



Forschungszentrum **L3S**  
Research Center

# MailRank: Using Ranking for Spam Detection

*Paul-Alexandru Chirita, Jörg Diederich, Wolfgang Nejdl*  
*L3S Research Center*  
*University of Hanover, Germany*  
*<http://www.l3s.de/~diederich>*  
*{chirita|diederich|nejdl}@l3s.de*

# Outline



Forschungszentrum **L3S**  
Research Center

- Problem of current spam detection schemes:
  - Maintenance: Adapting to spammer strategies
- Related Work: Spam detection schemes
- MailRank:
  - Construct email network from known email addresses
  - Use PageRank to compute reputation values
- Evaluation
  - Sparse input data
  - Attack resilience
- Future Work



- Enemy number one:



emails

- Spam detection

- Main problems of existing approaches:

- Maintenance

- Bootstrapping the system:
  - Initial input data (rules, dynamic keywords...)
- Running the system:
  - Adapting to spammers trying to circumvent the system

- Attack resistance

- Residual error rate

- Some remaining false positives and false negatives
  - I.e. spam not being recognized / non-spam being classified as spam

# Spam Detection I



Forschungszentrum **L3S**  
Research Center

- Content-based schemes: Analyze the subject / the body of an email message
  - Static keywords (taken from SpamAssassin):
    - 'Viagra', 'online pharmacy', 'click below', 'no cost', 'lowest price', '100 percent guaranteed', 'university diplomas', 'mortgage', 'millions of dollars', 'mas informacion', 'haga click aqui', number in the email 'subject' header
  - Cf. [http://spamassassin.apache.org/tests\\_3\\_0\\_x.html](http://spamassassin.apache.org/tests_3_0_x.html)
  - Learn more keywords dynamically
    - Bayesian network: Use training set of spam and non-spam
  - Email body structure:
    - HTML emails, extensive use of HTML markup,...
    - Numeric id in URL: 192.168.0.1:8080/login
- Example: SpamAssassin
  - Rather a framework (content-based plus more...)

# Spam Detection II



Forschungszentrum **L3S**  
Research Center

- **Pros:**
  - Content-based schemes can work pretty good
    - After an initial learning phase
- **Cons:**
  - Become a Spammer yourself...
    - Use html in emails
    - Work in a pharmacy (use `viagra`,...)
    - Have `true` business contacts: Amazon order confirmation classified as spam...
  - False positives (~0.6%)
  - **Permanent fight between adapting rules and new Spamming strategies**
    - Not too much German spam yet → no `German` rules currently
  - Still some false negatives (~6%; ~1% for well-trained Bayesian network)
    - Main problem: quality of the rules & permanent adaptation
  - Global rules (i.e., limited personalization)



- Header-based schemes: Analyze the SMTP protocol fields of the email message
  - Blacklists: Collect IP-addresses of spammers
  - Whitelists: Collect Email addresses of known non-spammers
  - Autowhitelists (in SpamAssassin):
    - Sum up the scores of previous emails (determined by a content-based approach) for one sender / IP address pair
    - Accept a single Spam message from a known sender having sent many non-spam messages before
- Pro:
  - Whitelists: Good way to protect people from your social network
  - Autowhitelist: Nice to allow your friends writing about 'viagra'
- Cons:
  - Maintenance (Cold start problem): Build up the {black|white}list
  - Autowhitelist: Mobile people with changing IP addresses...
  - Global blacklists: Difficult to maintain, easy to attack

# Spam Detection IV



Forschungszentrum **L3S**  
Research Center

- Sender authentication
  - Ensure that From: email addresses cannot be forged anymore
  - Add entry to DNS record for designated mail servers
    - SenderID (Microsoft), SPF (Pobox), DomainKeys (Yahoo)
- Pro:
  - Very important even for implementation of other approaches
    - E.g. for whitelists
- Cons:
  - Only a first step (Ex: Authenticated spammer on XX island)
  - Some inconveniences, e.g. for mobile workers



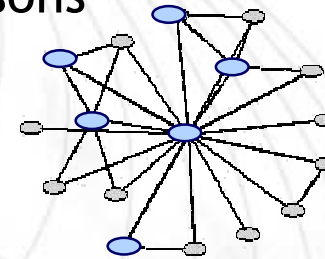
- Problem of current spam detection schemes:
  - Maintenance: Adapting to spammer strategies
- Related Work: Spam detection schemes
  - ➔ **Not sufficient**
- MailRank:
  - Construct email network from known email addresses
  - Use PageRank to compute reputation values
- Evaluation
  - Sparse input data
  - Attack resilience
- Future Work

