LENS:

LEveraging social **N**etworking and trust to prevent **S**pam transmission

Sufian Hameed (Uni Göttingen)
Xiaoming Fu (Uni Göttingen)
Pan Hui (Deutsche Telekom Labs)
Nishanth Sashtry (Uni Cambridge)

computer NETWORKS

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Introduction

Spam

- An open problem largely outnumbering legitimate emails
- In 2010, 89% of the emails were spams (262 billion spams daily) [1]
- Estimate cost of $130 billion in 2009 [2]
- Projected cost of $338 billion by 2013 [3]

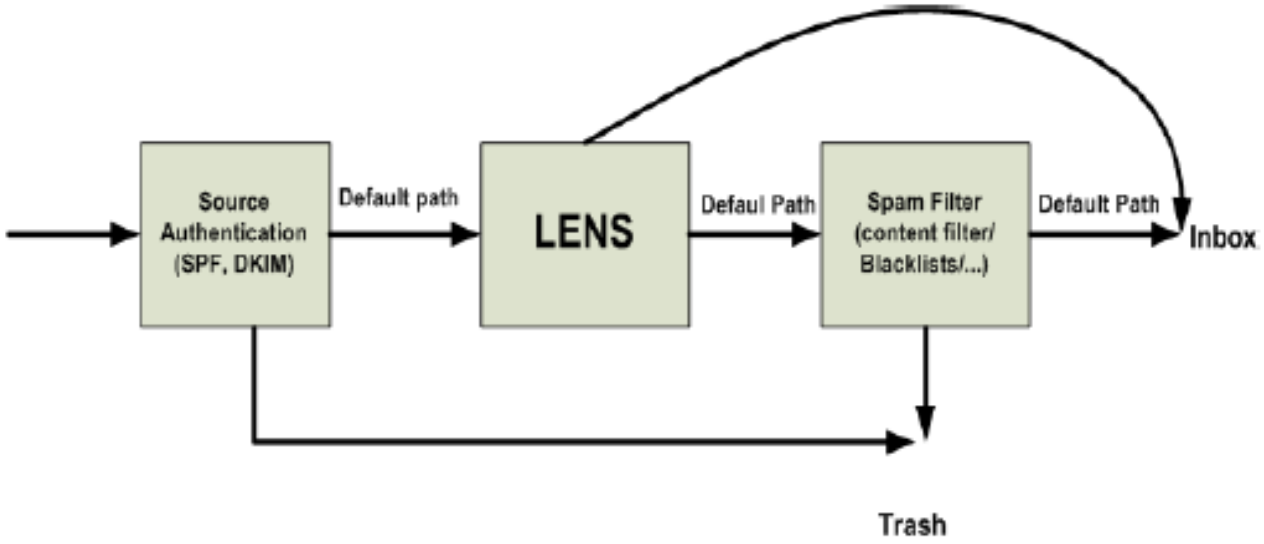
Common State-of-the-Art Strategies

- Filter spam at the recipient's edge.
- Content-based filtering has turned spam problem into false positive and false negative ones

Goal

Stop the Arms race: Prevent spam transmission during SMTP time and accept only ***legitimate email from legitimate users***

LENS



LENS leverages the social network of the recipient:

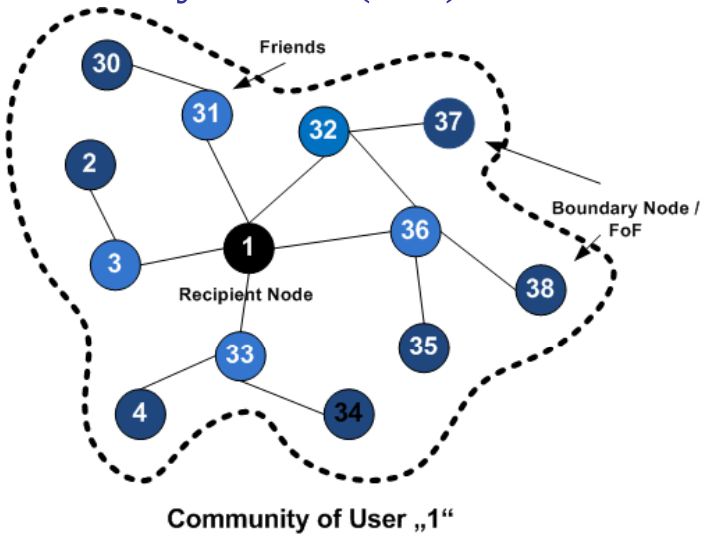
- Mitigates spam beyond recipient's social circles, by accepting only legitimate emails
- Filter at SMTP to prevent transmission at sender's edge
- Two types of communications:
  - Within recipient‘s community (social network)
  - Outside recipient‘s community (rest of the world)

Communication Within Community

- Emails within the community is delivered directly to the recipient.
- Community consists of two social hops
  - Friends and
  - Friends of Friends (FoF), also called boundary nodes (BN)

