迷惑メール対策の最新研究

山井 成良（岡山大学）

インターネット協会・迷惑メール対策セミナー（福岡）

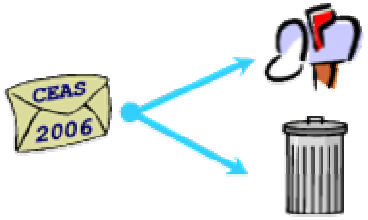
November 15, 2006

目次

- CEAS 2006報告
- 私の研究

CEAS 2006報告

- 正式名称
  - Third Conference on Email and Anti-Spam
- 開催日程
  - 7月27-28日開催
- 開催場所
  - Microsoft SVC conference Center, Mountain View, California
- 論文採択数
  - 27件/79件 (採択率：約34%)
- URL
  - http://www.ceas.cc



July 27 - プログラム1

- 9:00 am Introductory Remarks
- 9:05 am Invited Talk
  - The Underground Economy
- 9:50 am Abuse detection and control
  - Dynamic Port 25 Blocking to Control SPAM Zombies
  - Algorithmically Determining Store-and-Forward MTA Relays Using DomainKeys
  - Can DNS-Based Blacklists Keep Up with Bots?
- 10:30 am Break
- 11:00 am Counter-Attack
  - Spamalot: A Toolkit for Consuming Spammers' Resources
  - Breaking Anti-Spam Systems with Parasitic Spam
  - Observed Trends in Spam Construction Techniques: A Case Study of Spam Evolution

July 27 - プログラム2

- 11:55 pm Learning-based Filters I
  - Spam Filtering with Naive Bayes - Which Naive Bayes?
  - An Adaptive, Semi-Structured Language Model Approach to Spam Filtering on a New Corpus
- 12:45 pm Lunch
- 2:15 pm Email Enhancement
  - An Email and Meeting Assistant Using Graph Walks
  - Email Thread Reassembly Using Similarity Matching
  - "Sorry, I Forgot the Attachment:" Email Attachment Prediction
  - CC Prediction with Graphical Models
- 3:35 pm Break

July 27 - プログラム3

- 4:00 pm Social behavior
  - The Effects of Anti-Spam Methods on Spam Mail
  - Sender Reputation in a Large Webmail Service
  - Using E-Mail Social Network Analysis for Detecting Unauthorized Accounts
  - An Empirical Study of Clustering Behavior of Spammers and Group-based Anti-Spam Strategies
- 5:30 pm Reception
- 6:30 pm Banquet

July 28 - プログラム1

- 9:00 am Invited Talk
  - Web Spam: Current Techniques and Future Challenges (Hector Garcia-Molina and Zoltan Gyongyi)
- 9:45 am Group Discussion
  - Considering What's Missing — Building a Work List for the Email and Messaging Research Community
- 10:15 am Break

July 28 - プログラム2

- **10:45 am Corpora**
  - Modeling Identity in Archival Collections of Email: A Preliminary Study
  - Introducing the Webb Spam Corpus: Using Email Spam to Identify Web Spam Automatically
  - Annotating Subsets of the Enron Email Corpus
  - An Exploratory Study of the W3C Mailing List Test Collection for Retrieval of Emails with Pro/Con Argument
- **12:15 pm Lunch**
- **1:45 pm 4 Selected Lightning Papers**

July 28 - プログラム3

- 2:05 pm Organizational Impact
  - Teaching Spam and Spyware at the University of C@1g4ry
  - Deployment Experience: Rolling Out a New Antispam Solution in a Large Corporation
  - Using Early Results from the 'spamHINTS' Project to Estimate an ISP Abuse Team's Task
- 2:45 pm Break
- 3:15 pm Learning-based Filters II
  - Batch and Online Spam Filter Comparison
  - Learning at Low False Positive Rates
  - Online Discriminative Spam Filter Training
  - Fast Uncertainty Sampling for Labeling Large E-mail Corpora
- 4:45 pm Business Meeting
- 5:00 pm End

今回取り上げる発表

- Dynamic Port 25 Blocking to Control SPAM Zombies
- Algorithmically Determining Store-and-Forward MTA Relays Using DomainKeys
- Can DNS-Based Blacklists Keep Up with Bots?
- Spamalot: A Toolkit for Consuming Spammers' Resources

発表1 (1/3)

- Dynamic Port 25 Blocking to Control SPAM Zombies
  - *Jonathan Schmidt (Perftech Inc.)*
- Outbound Port 25 Blocking全面的導入の問題
  - サポートセンターへの問合せが頻発
- 提案方法
  - 1分当たり40コネクション以上でブロック
    - ブロック中はブラウザに説明を表示
  - 1秒間に5以上のコネクションがない状況が1分間続けばブロックを解除

