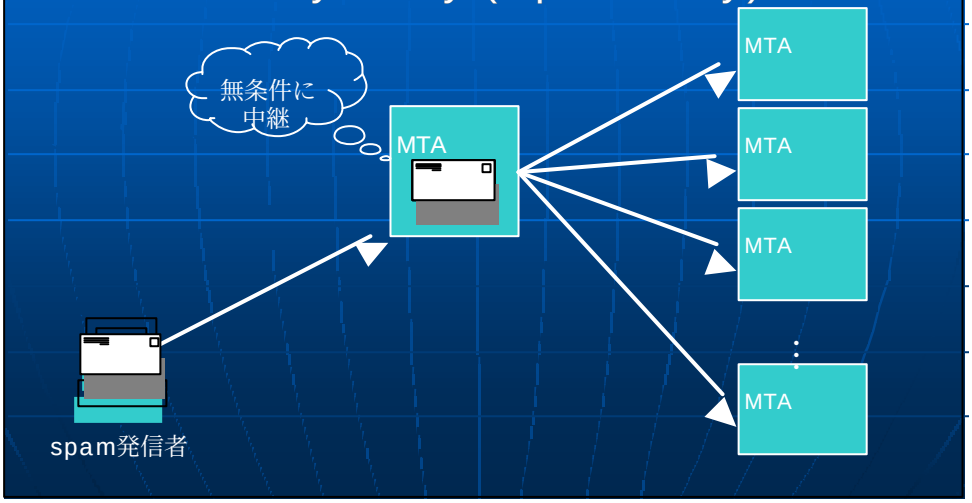
MUA: Mail User Agent
MTA: Mail Transfer Agent
MDA: Mail Delivery Agent
SMTP: Simple Mail Transfer Protocol
POP: Post Office Protocol
IMAP: Interactive Mail Access Protocol
DNS: Domain Name System
MX: Mail eXchange

電子メール配送の仕組み(2)



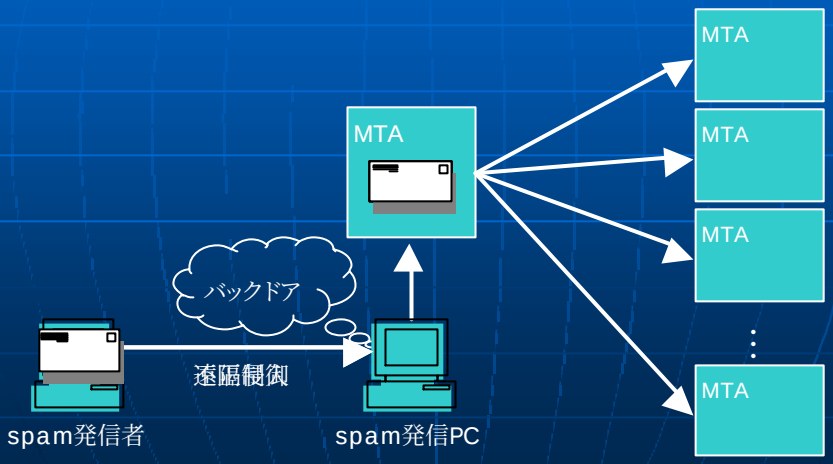
Spamメールの配送方法(1)

- Third Party Relay (Open Relay)



Spamメールの配送方法(2)