Community Formation

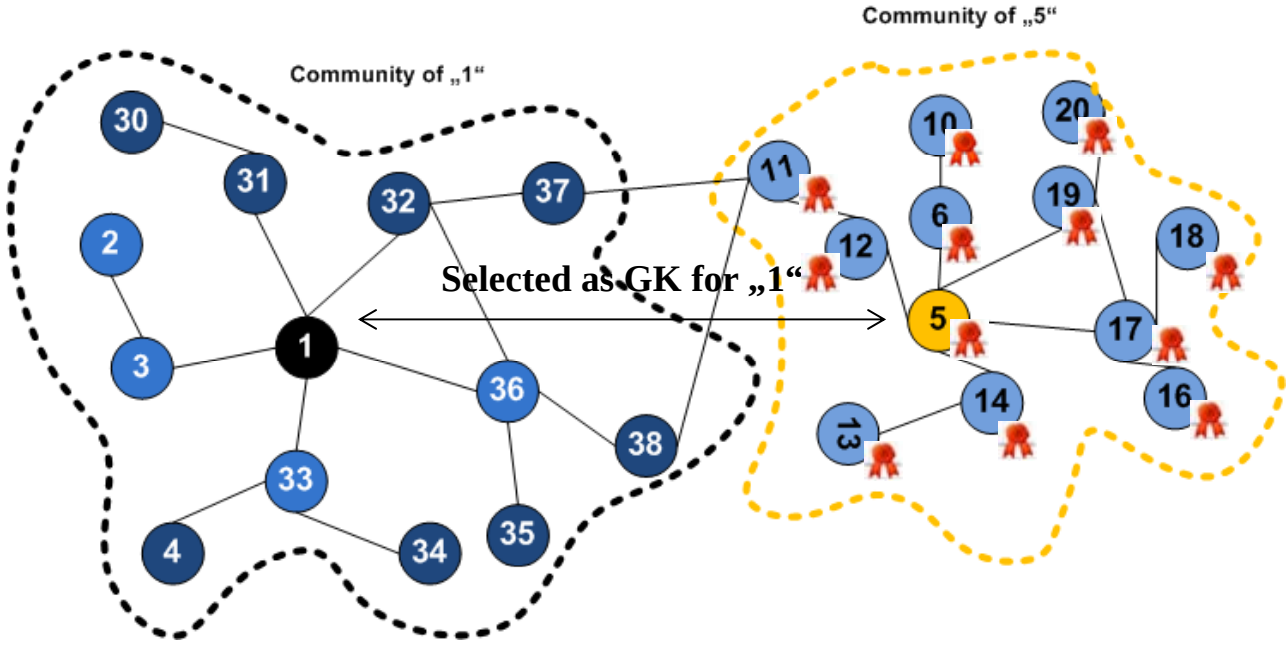
- A simple two step process
  - *Addition of friends*
  - *Addtion of FoF.*
- Process can be
  - ***Manual (User Involvement)***
  - *Automatic (Communication Pattern, Extraction from real Online Social Network*



- **Mail Server maintains the Community info of its users in CommList.**

Communication Outside the Community

- Mail Server selects Trusted/Legitimate users, called Gate Keepers (GKs) at various hop counts away for the recipient.
- Mail Server uses the GK to vouch for legitimate users outside the community of the recipient, by issuing un-forgeable vouchers.
- GK can only vouch for his immediate community.

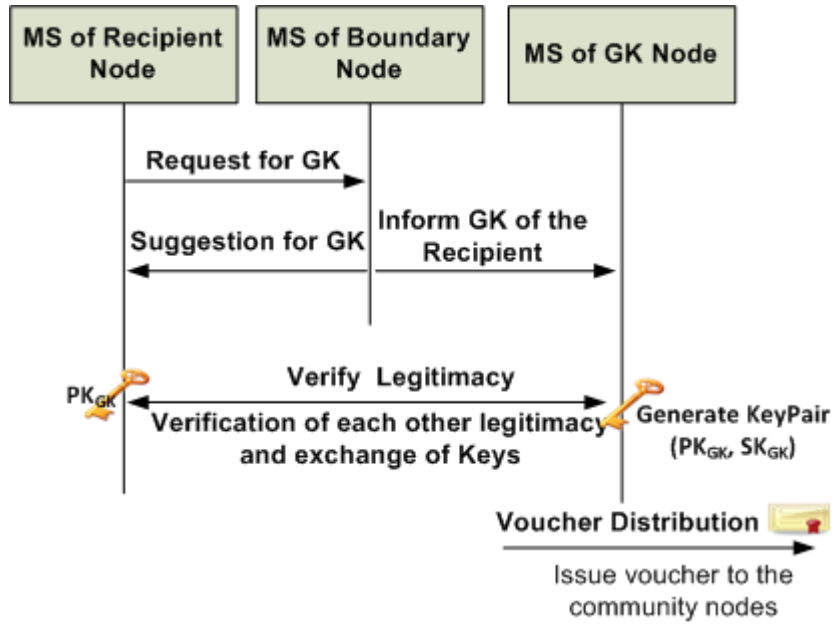


GK is a virtual entity; its selection and voucher distribution are system processes running on mail servers

GK Selection (1)