発表1 (2/3)

- 実験結果(1)
  - 利用者20,000人のサブネットで先行実験
  - 利用者からはサポートセンターへの苦情なし
  - 不正利用に対する外部からの苦情は大幅に減少
    - 導入前：1日100件程度
    - 導入後：全く苦情がない日も
  - OP25Bの影響を受ける正規利用者は5名
    - GUIでの申告メカニズムを予め用意
  - 95%のメールをブロック

発表1 (3/3)

- 実験結果(2)
  - 全利用者240,000人に対して適用
  - 不正利用に対する外部からの苦情は大幅に減少
    - 導入前：1日600件程度
    - 導入後：1日10件以下
  - 不正利用対策チームは解散
  - 1年後には利用者からサポートセンターへの苦情なし
  - ボット感染者も減少
    - 導入前：利用者全体の1%
    - 導入後：利用者全体の0.2%

発表2 (1/2)

- Algorithmically Determining Store-and-Forward MTA Relays Using DomainKeys
  - *Peter Ludemann, Miles Libbey (Yahoo! Inc.)*
- メール中継サーバの問題
  - メール中継サーバを経由すると…
    - メーリングリスト，メール転送サービス，メールゲートウェイ等
  - メールの発信元IPアドレスが不明確
  - IPアドレスベースの迷惑メール対策が無効
- 提案手法
  - メール中継サーバの検出

発表2 (2/2)

- 検出方法
  - 2種類の送信ドメイン認証を併用
    - デジタル署名ベース(DomainKeys等)
    - IPアドレスベース(SPF, Sender ID等)
  - 署名メールを送信した認証外サーバ＝中継サーバ
    - 署名の検証にはList-IDヘッダとReturn-Pathヘッダを除外
    - 例:

      To: user1@yahoo.com
      Return-Path: <user2@yahoo.com>
      Received: from 142.103.6.59 (EHLO alumni.cs.ubc.ca)
- 検証
  - 2006年1～3月にYahoo! Mailで実験
  - 8151個の中継用IPアドレスを検出

発表3 (1/4)

- Can DNS-Based Blacklists Keep Up with Bots?
  - *Anirudh Ramachandran, David Dagon, Nick Feamster (Georgia Institute of Technology)*
- 研究目的
  - DNSBLはどの程度信頼できるかを明らかにすること
- 評価基準
  - 完全性(completeness)
    - 迷惑メール発信IPアドレスのうち，どの程度を含んでいるか
  - 応答性(responsiveness)
    - 迷惑メール発信後から登録まで，どの程度の時間を要するか

発表3 (2/4)

- 調査方法
  - データ
    - SpamhausのDNSBLへの問合せ(1.5日間分)
      - 17台あるミラーサーバのうちの1つ
    - Bobaxボットネットのマスターの出すパケット(46日間分)
  - 調査手法
    - 既知のBobax感染ホストに対するDNSBLへの問合せに着目
      - 最初の問合せ時刻　迷惑メール送信活動開始時刻
      - 最初の正常応答時刻　DNSBL登録時刻

発表3 (3/4)

- **調査結果**
  - DNSBLへの問合せ
    - 問合せ全体: 81,950
    - Bobax感染IPアドレスに対する問合せ: 4,295
    - Bobax感染IPアドレス数: 2,042,991
  - 完全性
    - 46日間経過後でも4,295件中255件(6%)しか登録されず
    - ➡ 完全性は低い
  - 応答性
    - 255件中88件が1.5日間の途中で登録
    - 88件中34件は1回目の問合せ後に登録
    - ➡ 直ちに登録されるIPアドレスは少ない

発表3 (4/4)

- 調査結果 (続き)
  - DNSBL登録後のBobaxの振舞い
    - 目立った変化なし
  - 特定Bobaxホストに対する問合せ元IPアドレス/AS数
    - 60%以上のホストが1つのIPアドレス/ASから問合せ
    - 1つのホストが送信する迷惑メールは少数のドメイン宛
      - 迷惑メール送信者として報告される危険性減少のため
    - 10%程度のホストは多数のIPアドレス/ASから問合せ
      - DNSBLへの登録はこのうちの半分程度

発表4 (1/3)

- **Spamalot: A Toolkit for Consuming Spammers' Resources**
  - *Peter C. Nelson, Kenneth P. Dallmeyer, Lukasz M. Szybalski, Tom P. Palarz, Michael Wieher (University of Illinois at Chicago)*
  - http://acm.cs.uic.edu/~lszyba1/
- **研究目的**
  - 迷惑メール送信者 (人間!!) のリソースの消費
- **3種類のインテリジェント・エージェント**
  - Arthur (Nigerian)
  - Patsy (Web forms)
  - Lancelot (Phishing)・・・開発中