# MailRank: Assumptions



Forschungszentrum **L3S**  
Research Center

- People exchange emails with a limited number of persons
  - Example: The social (email) network of 5 persons
    - Connected
    - Share same contacts
    - Only an excerpt...
- Emails from within the social network of a person are no spam
  - With a high probability
  - Assuming the sender address of an email cannot be forged
    - Sender authentication approach assumed to be in place



# MailRank in a Nutshell



Forschungszentrum **L3S**  
Research Center

- Collect data about trusted email addresses
  - For privacy reasons: only hash-values
- Build a global email network
- Get information about locally unknown email addresses
  - But globally known by at least one other MailRank user
- Apply reputation algorithm: PageRank
  - Basic idea: Sending emails to someone = vote for her
    - No voting for spammer email addresses usually...
  - The more votes an email address receives, the higher the reputation score
- Spam detection: Identify good email addresses and bad ones
  - Emails from bad ones: spam; else: no spam
    - Blacklist server only know about the bad guys

# Advantages:



# MailRank

Forschungszentrum **L3S**  
Research Center

- Shorter individual cold-start phase
  - Only one global cold-start
- Partial participation
  - No need for everybody to participate
    - Power-law nature of email networks
- Stability of social networks
- High attack resilience
  - Similar to PageRank
- Can counter SPIT
  - Spam over Internet Telephony
- Personalization
  - Personalized PageRank schemes

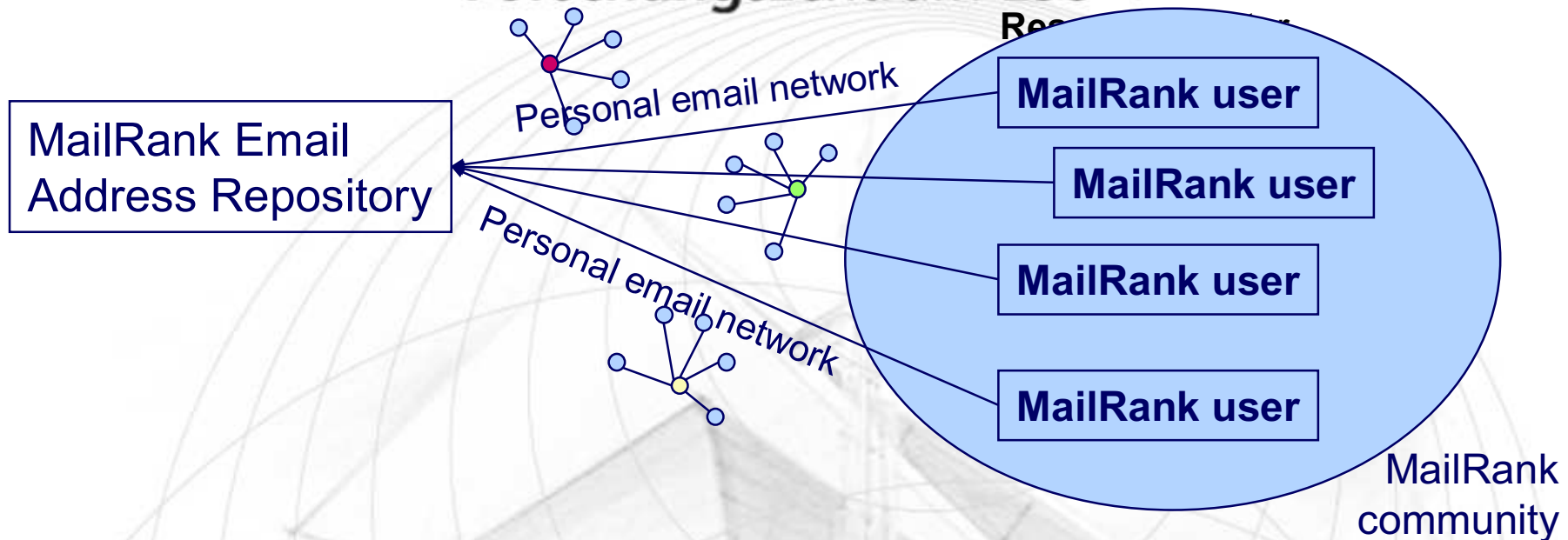
# Main Procedure: Step 1



# MailRank

Forschungszentrum L3S

Re...