- 計算機への不正侵入

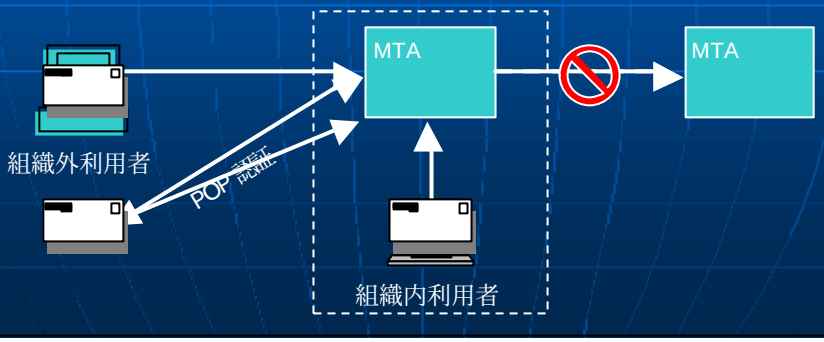


Spamメールへの対策

- ブロッキング
- フィルタリング
- 発信者詐称対策
- 法的対策

Spamメールのブロッキング(1)

- Third Party Relayの防止
  - 組織外から組織外への中継は原則として拒否
  - POP before SMTPとの併用



Spamメールのブロッキング(2)

- 発信者アドレスの確認
  - DNSを用いて発信者ドメインの存在を確認
  - 実在するドメイン名に詐称されると効果なし
    - 通常メールでも発信ホストと発信者アドレスが不一致の場合あり
  - 本来であれば不要な通信が発生
  - ネットワークが不安定だと，通常メールでも受信を拒否する場合あり

Spamメールのブロッキング(3)

- ブラックリストサービス
  - 代表例
    - MAPS RBL (Realtime Blackhole List)
    - ORBS (Open Relay Blocking System)⇒停止
  - DNSを利用
    - spamメール発信ホスト, open relayホストを登録
    - 登録ホストからのメールは無条件で受信拒否
  - トラブルも多い
    - 登録ホストからは通常メールも（ある日突然）拒否
    - 対策完了後も復旧に時間を要する
    - 詐称アドレスが実在した場合，巻き添えを食らう

Spamメールのブロッキング(4)

- Greylisting
  - 別名「お馴染みさん」方式
  - 初めてのMTAに対しては，一旦エラーを返す
    - エラー番号は400番台・・・一時的なエラー
  - 2回目以降はwhite listへ
  - Spam senderは再配送しないとの前提に基づく
  - 現状では結構効果的
  - 配送遅延が生じる点が問題

Spamメールのブロッキング(5)

- 発信者・発信ホストの認証
  - SMTPの拡張(RFC2554)
    - 新しいコマンドAUTHの追加
    - メール発信時に利用者を認証
    - 特別なソフトウェアが必要
  - auth-smtpd(岡山大・大阪市立大)
    - 計算機利用時・ネットワーク接続時の認証情報を利用
    - 強制的に発信者アドレスを書換えまたはSender追加
    - 岡山大では教育用計算機環境で利用
    - 大阪市立大ではダイアルアップ接続環境で利用

Spamメールのフィルタリング(1)

- 基本方針
  - メール受信後にspamメールかどうかを判断
  - spamメールは削除あるいは別に格納
- 代表的な方法
  - ルールベースフィルタリング
  - ベイジアンフィルタリング
  - 協調分散型spamデータベース

Spamメールのフィルタリング(2)

- ルールベースフィルタ
  - spamメールの特徴をルールとして記述
    - 本文中に「$」「viagra」など特定のキーワードを含む
    - FromとToが同じアドレスなど
  - マッチした場合，ルールに対応したスコアを加算
  - 一定のスコア以上のものをspamと判定
  - 欠点＝柔軟性の欠如
    - スコアの調整は可能だが限界が存在
    - 新たな手口には新たなルールが必要

Spamメールのフィルタリング(3)

- ベイジアンフィルタ(Bayesian filter)
  - キーワード(単語，3字組等)の出現率を学習
  - キーワードの種類に応じてspamメールを判定
  - ベイズ則 P(A|B)= P(A)P(B|A)/P(B)を利用
    - 事象A・・・メッセージがspamメールである
    - 事象B・・・メッセージがキーワードを含む
  - 有効なキーワードの例：
    - ff0000 ・・・ HTMLメールでの赤色フォントの指定
  - 新しい手口にもある程度対応可能

