Detecting Spam in VoIP Networks

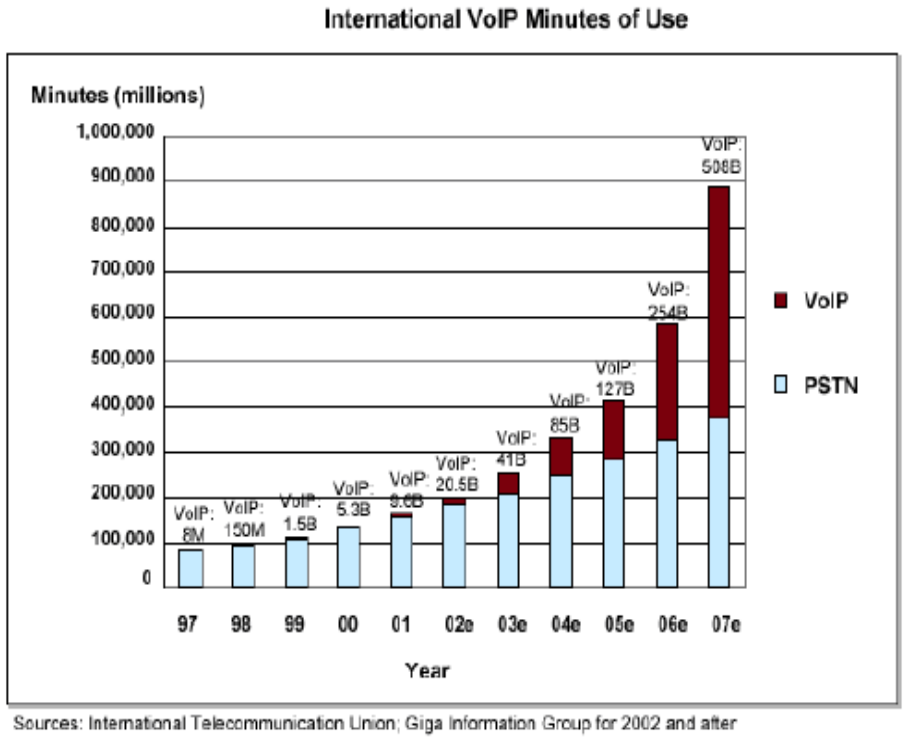
Ram Dantu

Prakash Kolan

Why use VOIP?

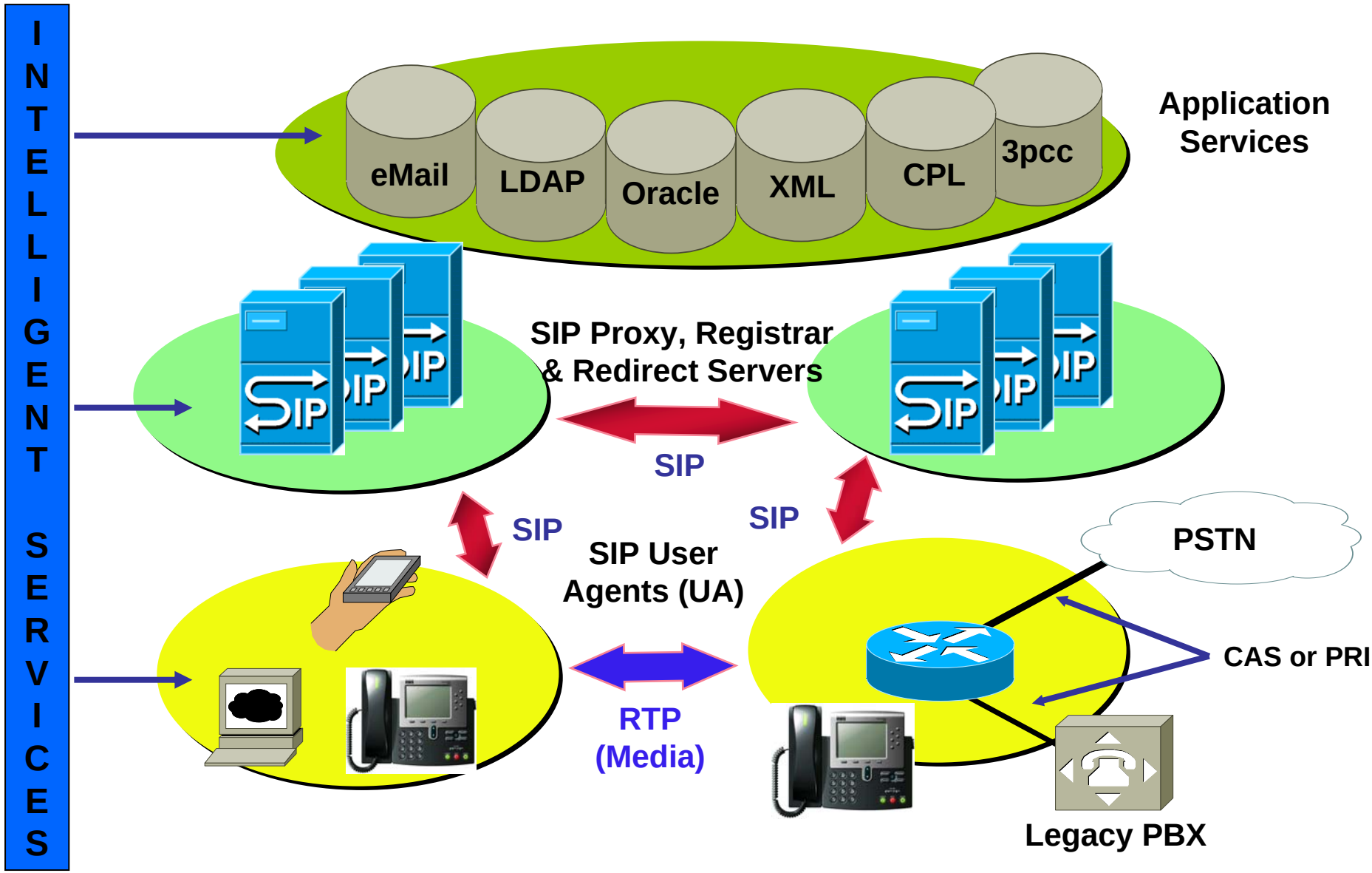
- More Multimedia Features
  - support for video-conferencing and video-phones
  - Easier integration of voice with applications and databases
  - **Convergence of voice and data communications**
- Cost
  - Long distance phone call costs virtually eliminated
  - PC + headset + software = telephone
  - Gartner: 90% of new corporate phones VOIP by 2008
  - TeleGeography 2005

    52% of online households said they were likely to subscribe to a flat-rate VoIP service package if it was priced at $30 per month