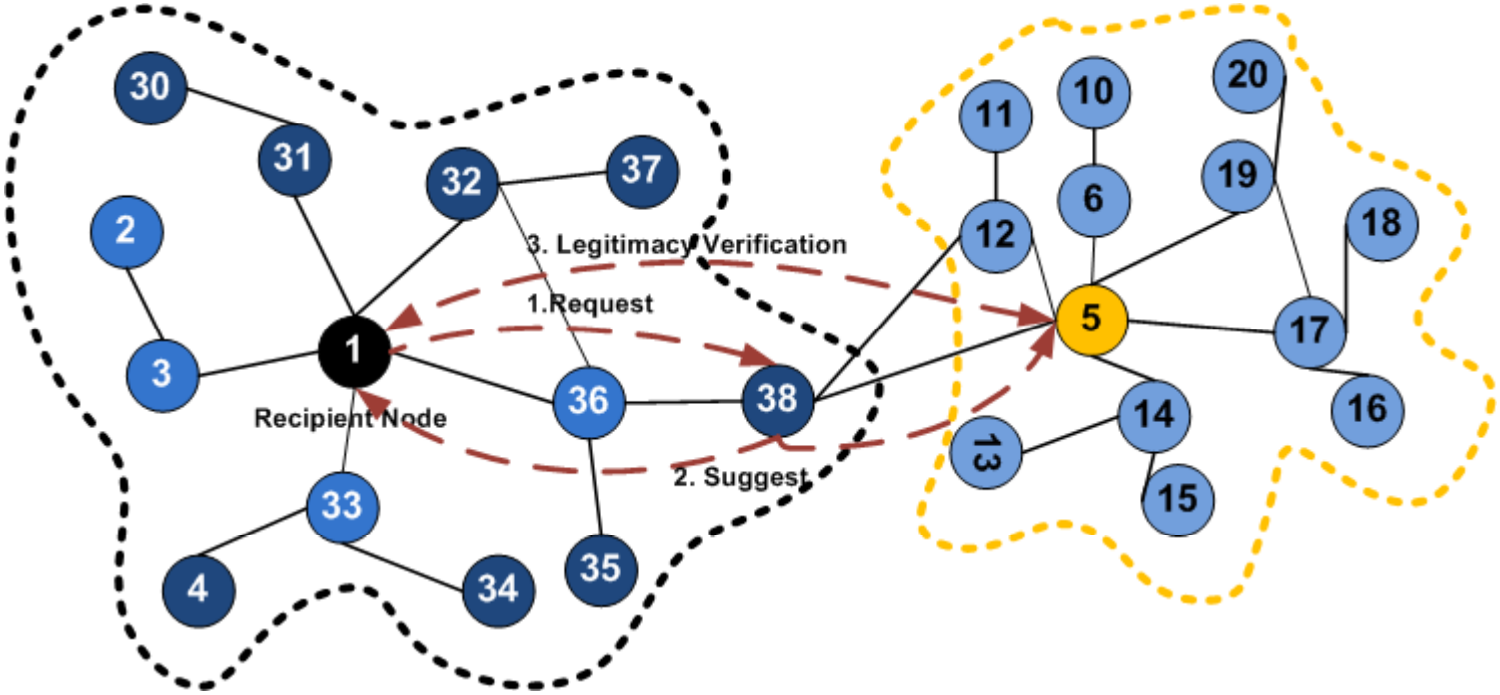
Stage 1 (Adjacent Communities)

- Run Transparently by MailServer
- Three Steps
  - Request
  - Suggestion
  - Legitimacy Verification



GK Selection (2)

Stage 1 (Adjacent Communities)