発表4 (2/3)

- Arthurの仕組み
  - できる限り長くメールのやり取りを続けるように返事を出す
  - ボイスメールを用意して電話をかけさせる
- 運用結果(7月28日)
  - Nigerian spam: 18件
  - Spammerから反応があったもの: 13件 (72%)
  - 最小会話長: 1通
  - 最大会話長: 44通
  - 平均会話長 (メジアン): 8通
  - ボイスメール数: 12件

発表4 (3/3)

- Patsyの仕組み
  - mortgageやdiplomaに関する迷惑メールが対象
  - Spammerの関心を引くようにWeb formを埋める
  - 専用メールアドレスやボイスメールに連絡させる
- 運用結果
  - 3月21日～7月18日の電話回数：　117件
    - 無言電話を除く

私の研究

- 発信者詐称spamメールに起因するエラーメール集中への対策手法
  - 第3回情報科学技術フォーラム(FIT2004)にて発表
  - FIT論文賞を受賞
- SMTPセッションの強制切断によるspamメール対策手法
  - 第5回情報科学技術フォーラム(FIT2006)にて発表
  - 船井ベストペーパー賞を受賞

SMTPセッションの強制切断によるspamメール対策手法

目次

- **はじめに**
- 従来のspamメール対策方法とその問題点
- SMTPセッションの強制切断によるspamメール対策手法
- 試作システムの実装と評価
- まとめ

研究の背景

spamメールの蔓延

- **資源の浪費**
- **選別コストの増大**

spam対策手法の種類と従来技法

- ブロッキング
  - spamメールの受信そのものを拒否
    - セッション開始時/途中の送信元MTA（メールサーバ）を判別
  - 代表的技法
    - tempfailing
    - DNSBL (DNS Black List)
- フィルタリング
  - メールを受信してからspamメールかどうかを判別
    - ヘッダや本文などを解析
  - 代表的技法
    - ベイジアンフィルタ
    - 分散協調spamデータベース

研究目的

- 従来技法の問題点の軽減
  - tempfailing
    - 未登録MTAに対して一時エラー
    - ✖通常メールの配送にも遅延が発生
    - ✖異なるMTAからの再送に対応不可
  - 分散協調spamデータベース
    - 登録済みspamメールとの同一性を判定
    - ✖spamメールの大量かつ早めの登録が必要