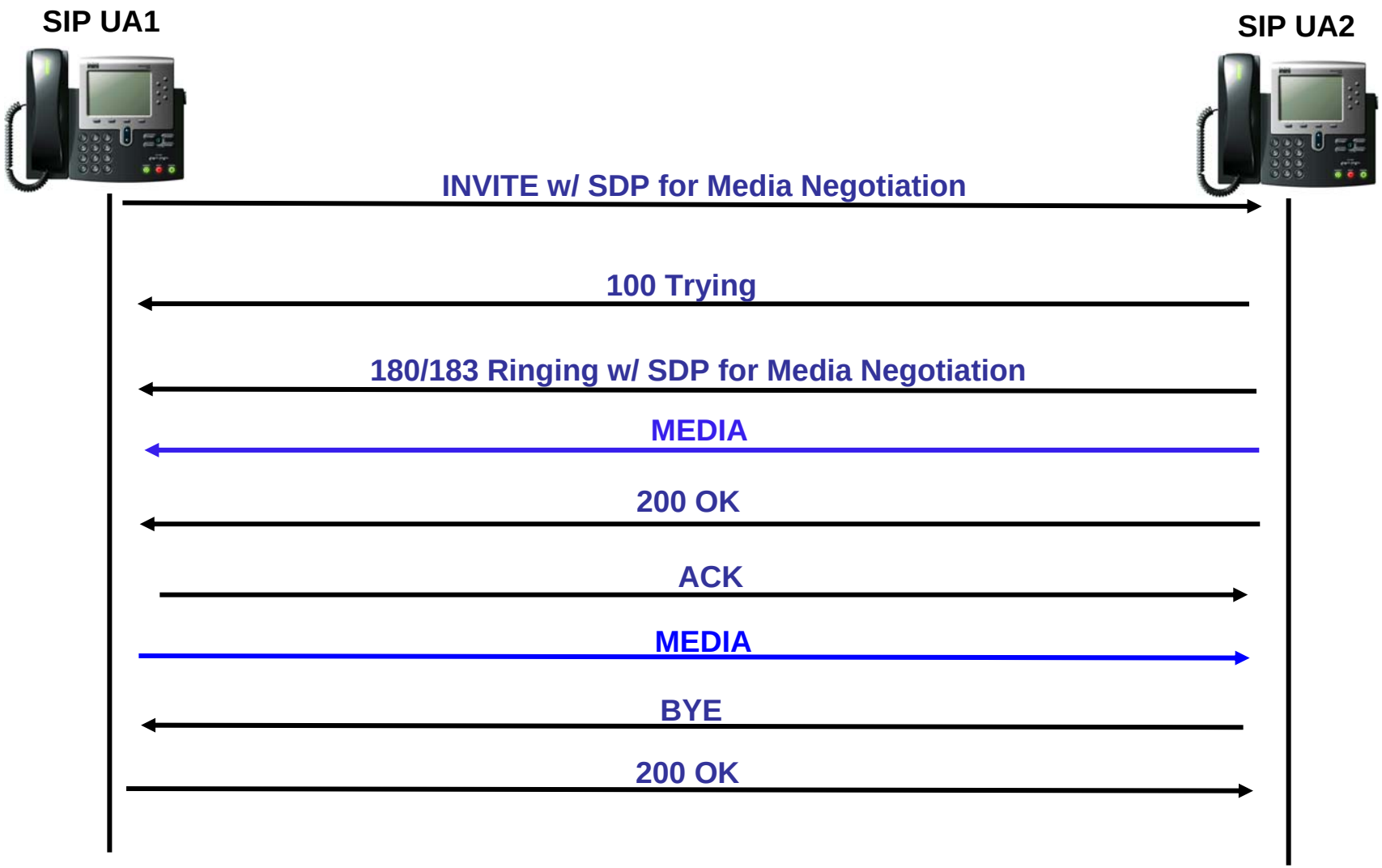


Overview of VoIP

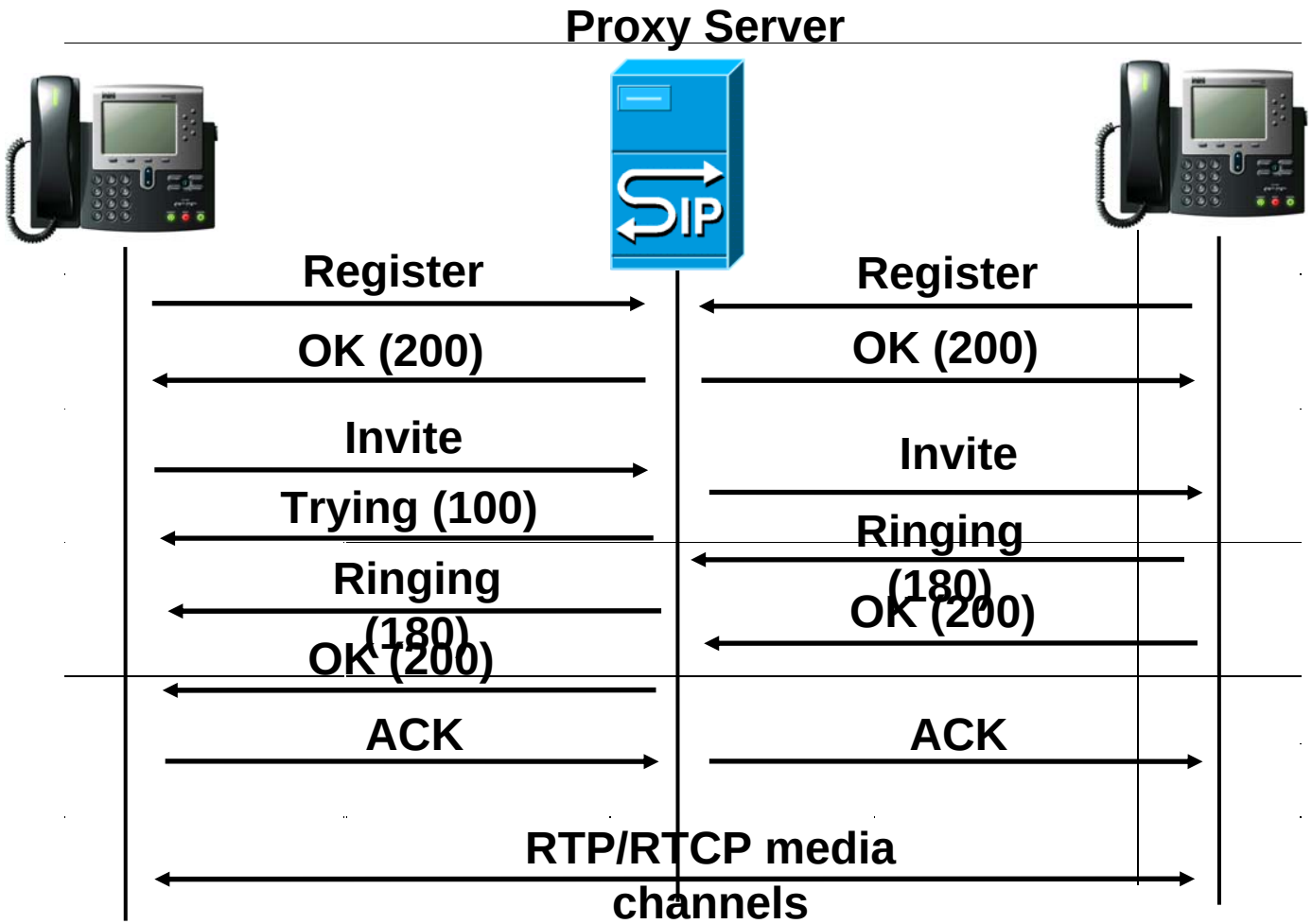
SIP based VoIP Architecture



Basic SIP Call-Flow



SIP Call Flow with Proxy Server



Issues

- VoIP is going to be a critical infrastructure for the nation
- Reliability and availability important for E911
- VoIP is real-time traffic. So QoS must be maintained
- VoIP inherits all properties of the IP protocol – including security weaknesses
- Separate signaling and media require different protection mechanisms thus complicating solutions
- FW and NAT traversal
- New/Changing Standards
- VoIP over WiFi/Cellular phones and mobility introduces new threats
- VoIP protocols/infrastructure used for IM, Multimedia, and Video
- Convergence of two global and structurally different networks (PSTN and VoIP) introduces new security weaknesses
- Threats like DDOS and Spamming are more damaging to voice compared to data networks
- Attack propagation; Threats and vulnerabilities will evolve and it is expected that they will impact (directly or indirectly) the existing PSTN.
- Attacks affect subscribers, service providers and carriers who are considered part of the National Critical Infrastructure

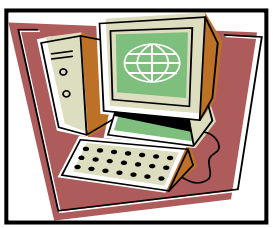
VoIP Security Workshops

- 1st VoIP Security workshop was held in Dallas, Texas as part of IEEE Globecom during December, 2004 (http://www.cs.unt.edu/~rdantu/VoIPSecurityWorkshop.htm)
- 2nd VoIP Security workshop was held in Washington DC during June, 2005 (https://www.csialliance.org/news/events/voip/VoIP_Agenda.PDF)
- More than 260 people have participated in these workshops. The participants include representatives from Department of Homeland Security, Department of Defense, the FBI, NSA, NIST, FCC, industry consortiums such as the International Packet Communications Consortium (IPCC) and SIP.EDU in Internet2, VoPSF, VoIPSA, and several telecommunications service providers, vendors and universities.