## 1. Populate the MailRank Repository

- All MailRank community members provide their personal email networks to the MailRank Email Address Repository
  - Use hash values to hide the social network of users
- Different sources of input data possible: Automatic extraction from
  - Sent-mail folders (initialization phase)
  - Ongoing email communication, log files (update phase)
- **High quality** since ,manual'

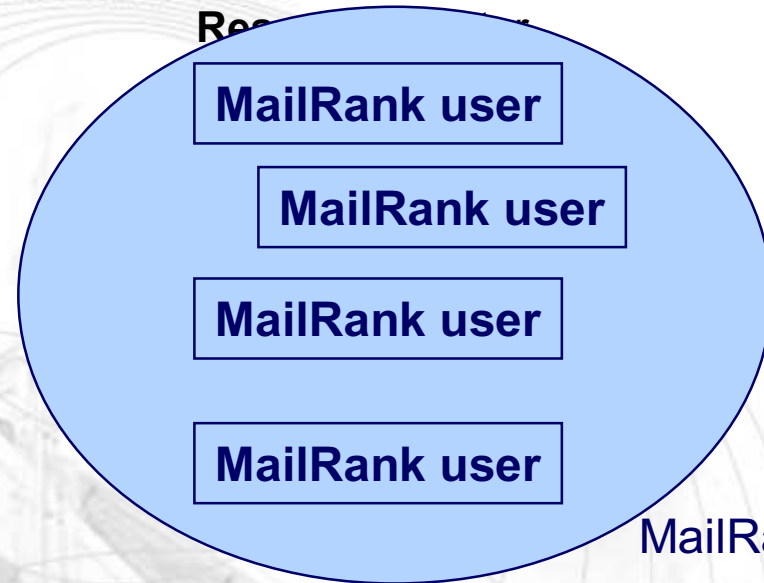
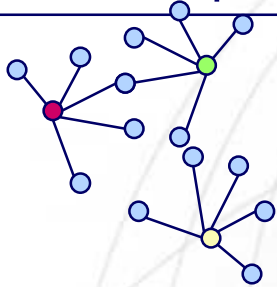
# Main Procedure: Step 2



# MailRank

Forschungszentrum L3S  
Re...

MailRank Email  
Address Repository



MailRank  
community

1. Populate the MailRank Repository
2. Connect the personal email networks to a global one
  - Not necessarily a connected graph...
  - [this is a feasibility study only: whole graph cannot be kept on a single server...]

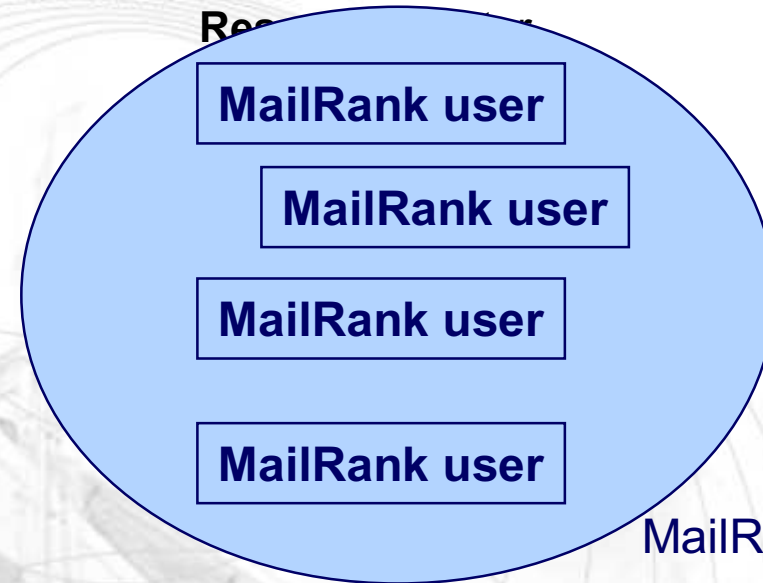
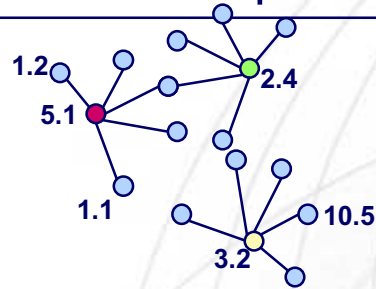
# Main Procedure: Step 3



# MailRank

Forschungszentrum L3S  
Re...