Spamメールのフィルタリング(4)

- 分散協調型spamデータベース
  - 既に判定済みのspamメールの再受信を排除
    - 同一メッセージが多数の利用者に(何回も)配送されることを逆利用
  - 利用者がspamメールをデータベースに登録
  - メール受信時に同一メッセージの存在を問合せ
    - 一定数以上の登録があればspamメールと判定
  - spamメールの認識率が低い点が問題
    - 登録までのタイムラグあり
    - 内容の一部変更に弱い ⇒ fuzzy checksum

Spamメールのフィルタリング(5)

- 性能評価
  - False negative
    - spamメールを通常メールと判定
  - False positive
    - 通常メールをspamメールと判定

  ※ False positiveのほうが問題
    - 重要なメールを見落とす可能性がある
    - 実例：spam苦情メール
  - 客観的な評価は難しい
    - 受信するメール集合やspam定義が人により異なる
    - spamの手口が時間とともに進化

Spamメールのフィルタリング(6)

- 性能の比較
  - 2002年9月 IBM developer worksより
    http://www.ibm.com/jp/developerworks/linux/021129/j_l-spamf.html

ルールベース

分散協調型

フィルタリング方法	非spam集合 (正解 vs. FP)	spam集合 (正解 vs. FN)
実際の数	1851 vs. 0	1916 vs. 0
SpamAssassin	1846 vs. 5	1558 vs. 358
3文字Bayesian	1849 vs. 2	1774 vs. 142
単語Bayesian	1847 vs. 4	1819 vs. 97
Pyzor	1747 vs. 0 (エラー4個)	943 vs. 971 (エラー2個)

Spamメールの発信者詐称対策(1)

- エラーメールの集中
  - 2億通×10% = 2000万通
  - 発生頻度小
    - 詐称アドレスが自組織ドメインの場合のみ発生
  - 発生時の影響大
    - MTA・ネットワークの過負荷
    - ディスクの大量消費（実在アドレスへの詐称の場合）
  - 詐称防止は困難
    - 郵便で差出人の詐称を防止することと同じ

Spamメールの発信者詐称対策(2)

- 2002年11月に国内ISPで発生した事例
  - 30万通以上のエラーメールが集中
  - 通常メールの配送遅延は最大15時間
  - 復旧に約2.5日間
    - 11/5 9:30am ～ 11/7 11:00pm
  - 恐らく実在アドレス
    - アドレスリスト中に含まれるものと推察

Spamメールの発信者詐称対策(3)

- 対策方法
  - 岡山大で現在開発中
    - 科学研究費補助金採択課題
      - 基盤研究(C)(2) 課題番号15500039
    - DNSのキャッシュの有無を利用
      - エラーメール大量受信時にMXレコードを変更
      - 通常やりとりのあるMTA
        - キャッシュヒット⇒ 通常のMTAで優先処理
      - エラーメールを送る大多数のMTA
        - キャッシュミス ⇒ エラーメール処理用MTAで処理

Spamメールに関する法的対策(1)

- 技術的なspamメール対策の限界
  - ブロッキング・フィルタリングでは spamメール発信者は不利益を被らない

  ⇒何らかの法的な対策が必要
- 法的なspamメール対策の実施国
  - 日本
  - アメリカ合衆国
  - EU
  - 韓国など

Spamメールに関する法的対策(2)

- 日本における法律
  - 迷惑メール対策法(2002年7月施行)
    - 特定商取引に関する法律の一部を改正する法律
      （改正特定商取引法）
    - 特定電子メールの送信の適正化に関する法律
      （特定電子メール法）

  ※ 広告メールを全面的に禁止するものではない
  （オプトアウト方式）
    - 広告は企業活動にとって必要
    - CM, ダイレクトメールとの比較

Spamメールに関する法的対策(3)