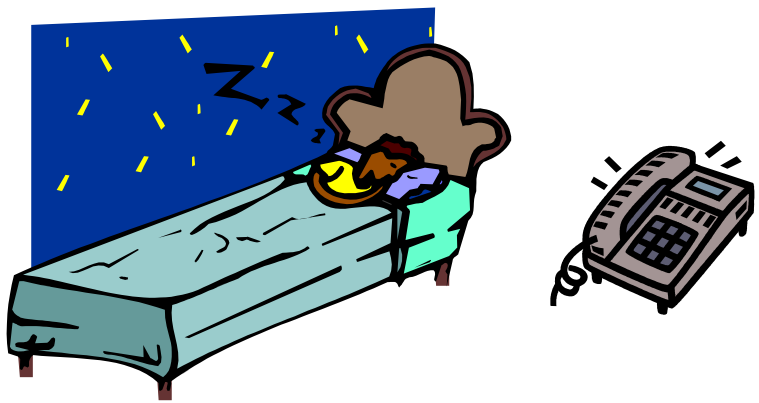
Voice Spam

- Voice Spam is different from E-mail Spam

E-Mail Spam at 2am

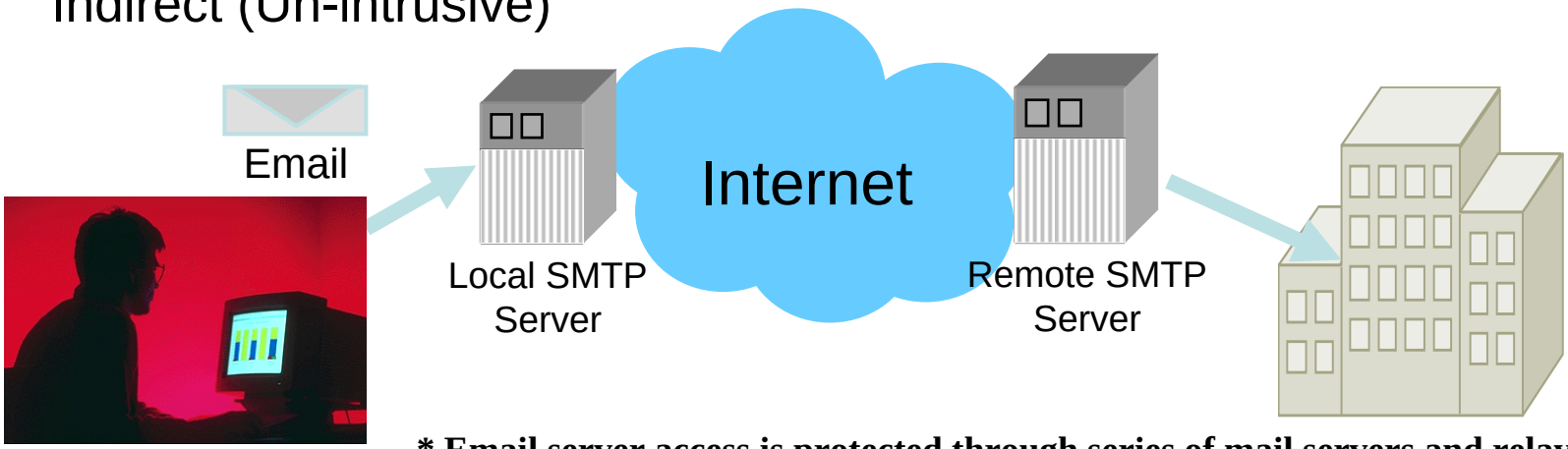


Voice Spam at 2am



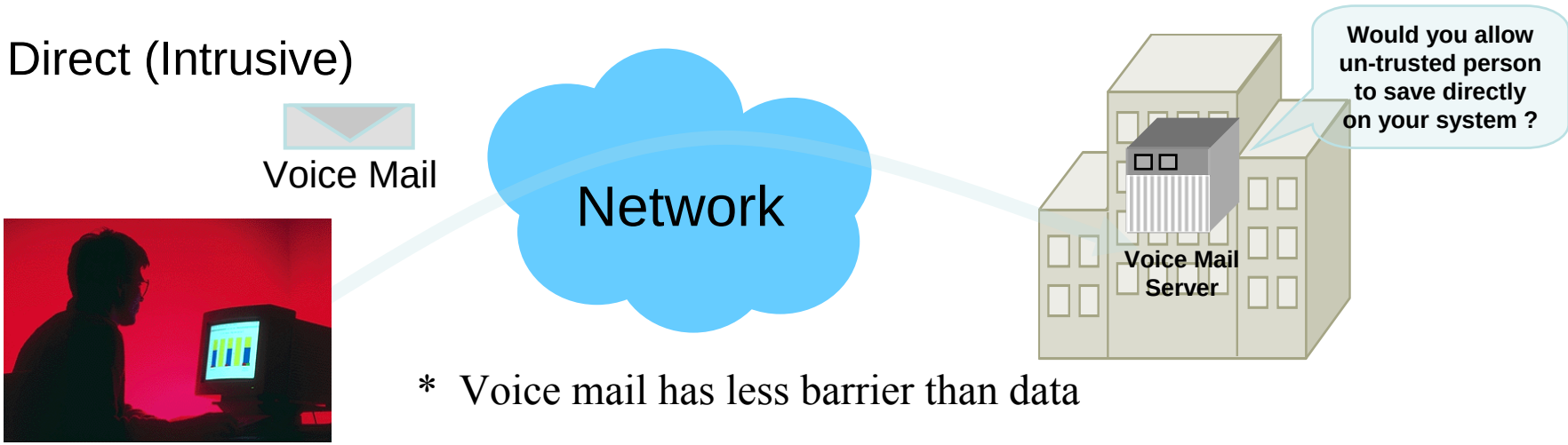
Email vs. Voice Mail

Indirect (Un-intrusive)



*** Email server access is protected through series of mail servers and relays**

Direct (Intrusive)



* Voice mail has less barrier than data

Spamming

- Different from e-mail spam
  - At 2AM, Received a junk e-mail sitting in Inbox.

    But, a junk voice call is a real nuisance.
  - Most E-mail filters rely on content analysis. But in Voice calls, it is too late to analyze media for spamming
- Voice Spam Detection – difficult
  - Headers for voice spam detection : “*from*” , “*contact*”. Are these enough ?
  - Detection in real time before the media arrives
- Spam is basically an unwanted call !!!!!

Solution

Functional Elements

- Elements in Spam Detection
  - Black and White Lists
    - Database of wanted and unwanted callers- Black & White Lists
    - Preference based on domains and contacts.
  - Bayesian Learning
    - Learning based on past history
    - Past history specifies spam or valid behavior about the caller
    - Update the learning based on past history, new evidence like feedback from end users.