MailRank Email  
Address Repository



MailRank  
community

1. Populate the MailRank Repository
2. Connect the personal email networks to a global one
3. Use PageRank to compute score for each known email address
  1. Determine a set of email addresses with a very high reputation
    - Called 'biasing set'
    - Manually / automatically / semi-automatically
  2. Run PageRank second time biasing only on the previously determined set
    - Assign initial ranks only to members of biasing set

# Main Procedure: Step 4

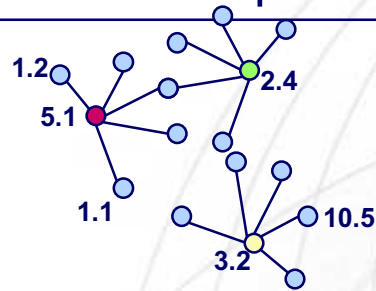


# MailRank

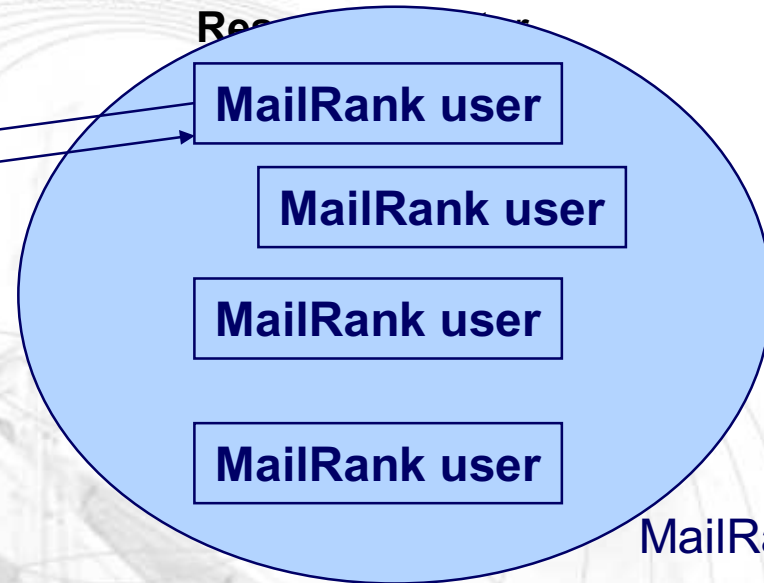
Forschungszentrum L3S

Re...

MailRank Email Address Repository



Joe@foo.com? ●  
Joe@foo.com: 2.4



MailRank community

1. Populate the MailRank Repository
2. Connect the personal email networks to a global one
3. Use PageRank algorithm
4. On arrival of an email from an (unknown) sender:
  - Retrieve the score of the sender email address
  - Validate known addresses once in a while (the scores might change)

## Step 4: Query the Server



MailRank

Forschungszentrum L3S

Research Center

Basically: Ask MailRank about locally unknown email addresses

- Result: Score for the email address
  - Identify 'good guys' with high probability
  - Cannot identify 'bad guys' explicitly
    - Need one threshold
      - If = 0: compute transitive closure from the biasing set





## Pros:

- High attack resistance (inherited from PageRank)
- Partial participation
  - No need for everybody to participate
    - Power-law nature of email networks
- Personalization: put your personal whitelist into the biasing set
  - Requires higher computational effort...

## Cons:

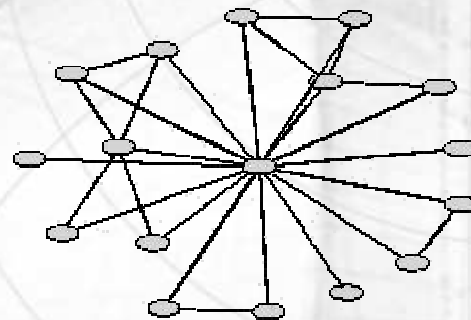
- Requires sender authentication to be in place
- Cold-start problem on a global scale
- Mailing lists / send-only email addresses
- New users (have to have a contact to other MailRank users)
- Virus-Attacks: Make good people voting for spammers

# Evaluation: Simulation Model



Forschungszentrum **L3S**  
Research Center

- Model email network as power-law graph
  - With exponential cut-off:
    - Maximal number of outgoing Links: 1500
      - The number of people you send email to is limited...
    - Minimal number of incoming Links: 5
      - Typically you have more than one 'friend'