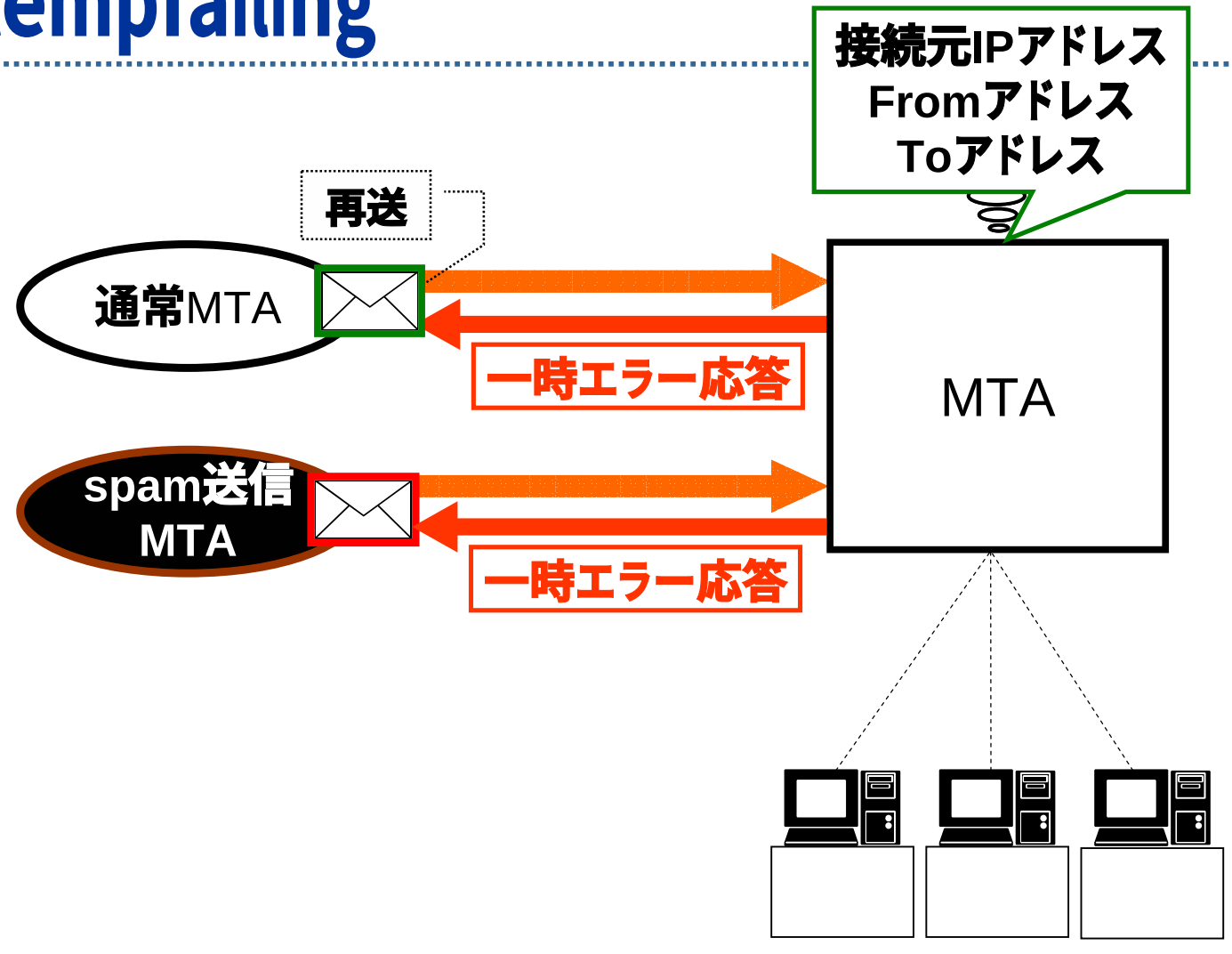
目次

- はじめに
- **従来のspamメール対策方法とその問題点**
- SMTPセッションの強制切断によるspamメール対策手法
- 試作システムの実装と評価
- まとめ

tempfailing

- 一時エラー時の動作の違いを利用
  - 通常のMTA
    - 確実性を重視
    - 一時エラー時にメールを再送する
  - spam発信MTAの多く
    - スループットを重視
    - 一時エラー時でもメールを再送しない

tempfailing



tempfailing

- **本技法の問題点(1)**

RFC2821

SMTPセッション中に一時エラー応答を受け取ったメール

⇒ 送信者は一定時間（30分以上）後に再送しなければならない

問題点（1）：通常メールの受信に遅延が発生

tempfailing

- 本技法の問題点(2)