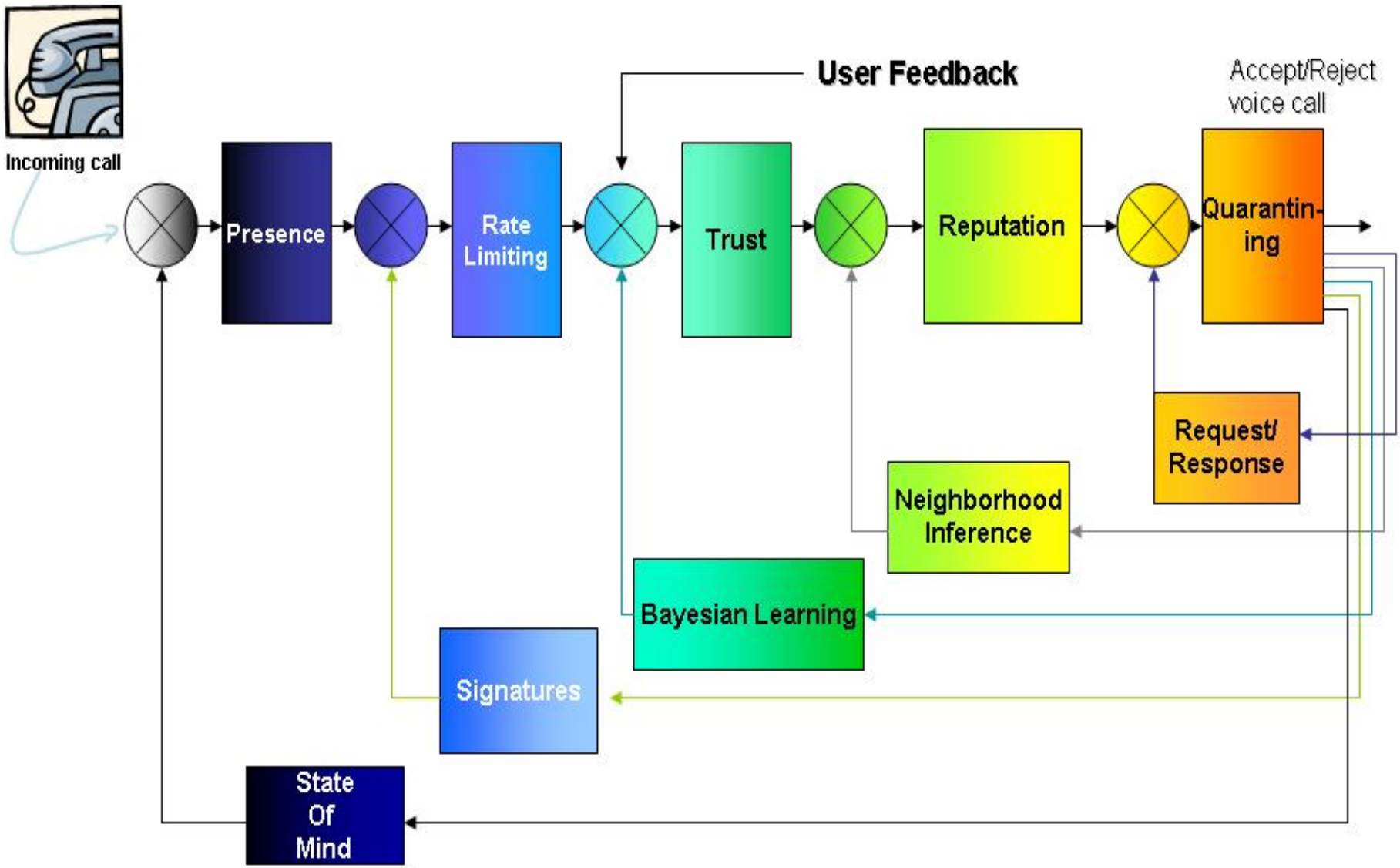
Inference

- Trust
  - Trust is a a derived entity.
  - Depends on
    - Caller’s past behavior with callee
    - Spam and valid calls from the caller
  - The identifiers having extreme values of Trust can be added to Black and White lists.
    - This is time bound as few spam calls after a good behavior might go unnoticed, or vice versa.
  - Trust grows slowly with time for a valid behavior and drops fast for a spam behavior. Additive growth and multiplicative descent.

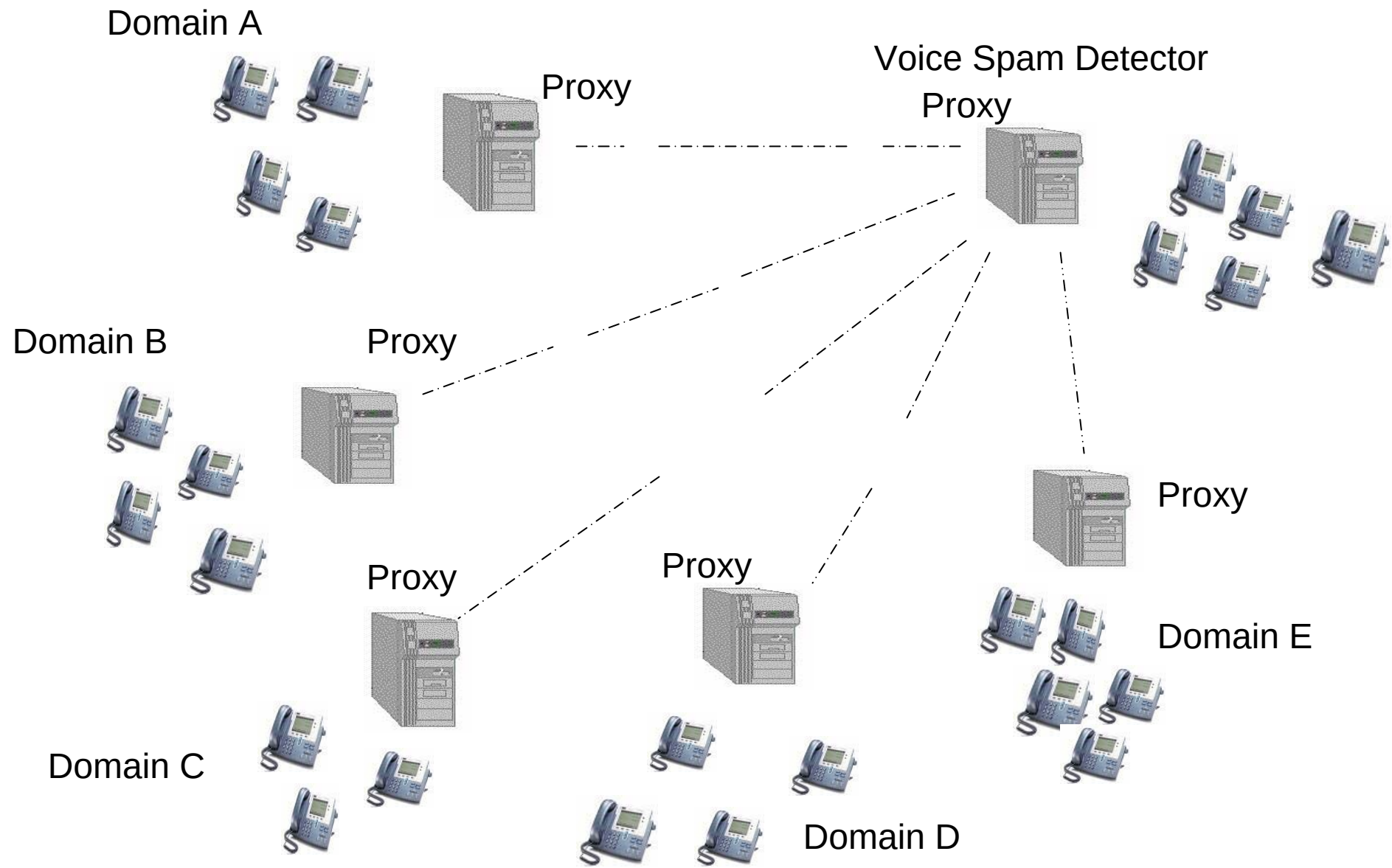
Functional Elements Contd..

- Inference
  - Inference from trusted neighbors
  - Reputation coefficient for each neighbor
  - Bayesian Inference of reputation for known proxy/neighbor topology
  - Update Trust and Reputation coefficients based on calling patterns
- Presence
  - Called party preferences and interests
  - Presence of situation – State of mind and mood of called party

Functional Elements of VSD



Experimental Setup



Traffic Profile

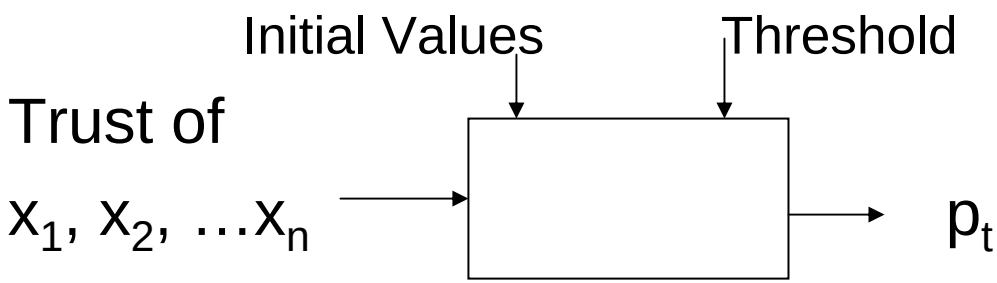
- Five domains, 35 hosts per each domain and 100 users for each host
- Average rate of 8 calls per minute
- Neither the VSD nor the called-users have any idea regarding the call generation process.
- Randomly select a set of users, hosts and domains as spammers before the start of the experiment
- *A button called “SPAM” included in each hard phone in the receiving domain for giving feedback to the VSD.*
- Randomly select a set of users, hosts and domains as spammers before the start of the experiment