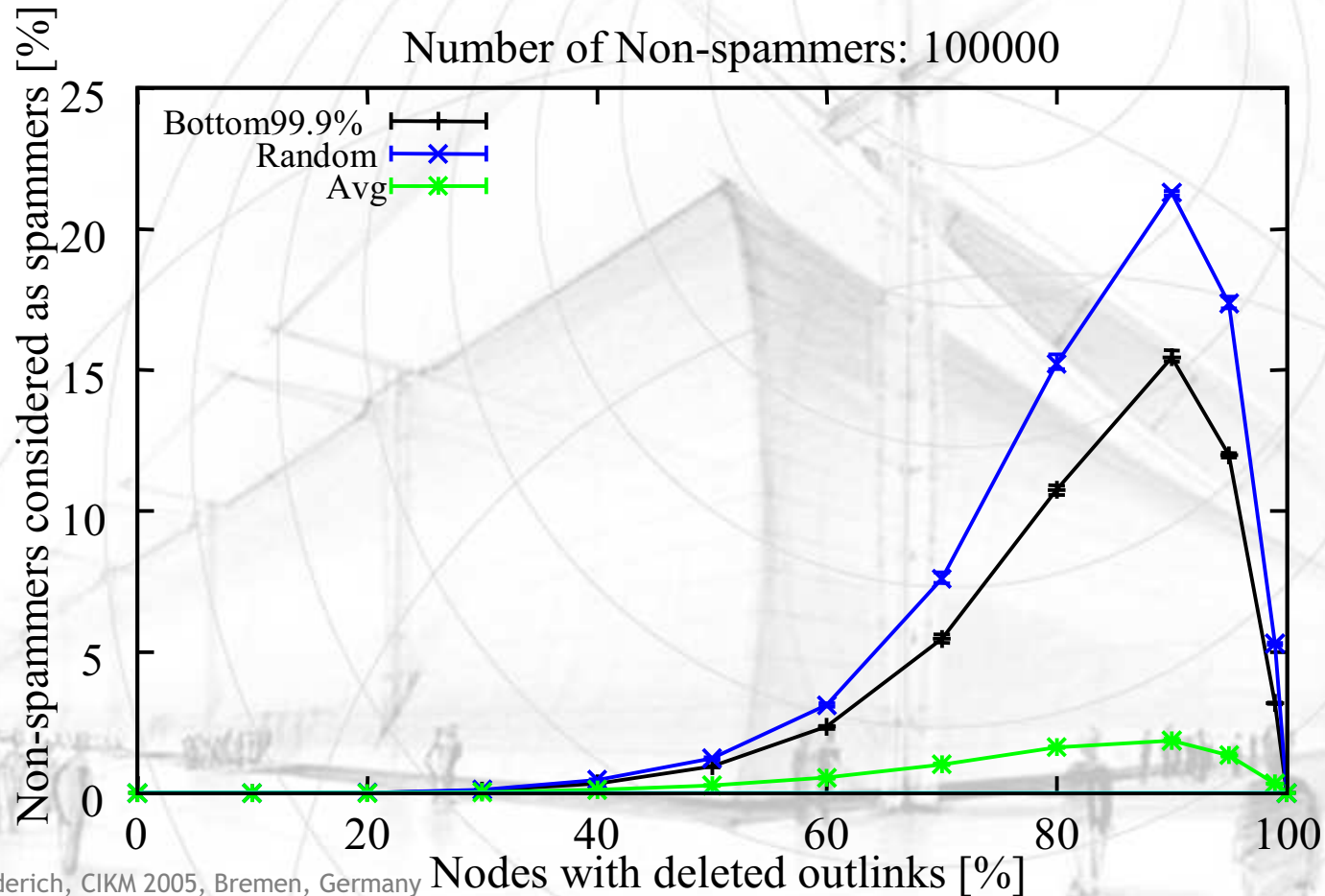
# Evaluation



Forschungszentrum L3S

Research Center

- All spammers successfully recognized
- Works well even in case of sparse email networks
  - I.e. only few people participate