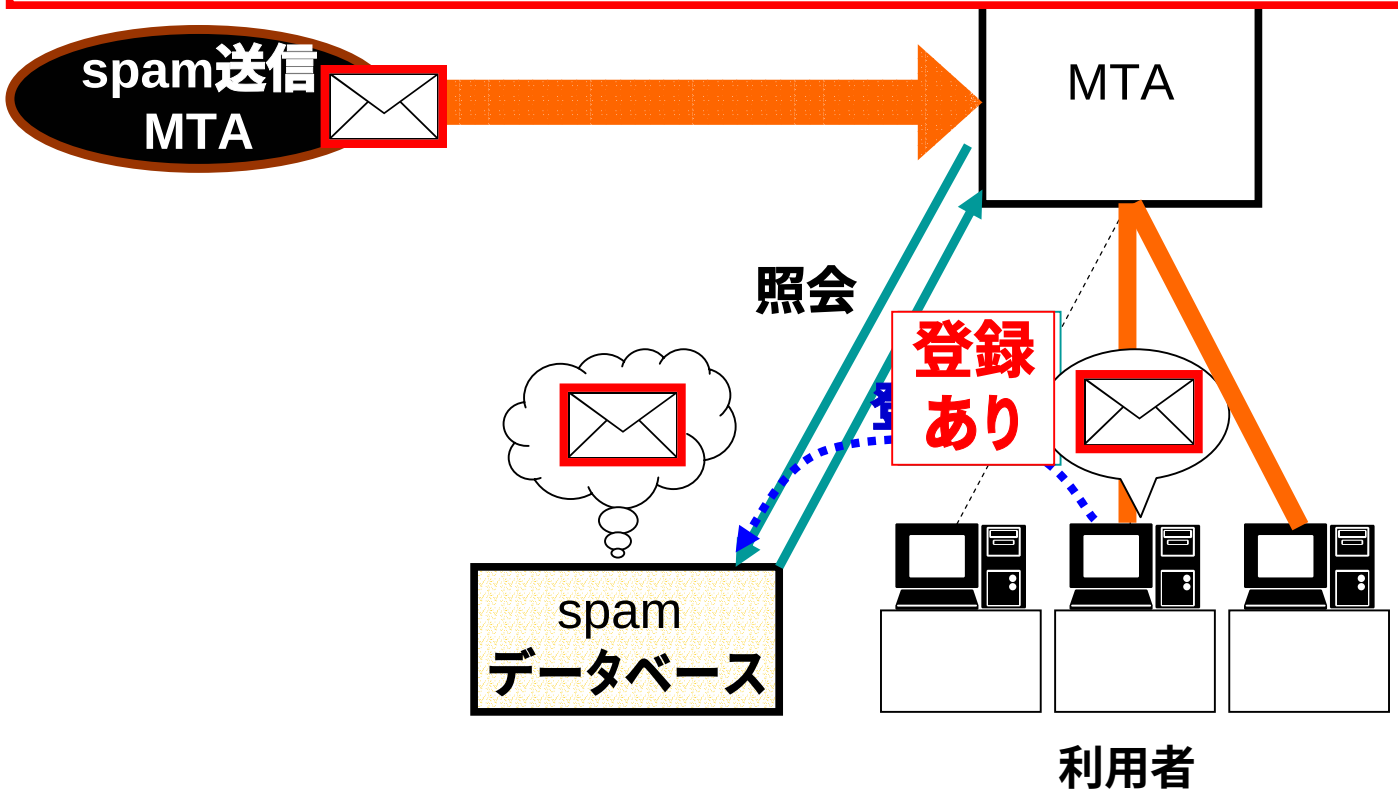
再送の判定

- 接続元IPアドレス
- Fromアドレス
- Toアドレス

問題点（2）：初回時と異なるMTAからの再送も受信拒否

分散協調spamデータベース

問題点：「実在」利用者の「既読」メールのみ登録可能



目次

- はじめに
- 従来のspamメール対策方法とその問題点
- **SMTPセッションの強制切断によるspamメール対策手法**
- 試作システムの実装と評価
- まとめ