Trust Computation

History regarding participating entities – $x_1, x_2, \ldots x_n$

$p_t =$ $$\frac{P(C = spam)\prod_{i=1..n} P(x_i = spam)}{\sum_{k=spam,valid} P(C = k)\prod_{i=1..n} P(x_i = k)}$$

$x_1, x_2, \ldots x_n$ are updated after every call



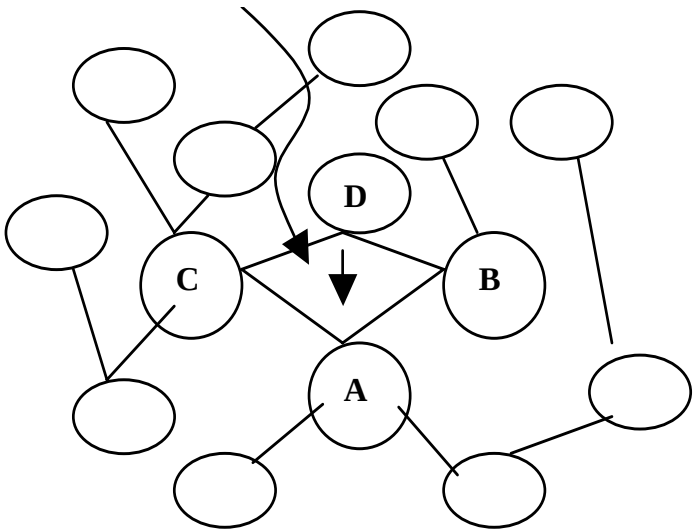
Reputation Inference

Reputation Coefficient R(*a*,*b*) - Reputation of *b* on *a*

In the above figure R(*a*,*c*) = Θ(R(*a*,*b*),R(*b*,*c*)) for all *b* in trusted neighbors of *a*

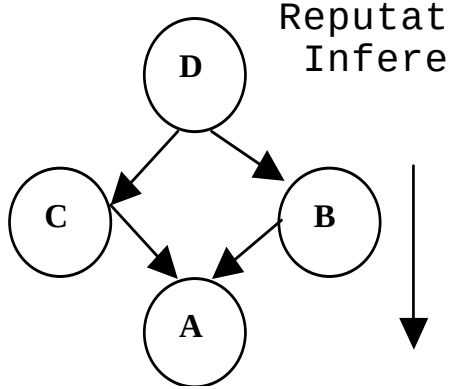
Θ represents the Bayesian Inference function. This inference is carried out using inference graph as shown below.

Call from a host in Domain D to a host in Domain A.



Domain Proxies

Reputation Inference



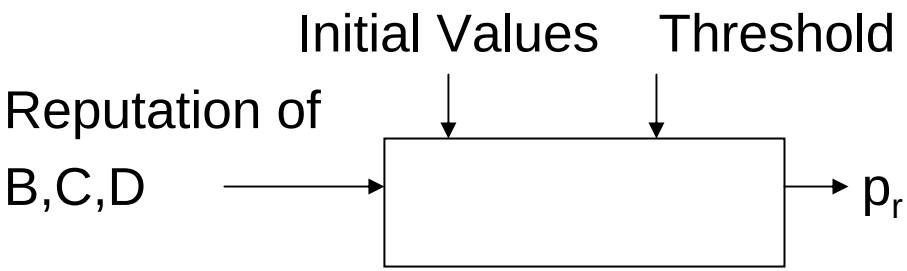
Inference Graph of Network topology

AP, BP, CP and DP are the proxies of domains A,B,C and D respectively

Reputation Inference

For a call from a domain

The reputation is updated after every call based on user feedback. The update is positive for a valid call and negative for a spam call.



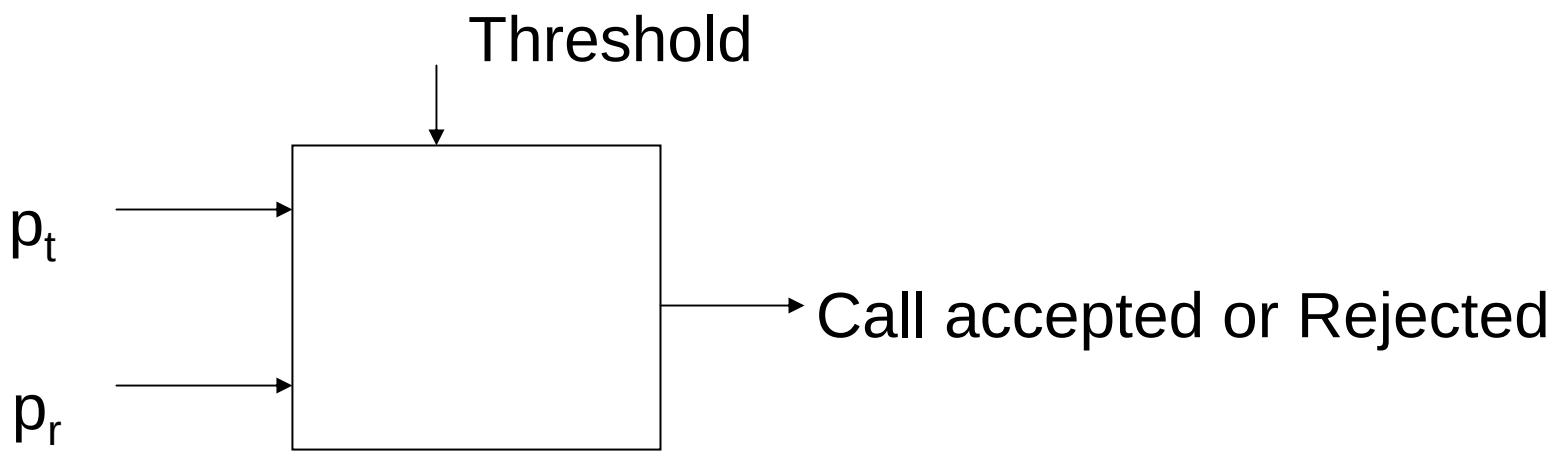
All the generating and intermediate domains routing the call are updated based on their respective learning.