- 迷惑メール対策法の比較

法律名	改正特定商取引法	特定電子メール法
担当官庁	経済産業省	総務省
規制対象	事業者	メール発信者
表示義務	1. メールアドレス 2. 未承諾広告※ 3. オプトアウト方法	1. 未承諾広告※ 2. 氏名・住所 3. 発信アドレス 4. 受信アドレス
禁止事項	拒否者への送信	・拒否者への送信 ・架空アドレスへの送信
罰則	・2年以下の懲役 ・300万(法人は2億)円以下の罰金	50万円以下の罰金

Spamメールに関する法的対策(4)

- 迷惑メール対策法の効果 ⇒ 殆どなし(3件)
  - 殆どの広告メールは表示義務に違反
    - 違反者の調査が困難
      - 発信者情報の欠落
      - 多くの場合，携帯電話から発信 ⇒ 追跡が困難
    - 違反しても直ちには処罰されない
      - 措置命令に違反した場合に初めて罰金・懲役
  - 平成15年10月9日に初めて2社が行政処分
    - 件名に「未承諾広告※」「※未詳諾広告※」などと表示
    - 平成14年8月頃から平成15年9月頃まで発信
    - 平成15年6月以降は送信者情報表示義務にも違反

Spamメールに関する法的対策(5)

- 迷惑メール対策法の効果（続き）
  - 多数の抜け道が存在
    - 登録者への送信は規制の対象外
      - 登録会員への送信を装う
    - 広告メールが対象
      - 友人からの情報交換メールを装い，URLのみ表記

⇒ 「優良」迷惑メールしか効果なし

- 「正直者が馬鹿を見る」状態

Spamメールに関する法的対策(6)

- 米国での迷惑メール対策法
  - CAN-SPAM法(2004年1月施行)
    - 表示義務
      - オプトアウト方法の提示
      - 有効な返信アドレス
      - 広告メールの表示
    - 禁止事項
      - 発信元詐称
      - 偽の件名の表示
      - 自動的なアドレス収集・発信用アドレスの取得
      - ラベルを含まない「性的内容が中心の素材」の発信
    - 罰則あり

Spamメールに関する法的対策(7)

- 米国での迷惑メール対策法(続き)
  - 条件を満たすspamメール発信は合法
    - spam業者にとってはクリスマスプレゼント
  - 州法に優先
    - カリフォルニア州(2004年1月施行)など３６州で制定
      - より強力なオプトイン（事前登録）方式が事実上無効に
      - 但し，オプトイン方式は合衆国憲法に反するとの指摘あり
  - CAN-SPAMは事実上 “You CAN SPAM”
    - 本当はControlling the Assault of Non-Solicited Pornography and Marketing Act

Spamメールに関する法的対策(8)

- EUでの迷惑メール対策法
  - 2003年10月 31日発効
  - 原則的にはオプトイン方式
    - 国内法も遵守する必要あり
  - マーケティング業者やプロバイダが対処に苦慮
  - 効果は不明

Spamメールに関する法的対策(9)

- まとめ
  - オプトアウトとオプトイン
    - 「表現の自由」「検閲の禁止」との兼ね合い
    - ユーザあるいは優良業者のいずれかが苦慮
  - 国外から来るspamメールへの効果
    - 国際法なし
    - 国により規制がばらばら
  - 効果は疑問だが，規制は必要
    - 現在は「正直者が馬鹿を見る」状態

最近のspamメールの手口(1)

- spamメール発信の手口
  - spam業者とウィルス作者の協力
    - 例：Sobig，MiMail.F
    - 感染成功時にバックドアを仕掛け，作者に通知
    - 機会を見てspamメールの発信に利用
  - ヘッダ偽造技術の向上
    - 発信ホストの秘匿化

最近のspamメールの手口(2)