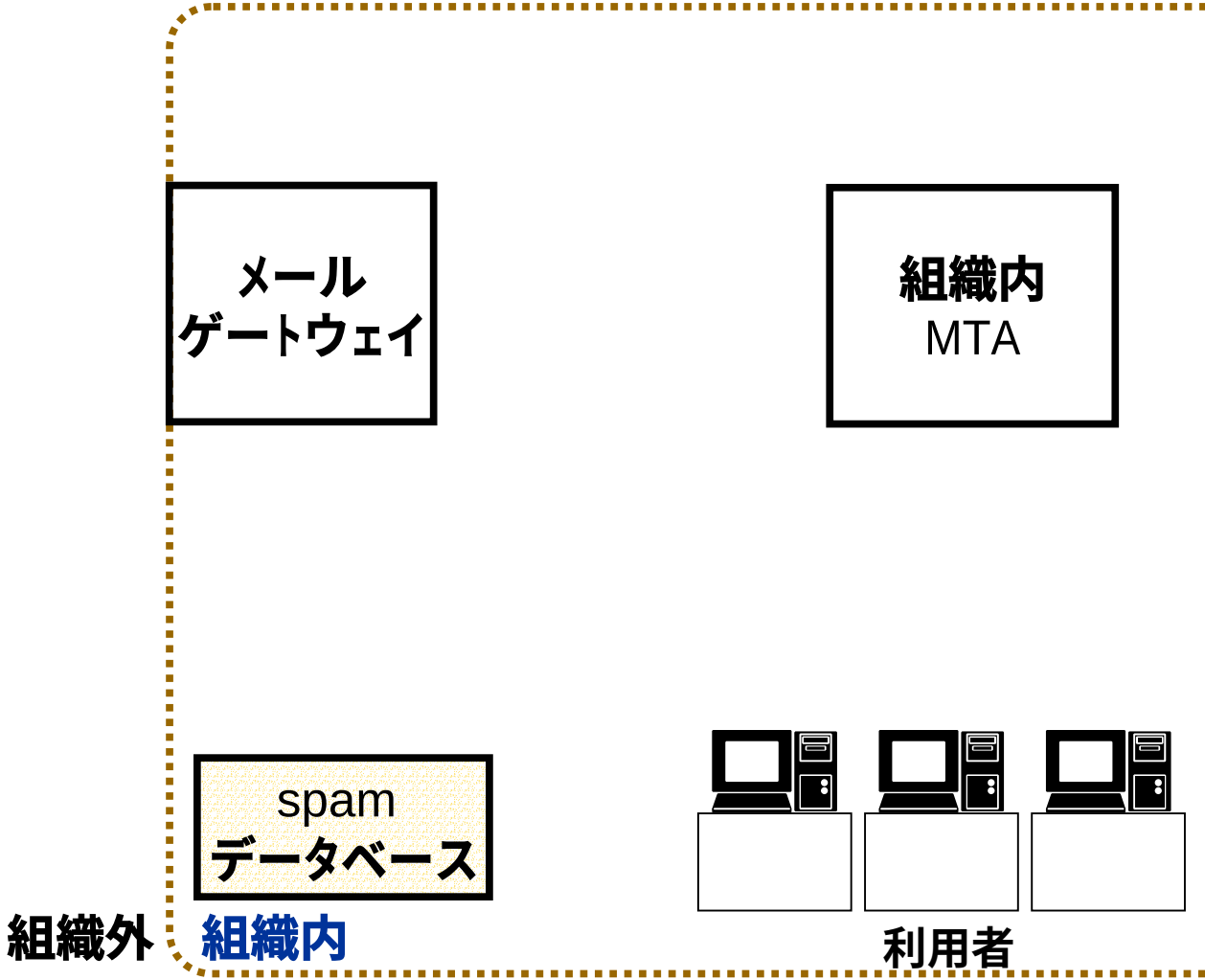
従来技法の問題点（まとめ）

- 問題点1
  - 通常メールの受信に遅延が発生
- 問題点2
  - 初回時と異なるMTAからの再送も受信拒否
- 問題点3
  - 「実在」利用者の「既読」メールのみ登録可能

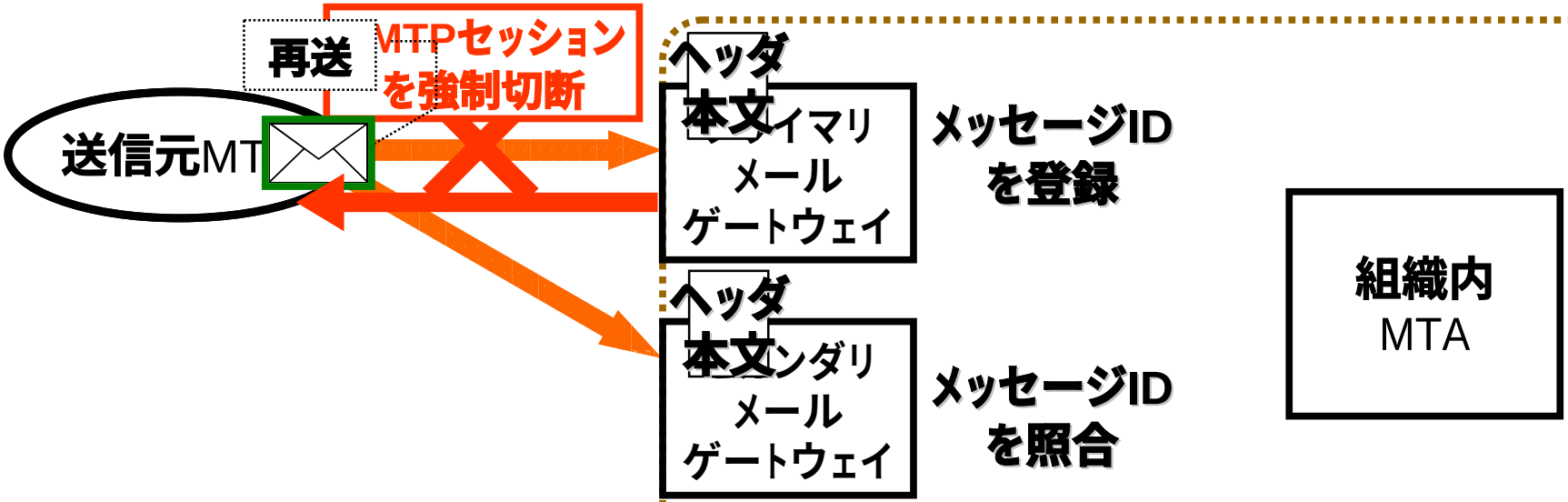
提案手法の概要

- 2台のメールゲートウェイを導入
  - 一時エラー時に直ちに2台目に再送可
- SMTPセッションを強制終了
  - ヘッダ・本文の取得が可能
  - 再送判定にヘッダ情報が利用可
- 宛先不明・未再送メールを自動登録
  - 大量かつ早期の登録が可能

想定環境

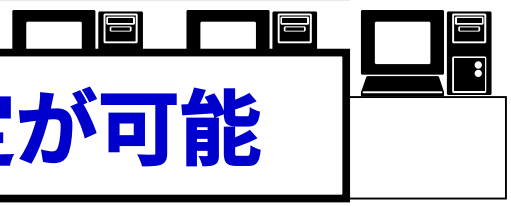


システム構成および動作(1)