Message Probability

Message Probability

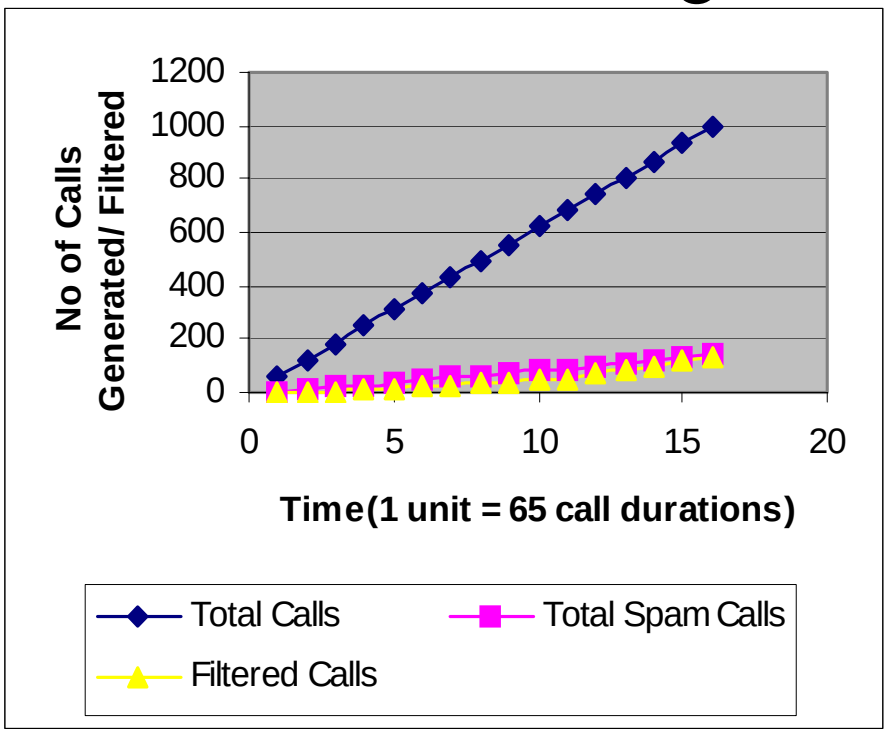
$$p = f(p_t, p_r)$$

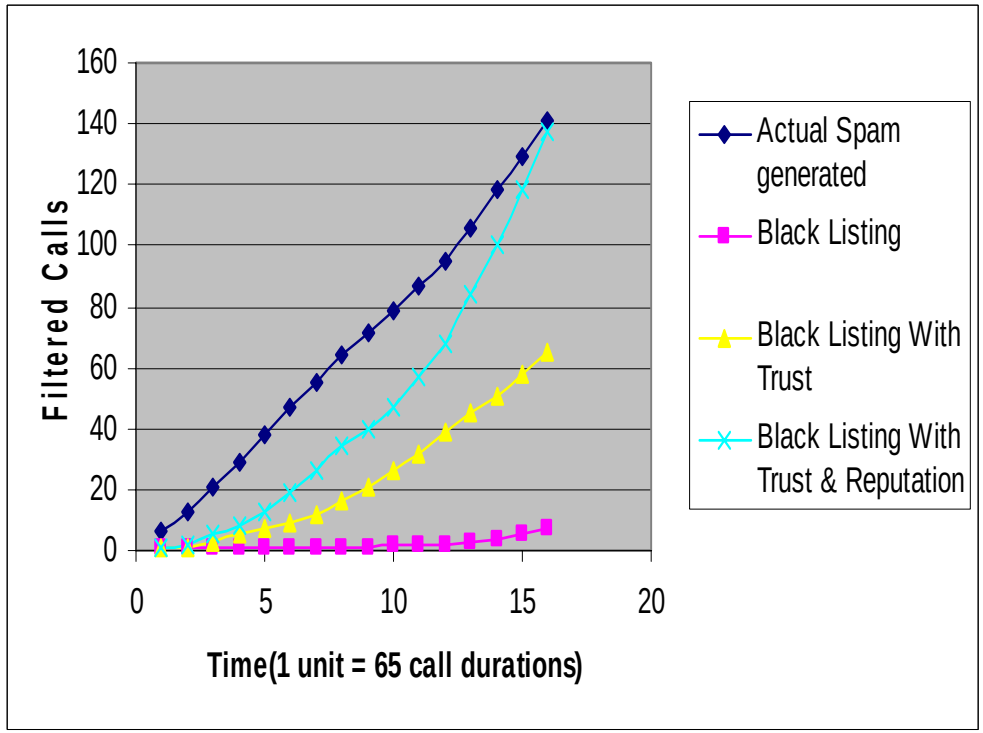
If p>T, call = spam else call = not spam