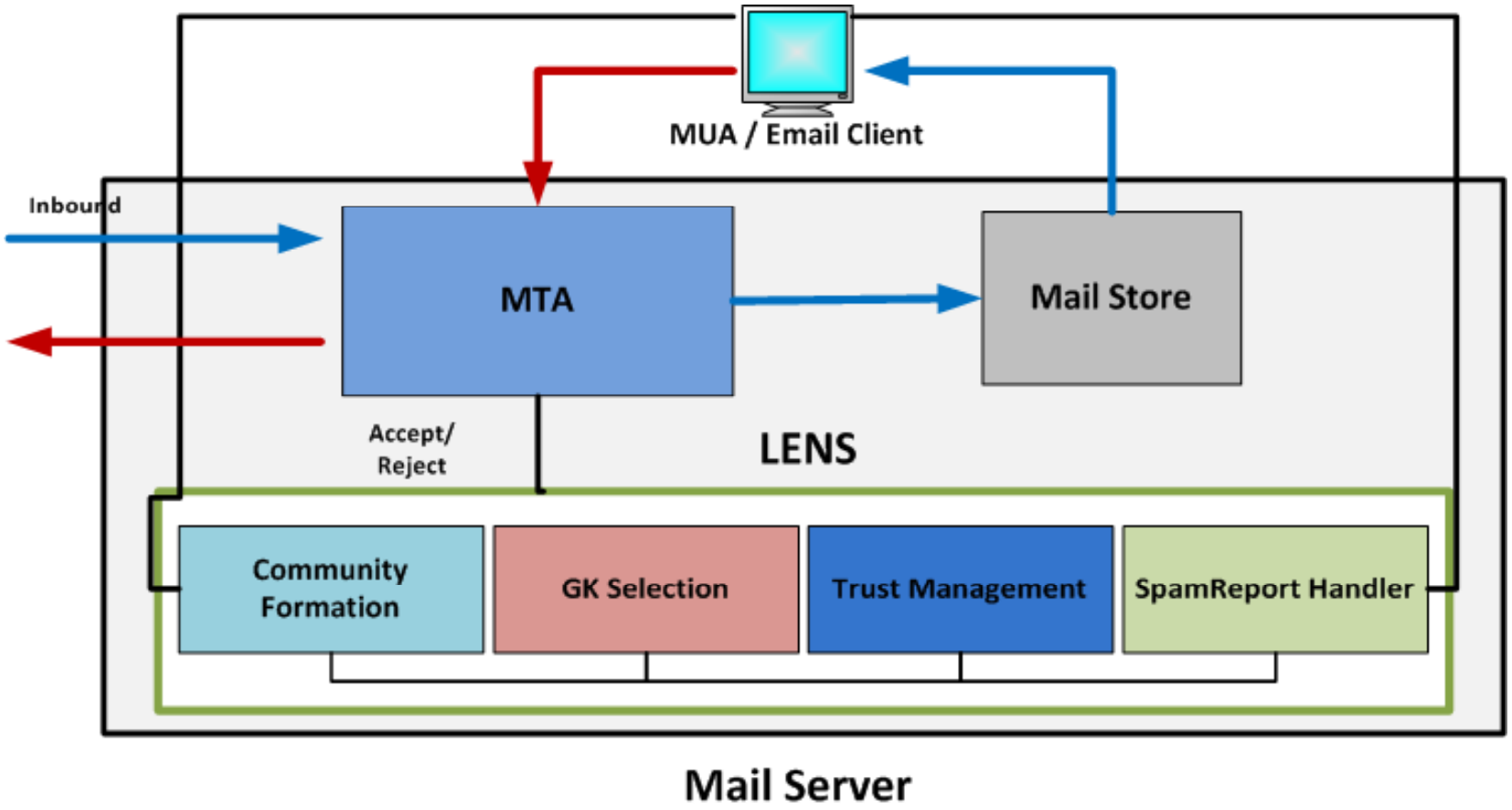


GK Legitimacy

Legitimacy Verification of GK is a 2 step

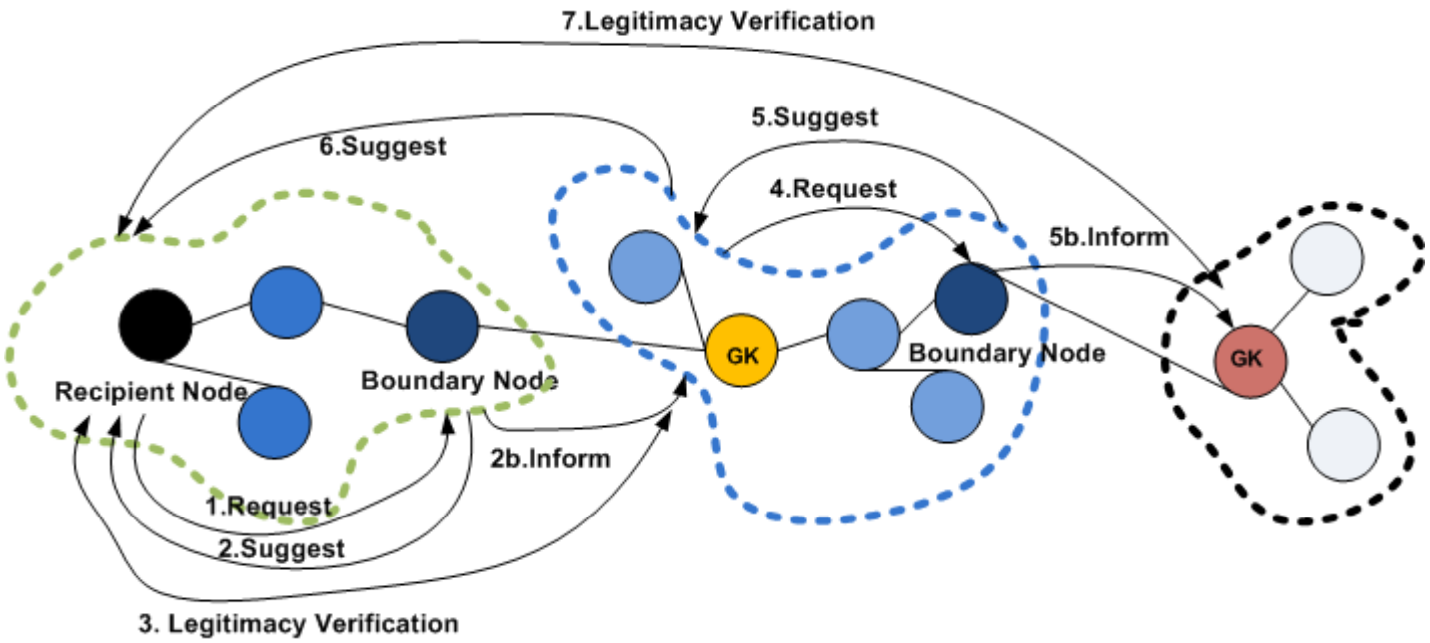
1. Legitimacy verificaton of the email service provicer
   - Identity verification using Certification Authority (done by all legitimate email provider, companies and universities)
   - Trust and Reputation measured over time
2. Legitimacy verification of a User (potential GK)
   - Based on the Trust Ratings of the user
   - Trust Rating is increased if a user is voted (receive emails) from other legitimate users