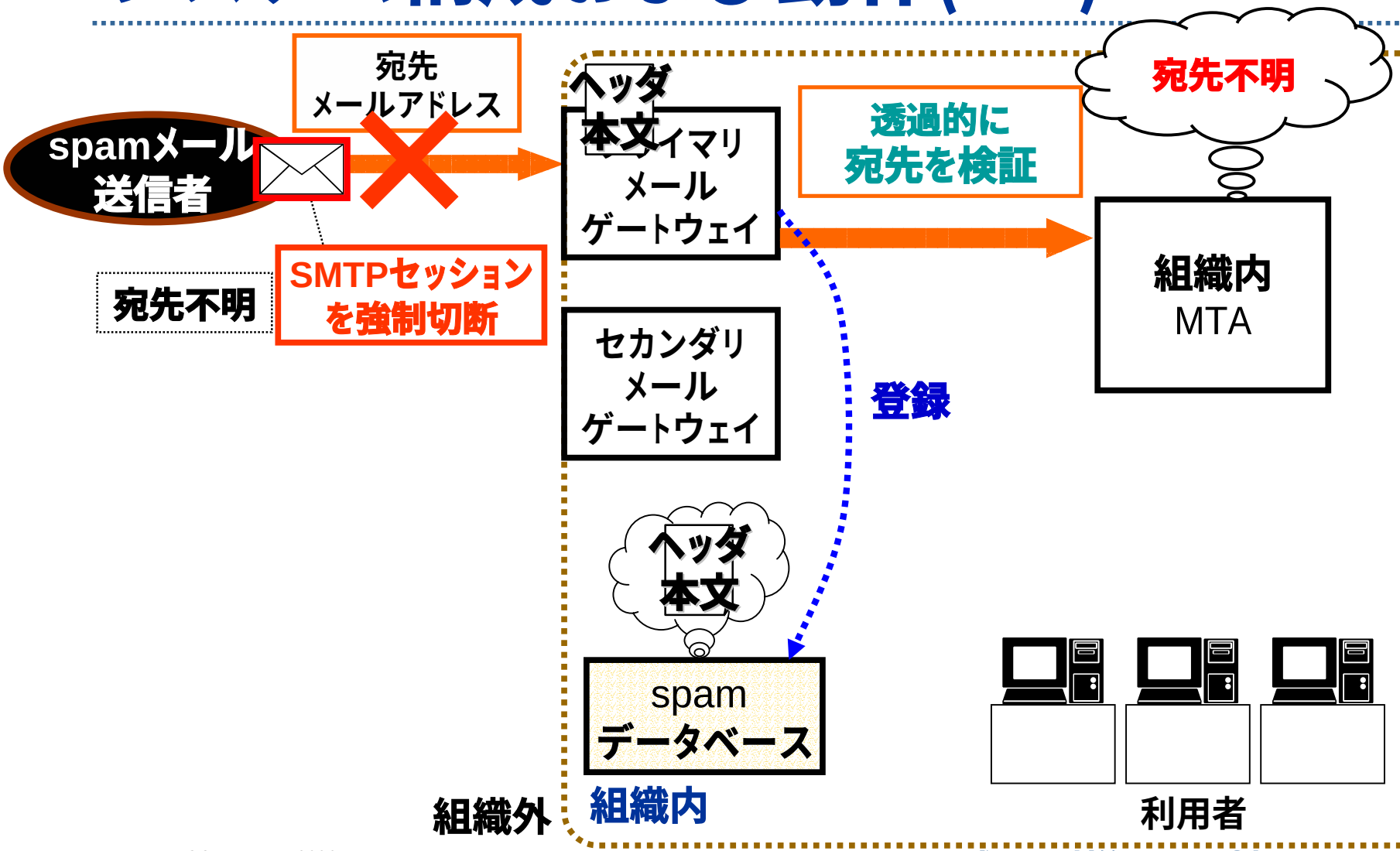
通常メールの受信における遅延発生を防止

IPアドレスに基づかない再送判定が可能

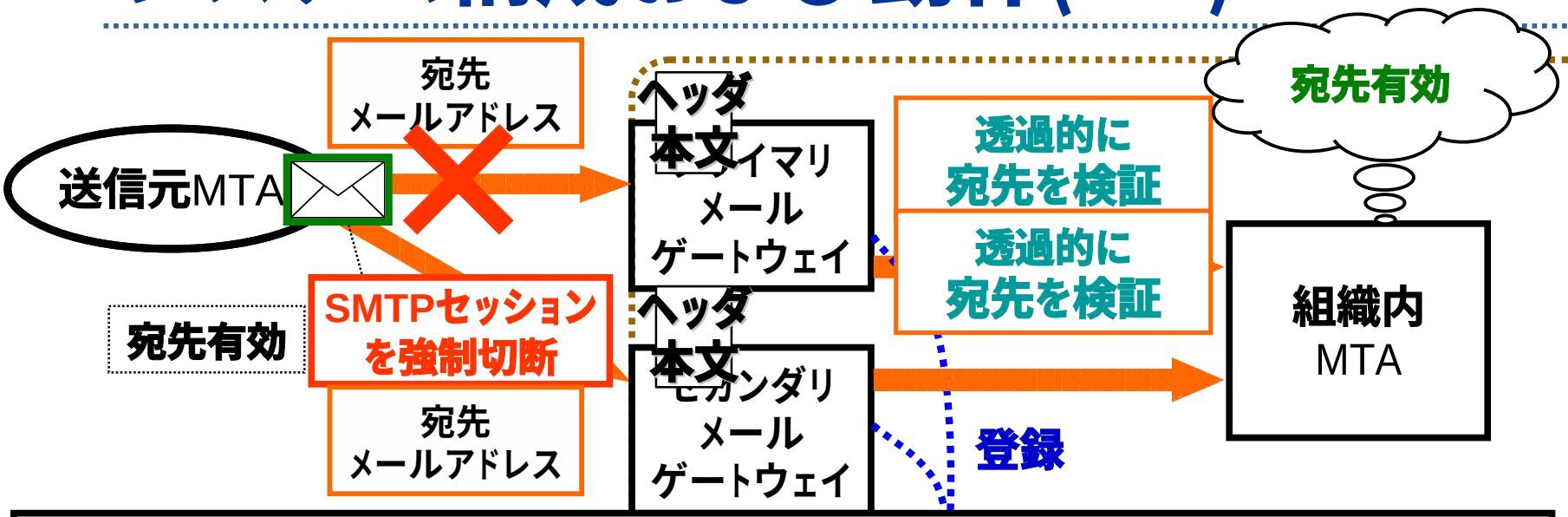


組織外　組織内　利用者

システム構成および動作(2-1)



システム構成および動作(2-2)



宛先不明and/or未再送メールを自動登録可能

- **分散協調spamデータベースの性能向上**

目次

- はじめに
- 従来のspamメール対策方法とその問題点
- SMTPセッションの強制切断によるspamメール対策手法
- **試作システムの実装と評価**
- まとめ

試作システムの実装

- プラットフォーム
  - OS: FreeBSD
  - MTA: sendmail
  - spamデータベース：DCC
- メールゲートウェイでの処理
  - ヘッダおよび本文受信後にRSTパケット送出
  - 再送判定=(MAIL From, RCPT To, Message-ID)
  - 宛先不明かつ未再送メールをDCCに登録
    - 宛先不明・・・利用者の承諾が不要
    - 未再送・・・・・メーリングリストを登録対象から除外

動作試験

1. MTAから通常メールを送信（再送あり）
   - 初回送信の失敗後，セカンダリメールゲートウェイへ瞬時に再送されることを確認
2. 送信失敗後、異なるMTAから同一メールを再送
   - メールが正常に受信されることを確認
3. MTAから宛先不明メールを送信（再送なし）
   - spamデータベースへの登録を確認
4. MTAから宛先不明メールを送信（再送あり）