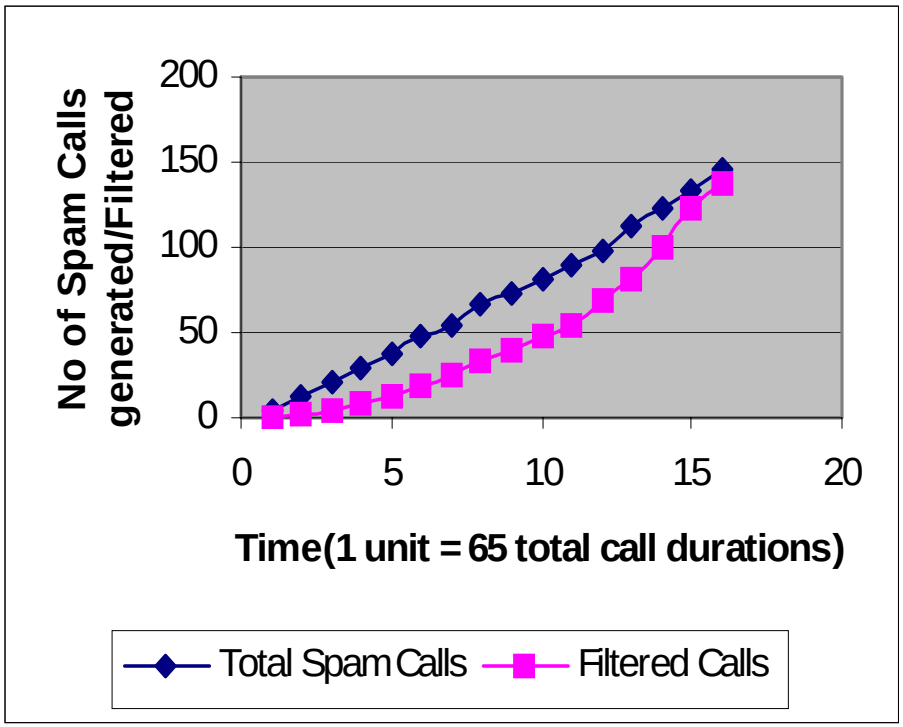
Results

Collaboration between different methodologies in the architecture

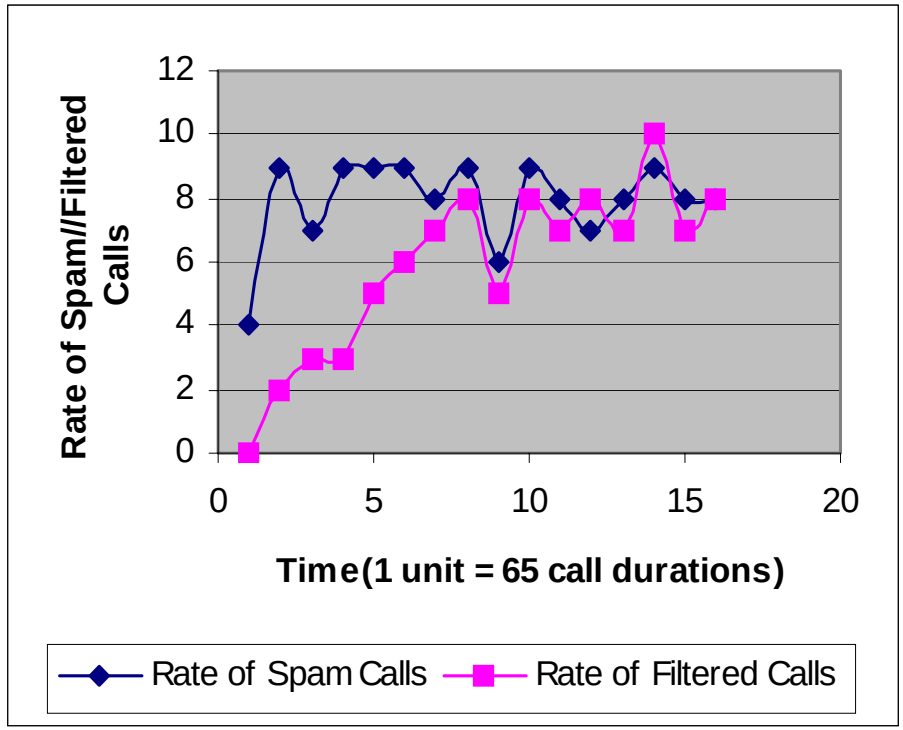




Accuracy of VSD during learning and lock-in period

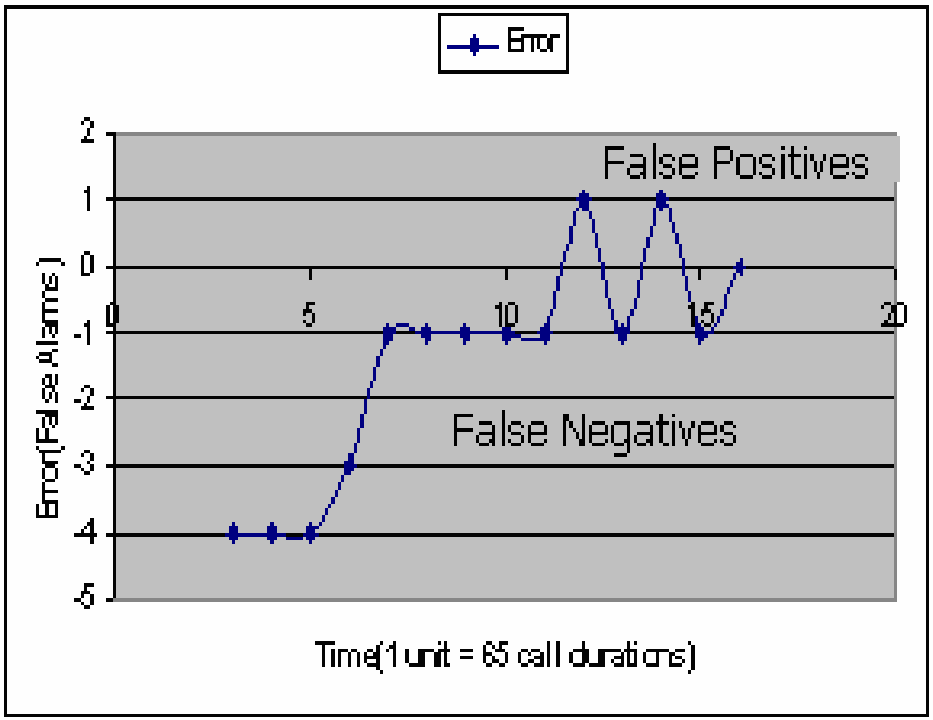


Accuracy of VSD During Learning Period

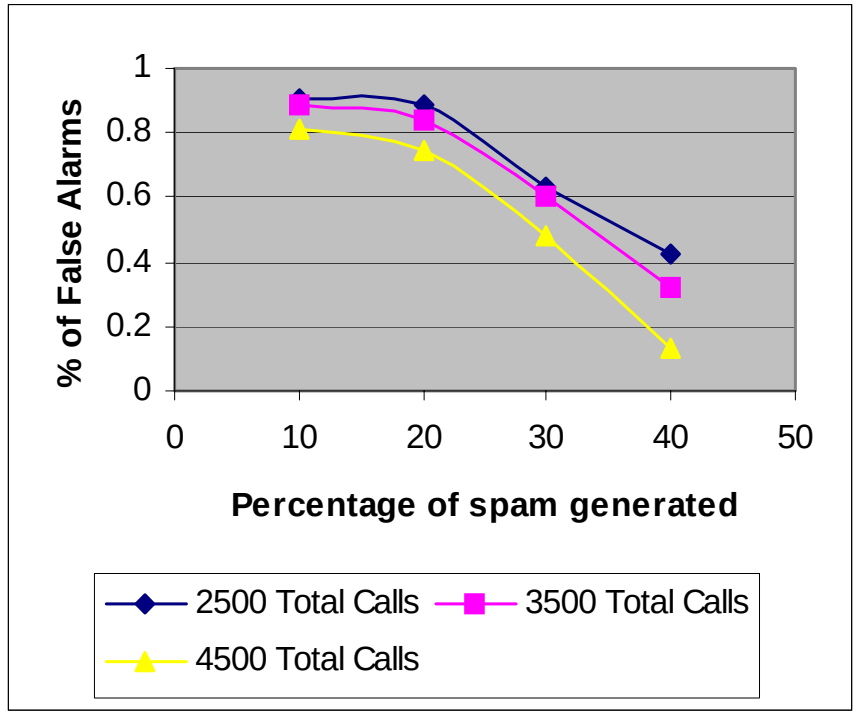


Accuracy of VSD During Lock-In Period

Sensitivity Analysis

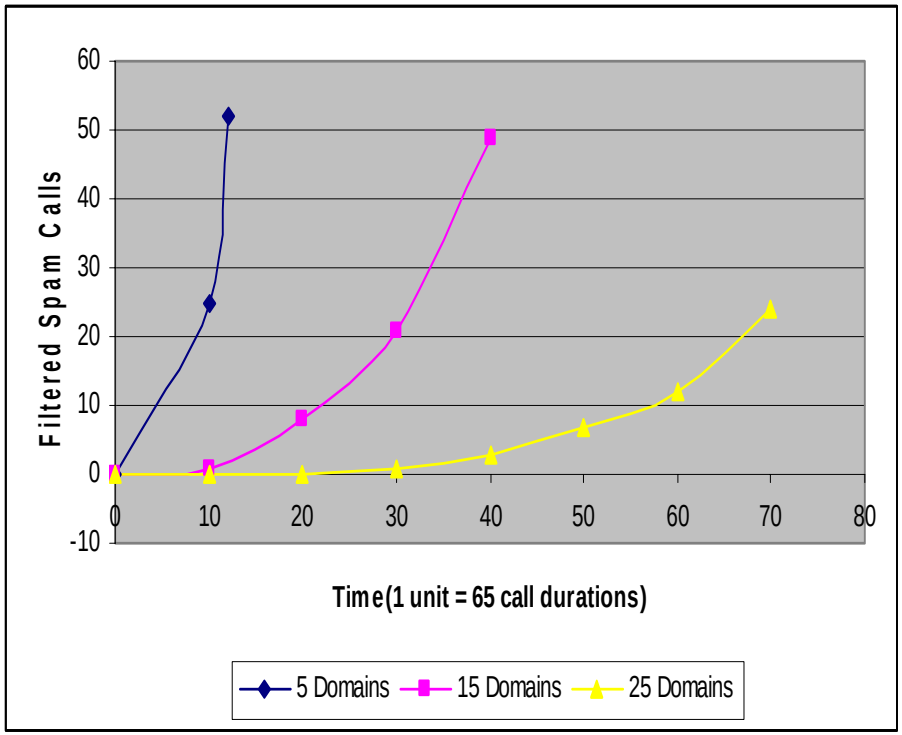


Accuracy of VSD

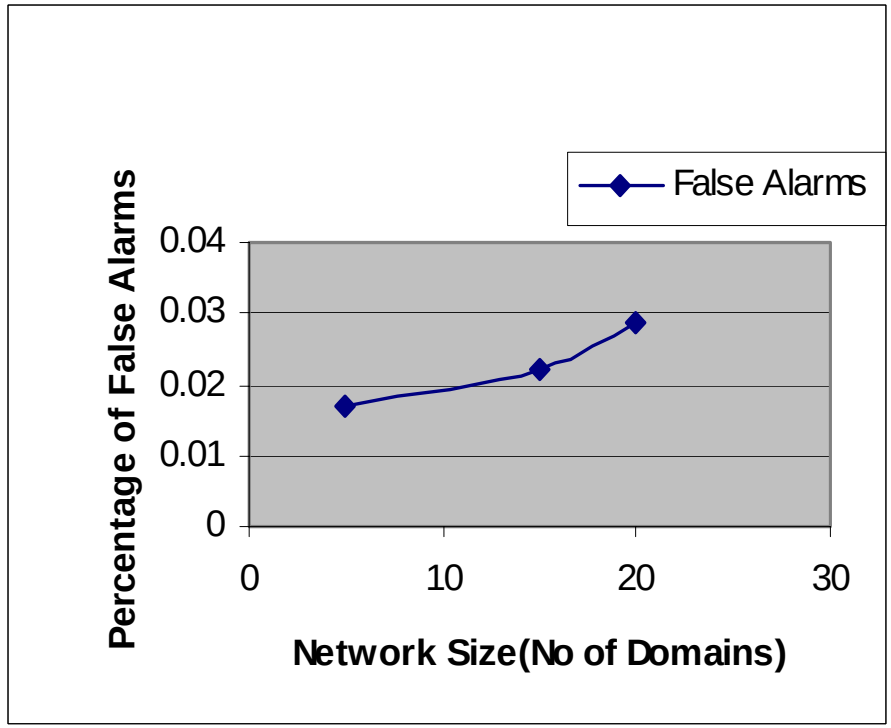


Spam Volume Vs. Accuracy

Sensitivity Analysis

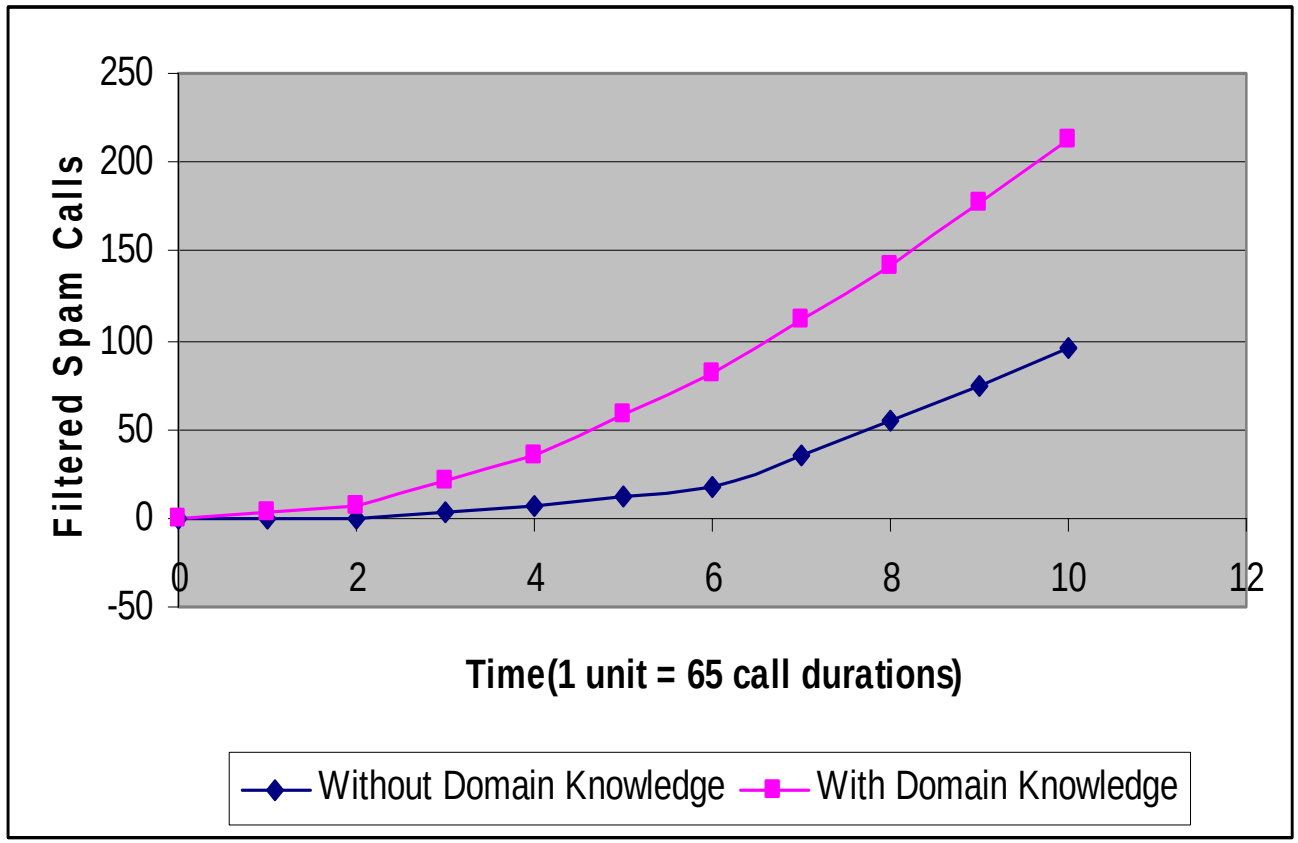


Network Size Vs. Accuracy

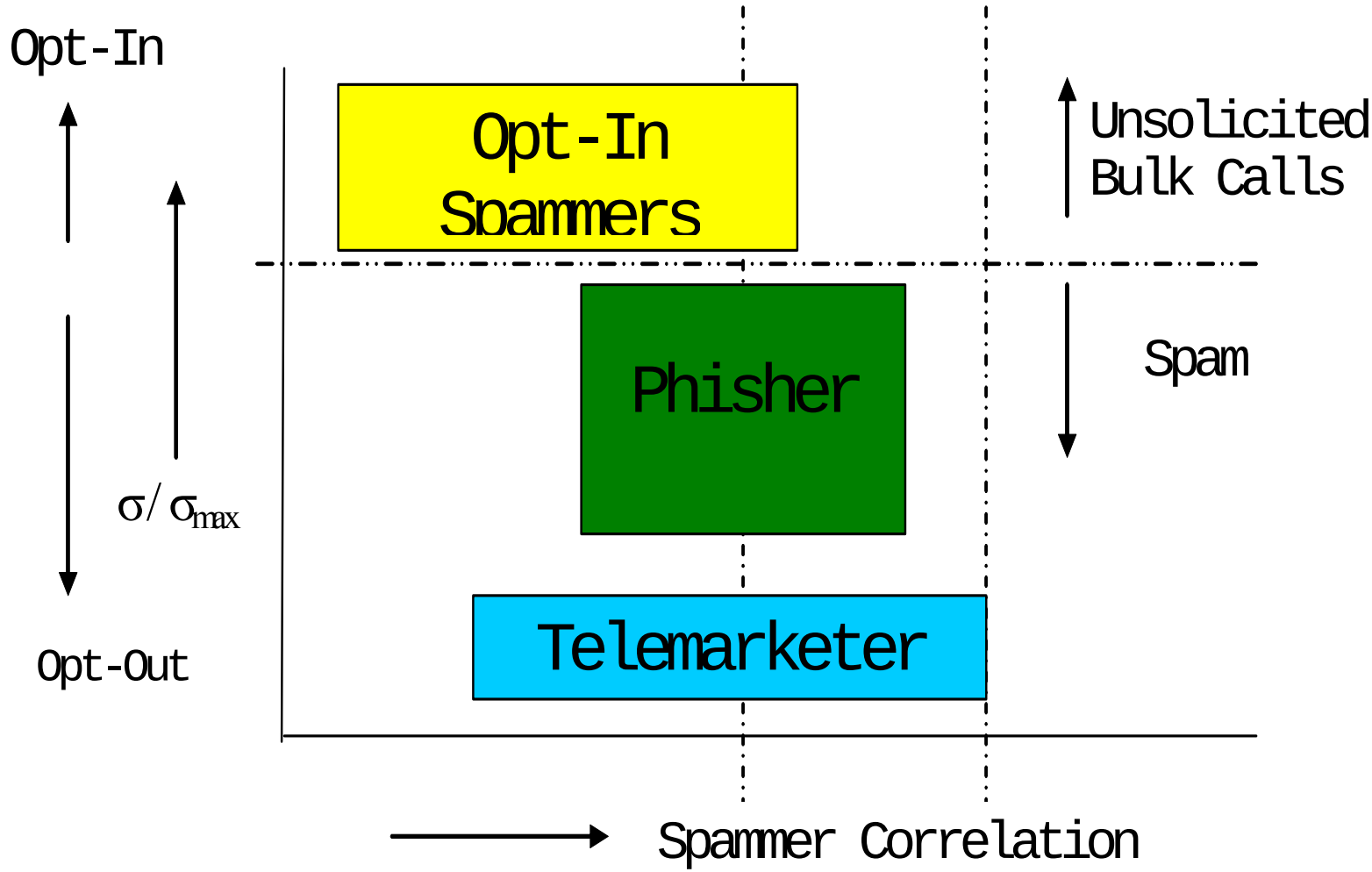


Network Size Vs. False Alarms

Using Domain-specific knowledge



Learning From Others Experience



Spammer Classification based on their correlation and spam distribution

Conclusions

- An architecture based on collaboration between different functional elements is proposed
- We have used open source proxy server, soft clients and hard phones for the experiments
- A combinations of black/white lists, trust and reputation functions and media quarantining are used
- Spammer classification technique is developed and supplemented with the above results
- From our observation of the logs it takes 3 spam calls to confirm it is a spam and the 4th call can be accurately identified as the spam.