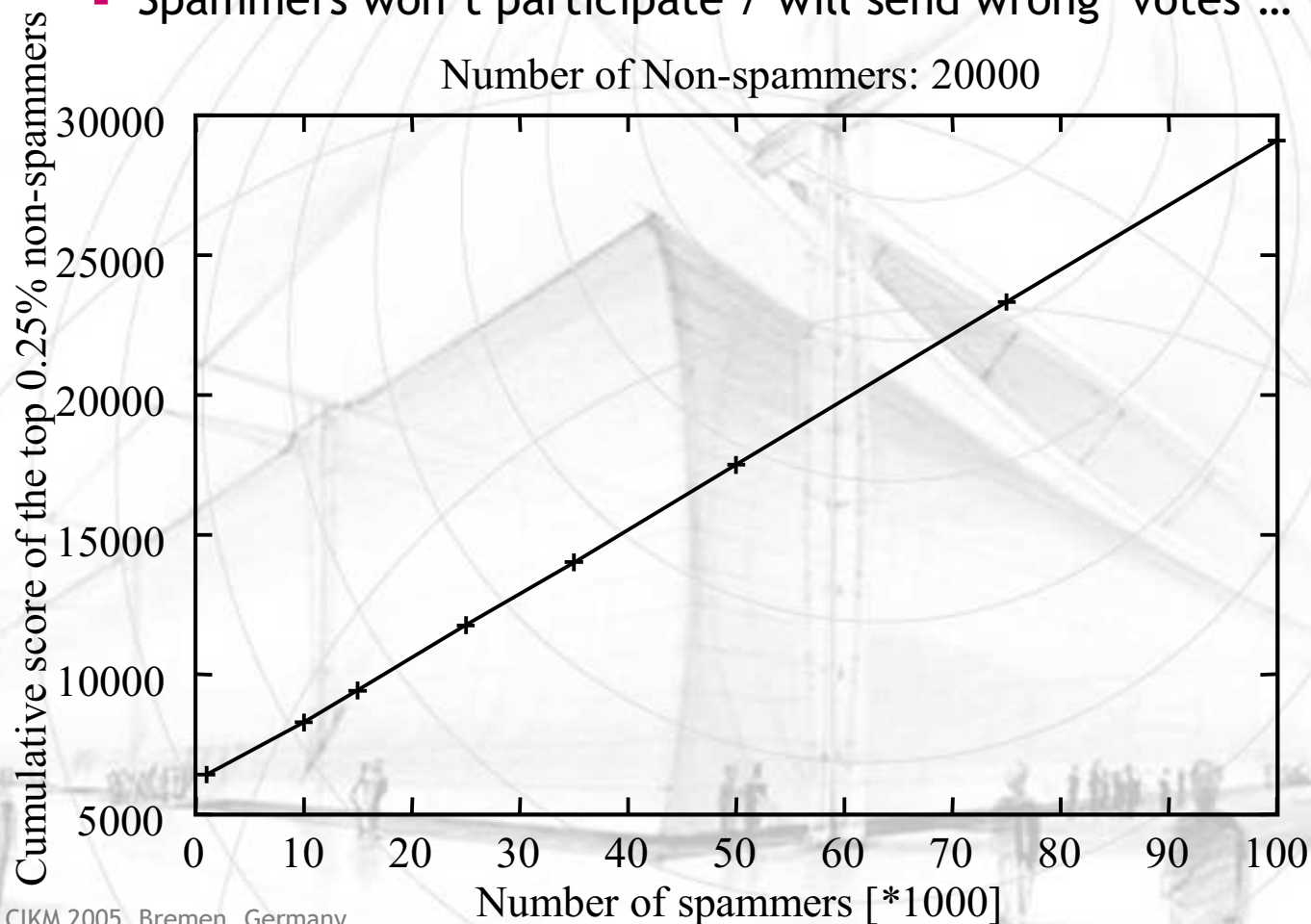


# Evaluation: Spammer's influence



Forschungszentrum **L3S**  
Research Center

- Nice result: the more spammer participate (=vote), the better
  - But need to get the data from mail servers
    - Spammers won't participate / will send wrong 'votes'...