   - spamデータベースからの登録抹消を確認

性能評価(1)

- 実験環境
  - 近々廃止予定の学内ドメイン宛メールを試作システムで処理
- 対象メール
  - 大半が宛先不明 (卒業生宛を含む)
  - 宛先有効メールもspamの可能性大
  - 卒業生宛のメーリングリストも存在
- 実験期間
  - 「2006年1月29日 ～ 2月5日」の7日間

性能評価(2)

- 実験結果(1)

全メール	54,719通
tempfailingにより排除できたメール	44,303通
tempfailingにより排除できなかったメール (再送されてきたメール)	10,416通

- 81%のメールをtempfailingにより排除
  - 従来の技法と同等の数値
  - ➡ tempfailingの有効性を確認

性能評価(3)

- 実験結果(2)

tempfailingにより排除できなかったメール (再送されてきたメール)	10,416通
宛先不明メールの登録で新たに 排除できたメール	2,180通

- 再送メールの21%を
  分散協調spamデータベースで排除
  - 7日間の宛先不明メールのみ登録
    - さらなる改善が見込まれる

目次

- はじめに
- 従来のspamメール対策方法とその問題点
- SMTPセッションの強制切断によるspamメール対策手法
- 試作システムの実装と評価
- **まとめ**

まとめと今後の課題

- SMTPセッションの強制切断
  - ヘッダや本文を事前に取得可能
  - ヘッダ情報に基づく再送判定が可能
  - 分散協調spamデータベースに大量かつ早期に登録可能
- 今後の課題
  - 長期の実運用を通じての性能評価