LENS Architecture



GK Selection (2)

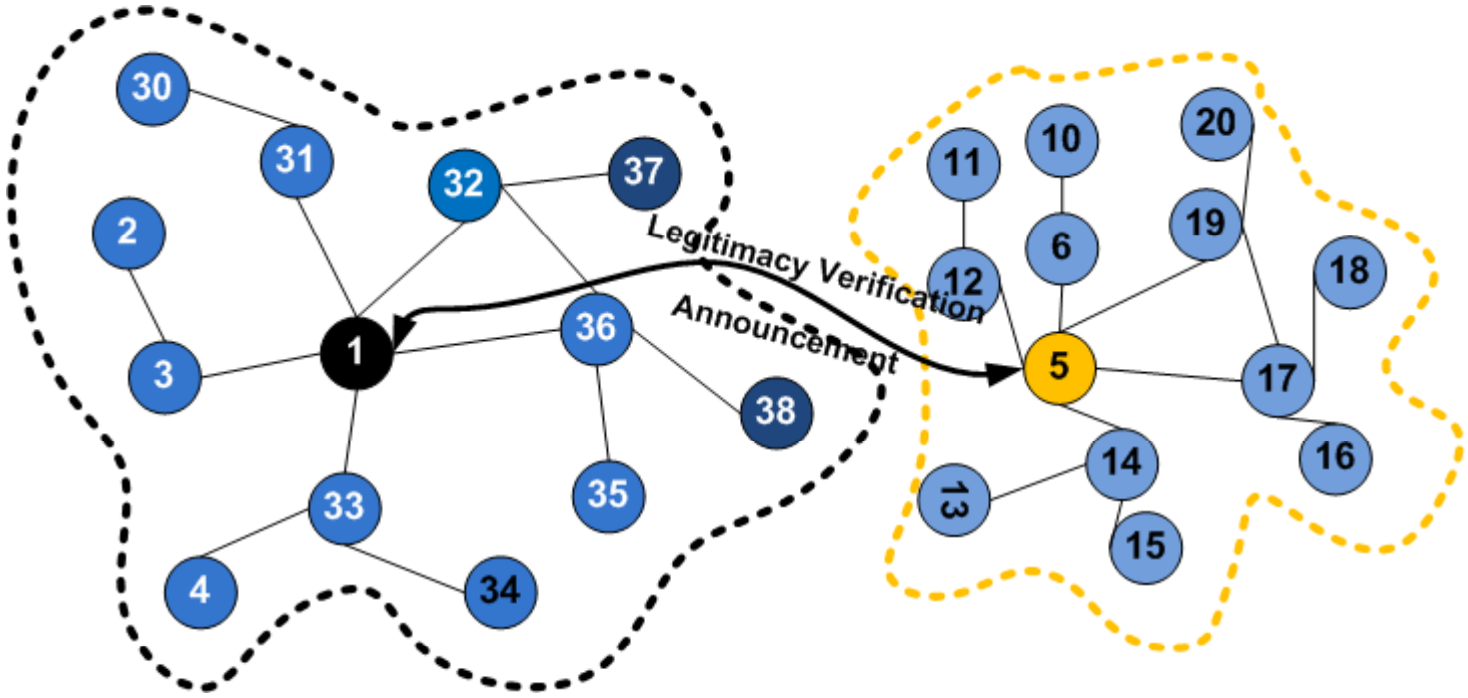
- Stage 2 (Beyond Ajacent Communities)



GK Selection (3)