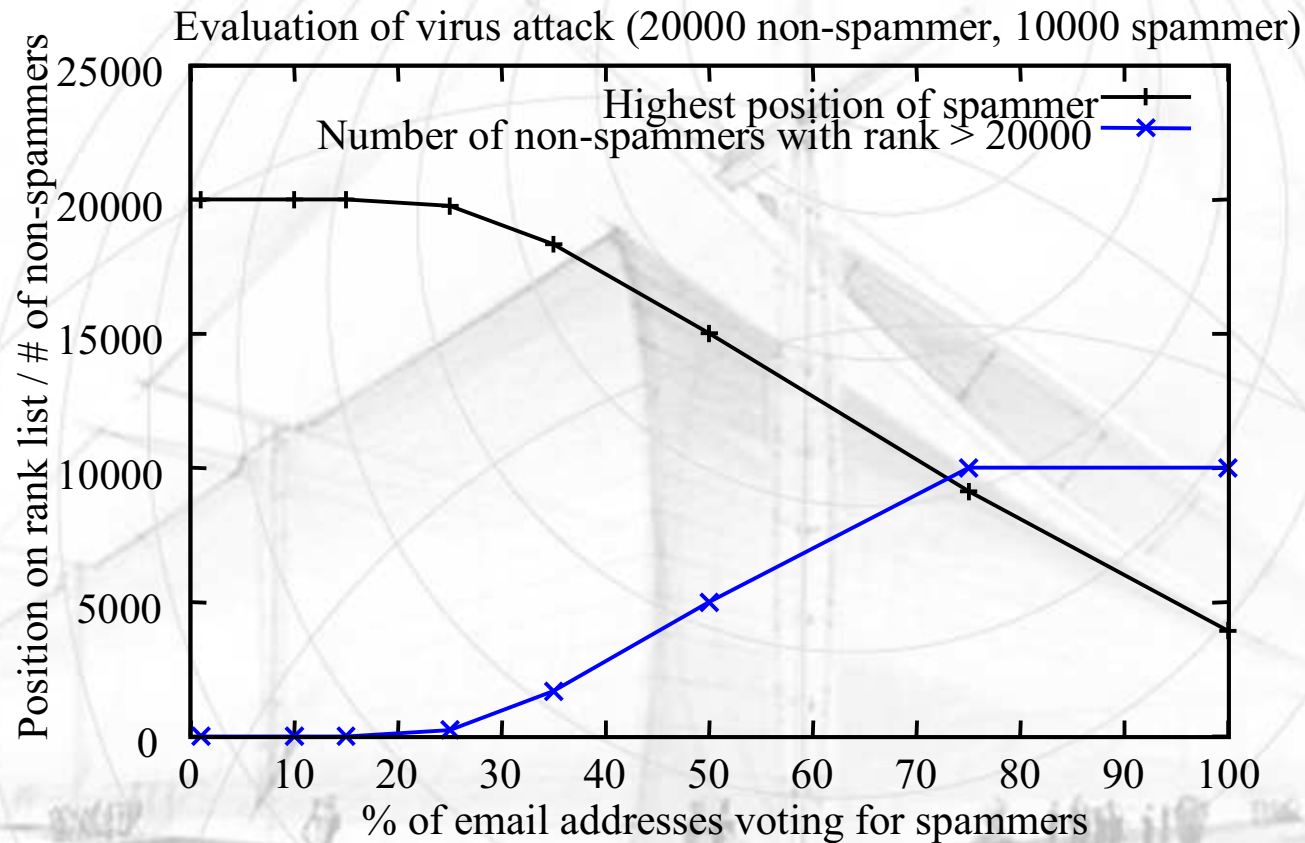


# Evaluation: Virus infection



Forschungszentrum L3S  
Research Center

- Can withstand even a certain number of infected hosts
  - Viruses, worms



■ But only under the assumption that top 10% immune!

# Summary



Forschungszentrum **L3S**  
Research Center

- MailRank:
    - Connect personal email networks and apply PageRank to detect spammer
    - Use high-quality input data (e.g. from sent-mail folder, log files)
      - 'manually' created → high quality
        - but no additional efforts
- Works in case of sparse participation & difficult to attack

# Future Work



Forschungszentrum **L3S**  
Research Center

- Input data:
  - What if people answer to spam?
    - No automatically generated emails (vacation) into sent-mail folder...
    - Similar effect as virus infection
- Threshold value to identify spammer addresses
  - Depends on personal taste...
- Distributed version of the email repository
  - Need distributed computation of PageRank
- More anonymization efforts?



**Thank you for your attention!**

**Questions?**

**I have two...**

# Straw poll



Forschungszentrum **L3S**  
Research Center

- Who would be willing to disclose his/her (anonymized) address book if it would help to counter spam?
- Who would do so only if the address book data would be merged with other data before being sent to the MailRank server (e.g., merging within an organization, effectively merging all email-addresses within an organization to a single node in the graph)?