- フィルタ回避の手口
  - 件名や本文をMIMEエンコードする
  - 本文を圧縮する
  - キーワードの綴りを変える
    - 例：Vi@gra, V i a g r a, rem0ve
  - HTMLのコメントを挿入する
    - 例：Via<!--random string-->gra
  - 文字を画像として表示する
  - 無関係な単語を含む
  - 苦情メール・エラーメールを装う

最新のspamメール対策(1)

- Spamメールのブロッキング(1)
  - Sender ID (AOL, MS)
    - SPF (Sender Policy Framework) + Caller ID
      - 対象がエンベロープ(SPF)かヘッダ(Caller ID)かの違い
    - ドメインに対して発信可能なIPアドレスをDNSに登録
      - 例：gnu.org text = "v=spf1 ip4:199.232.76.160/27 ip4:199.232.41.0/28 ?all"
    - 受信時に，発信IPアドレスとDNSの登録内容を照合
      - 例：199.232.41.1からwho@gnu.orgが送ったメール
        ⇒ 正当
      - 例：192.168.1.2からwho@gnu.orgが送ったメール
        ⇒ 詐称の可能性あり(正確には“?all”=中立)

最新のspamメール対策(2)

- Spamメールのブロッキング(2)
  - DomainKeys (Yahoo!)
    - DNSにMTAの公開鍵を登録
    - 発信MTAはメッセージに署名を添付
    - 受信時に署名の正当性を確認
      - 発信MTAの詐称を防止
  - Sender ID, DomainKeysの問題点
    - これらの仕組みを導入したドメインしか効果なし
    - 全てのMTAに導入されるまでには長時間が必要

最新のspamメール対策(3)

- Spamメールのフィルタリング(1)
  - 本文中のURLの確認
    - 東大情報基盤センターと株式会社ディープソフトの共同研究
    - 本文中に含まれるURLの「中身」を確認
      - 自動的にページを取得
      - ページ中のキーワードによりspamかどうかを判定
    - 問題点
      - ページ中にキーワードが含まれる保証がない
      - ネットワークへの負荷や処理時間の増加が大きい
      - URLへのアクセスで退会してしまうMLなども存在

最新のspamメール対策(4)

- Spamメールのフィルタリング(2)
  - 分散協調型spamデータベースの遅延評価
    - 岡山大総合情報処理センターなどで開発中
    - メール受信時ではなく, 読出し時にチェック
      - 遅延評価によりspam判定率の向上を期待
    - MUA側ではなくMDA(POPサーバ)側でチェック
      - 余分な通信を発生させない
    - 問題点
      - どこまでspam判定率が向上するか

最新のspamメール対策(5)

- Spamメールに関する法的対策
  - 特定電子メール法(総務省)の改正
    - 以下のURL参照
      http://www.nikkei.co.jp/news/keizai/20040531AT1F3000330052004.html
      http://www.asahi.com/tech/asahinews/TKY200406060122.html
    - 2005年の通常国会への提出を検討中
    - 「直罰規定」を検討
    - 捜査の迅速化
      - 最初から警察による捜査を実施
      - 警察によるログの押収が容易になるとの懸念も

おわりに(1)

- 当分は「いたちごっこ」状態
  - 例：最近のSpamAssassinでのスコア = -4.9
- 多方面からの取組みが必要
  - ブロッキング，フィルタリング，法律
  - 計算機におけるセキュリティ対策・ウィルス対策
  - インターネットの持つ匿名性への対策
- 国際的な枠組みが必要
  - 国外からのspamメールへの対処

おわりに(2)

- 今後はアジア(特にCJK)が問題か
  - 現在のフィルタリングは殆ど英語が対象
  - 現実に中国・台湾からのspamメールは多い
  - 韓国は既にspam大国
    - ISPが自衛のためフィルタリング
  - 日本は携帯電話でのspamメールが社会問題化

⇒ 皆さんの活躍に期待

おわりに(3)

- Spam対策メーリングリスト
  - anti-spam-request@cc.okayama-u.ac.jp 宛に以下の内容のメールを送信

    subscribe
    end