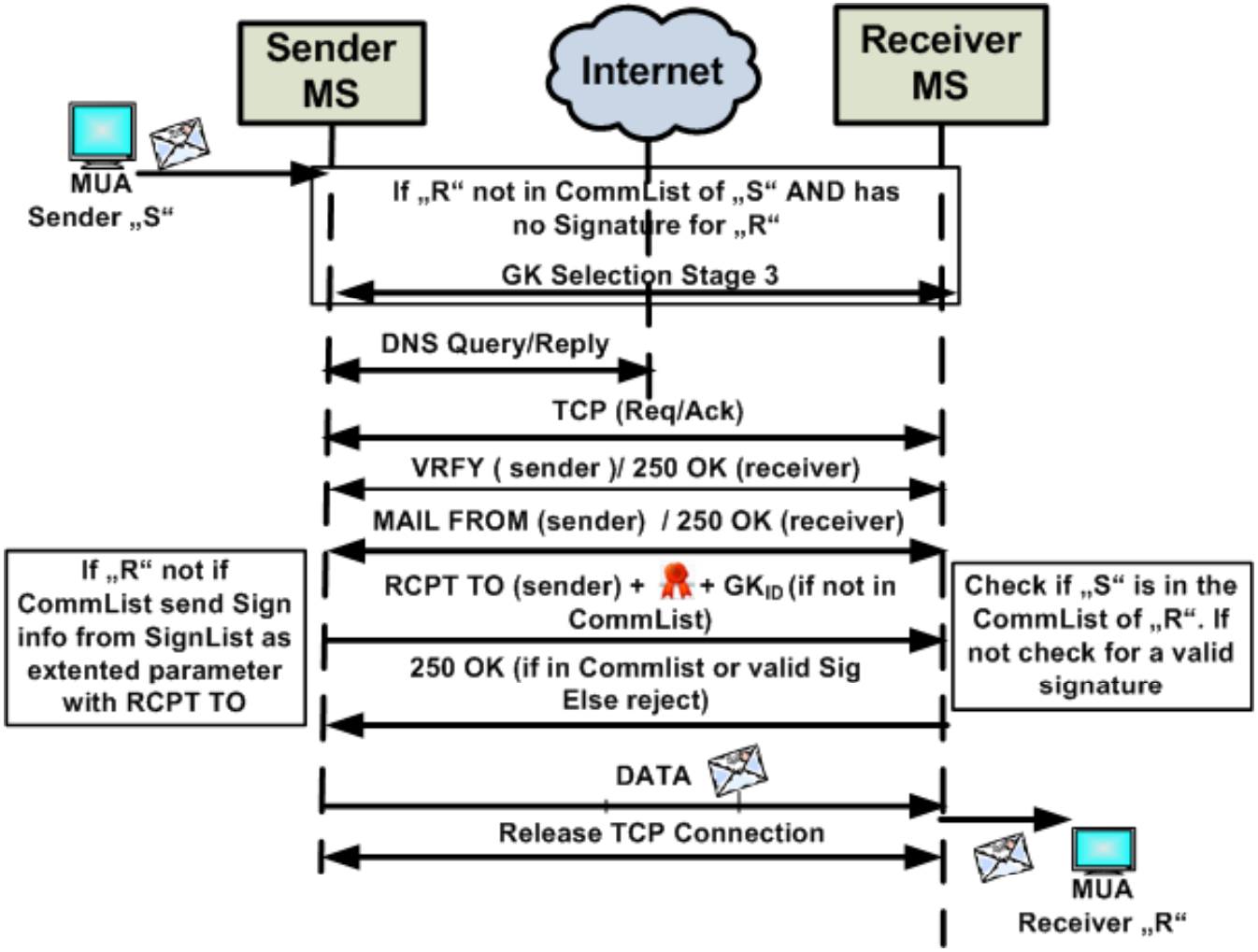
Stage 3 (New Communication)

- Two Steps
  - Announce
  - Legitimacy Verification



Apply sender rate limit if reputation and trust ratings of user is low

Email processing with LENS



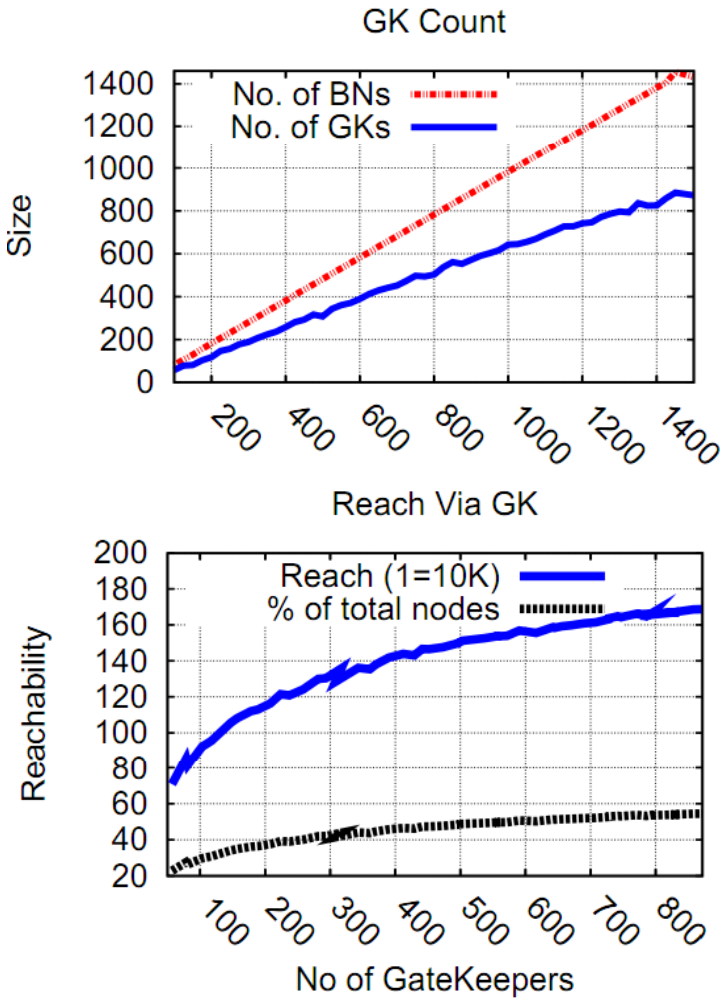
Evaluations

Evaluation metrics:

- # of GKs for receiving messages
- Reachablity of recipient via GK
- Computational complexity of email processing with LENS

Experiment on Facebook Dataset

- 3.1 M users, 23 M edges [3]
- Randomly select 4K users of community size 100-1500
- Number of GKs between 56-880
- Reachability between 710K - 1.7 million (23-55%)



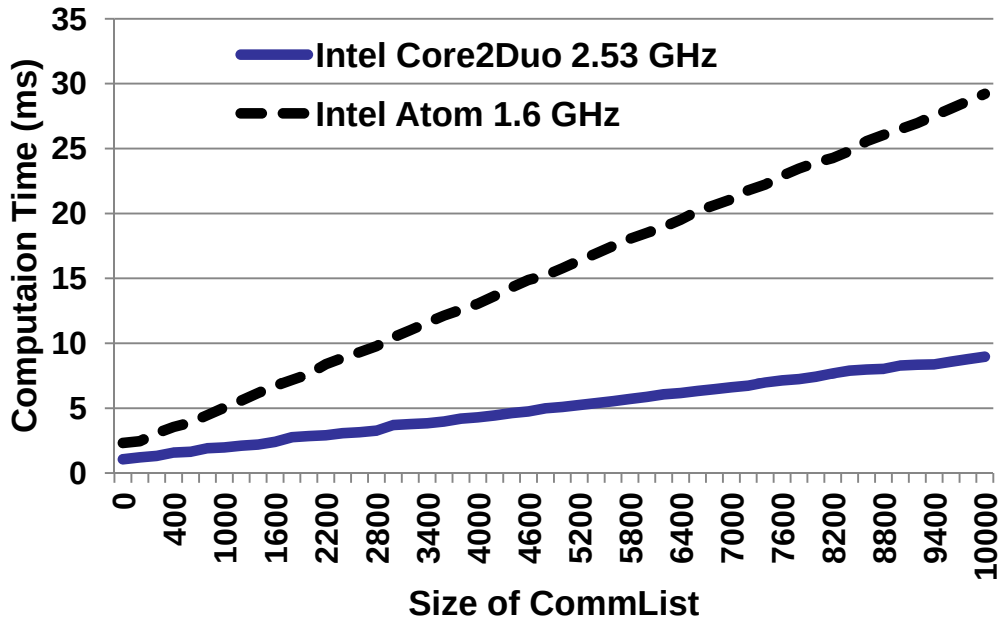
Reliable email delivery from millions of potential users is possible using GKs in the order of hundred.

Computation Complexity

- To analyze the computational complexity of email processing, we augment a standard mail processing system with LENS and measure its impact
- Used two MSes running Mail Avenger SMTP server on top of the Postfix MTA, connected via a local area network
- One of the MS is an intel atom 1.6 GHz with 1 GB ram and the other one is an intel core2duo 2.53 GHz with 4 GB ram
- We measure the effect of;
  - community size,
  - complexity of signature/voucher verification and
  - size of PkList on the email processing using LENS.

Note that in our experiments we only measured computational overhead (in terms of time) on the SMTP transaction (i.e. RCPT TO)

Effect of CommList



•On receiving the RCPT TO command, the recipient's MS checks whether the sender is in the recipient's Community

•With the community size of 10K,

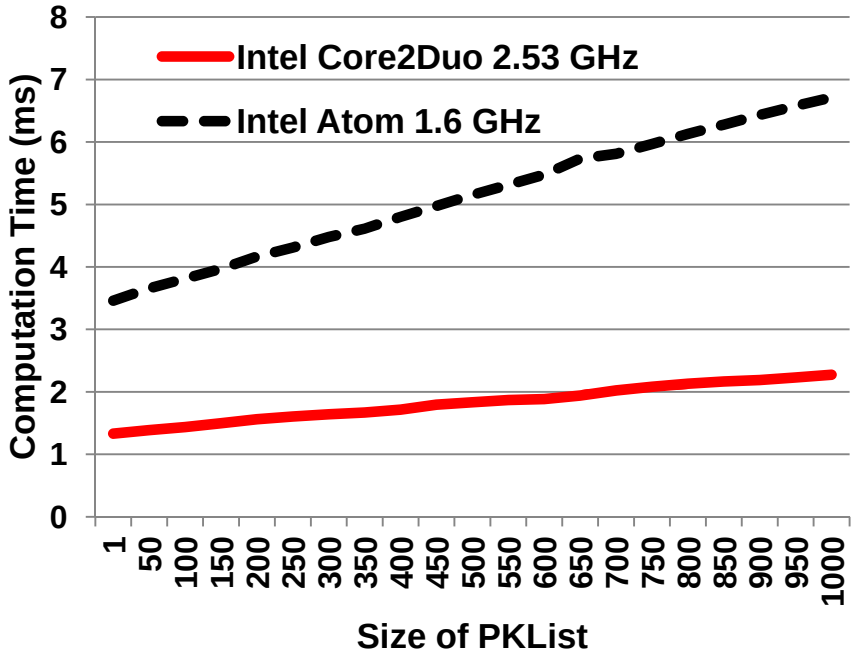
–overhead at one MS (intel atom 1.6 GHz) is ~30 ms

–overhead at the other MS (intel core2duo 2.53 GHz) is ~9 ms

Complexity of Voucher/Signature Verification

- The cost of checking signature has been has been so far considered substantial
- We used OpenSSL (www.openssl.org) speed measurement to study the computational cost of voucher generation/verification
- It takes approximately **0.13 ms to sign** and **0.07 ms to verify** a message using 1024 bits RSA and SHA-1 on low end MS (intel atom 1.6 GHz).

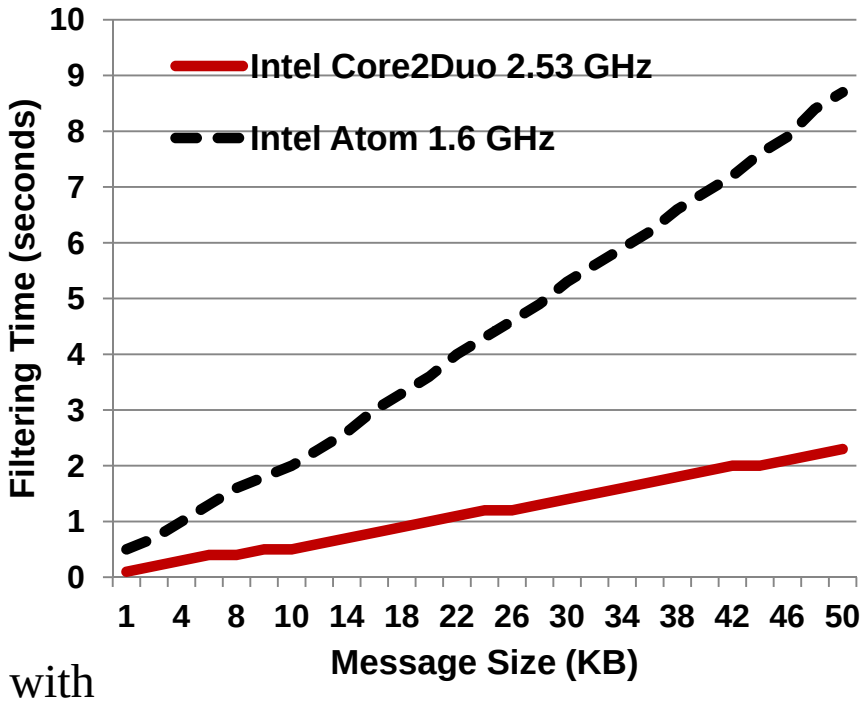
Effect of PKList



- Voucher verification, when integrated in LENS for email processing, depends on the size of PKList as well.
- The size of the PKList of a recipient is the same as the number of its GKs
- The computational overhead on both the MSs is only few milliseconds (ms).
  - With the PKList size of 1000 entries:
    * computation on one MS (intel atom 1.6 GHz): 6.7 ms
    * computation on the other MS (intel core2duo 2.53 GHz): 2.3 ms

Complexity of Content-Based Filtering

- Un-like LENS, Content-based filtering (SpamAssassin) is dependent on message size as well
- We varied the message size from 1 KB to 50 KB and analyzed the performance
- SpamAssassin takes 0.1 sec on intel core2duo 2.53 GHz machine and 0.5 sec on intel atom 1.6 GHz for a message size of 1 KB
- SpamAssassin is linear to the size of the message and the filtering time increases with an increase in the message size
- SpamAssassin takes 2.3 sec on intel core2duo 2.53 GHz machine and 8.7
- sec on intel atom 1.6 GHz for a message size of 50 KB



Conclusion

- Evaluations on OSN prove LENS to be scalable with large fraction of users. Reliable email delivery from millions of potential users is possible using GKs in the order of hundred.
- LENS is computationally efficient. LENS is fast in processing emails and it is 2-3 orders of magnitude faster than SpamAssassin
- Scales well with increasing size of the lists (CommList, VoucherList and PKList), and even on a low end processing machines the overheads are in couple of milliseconds

Open issues:

- Incremental deployment considerations
- Clear threat model
- Intensive system microbenchmaking
- Latency of GK selection in Stage 3

References

[1] http://royal.pingdom.com/2011/01/19/email-spam-statistics/

[2] Ferris Research, *http://ferris.com/2009/01/28/cost-of-spam-is-flattening-our-2009-predictions*.
http://www.redcondor.com/company/

[3] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy, and Ben Y. Zhao. User interactions in social networks and their implications. *EuroSys 2009*.