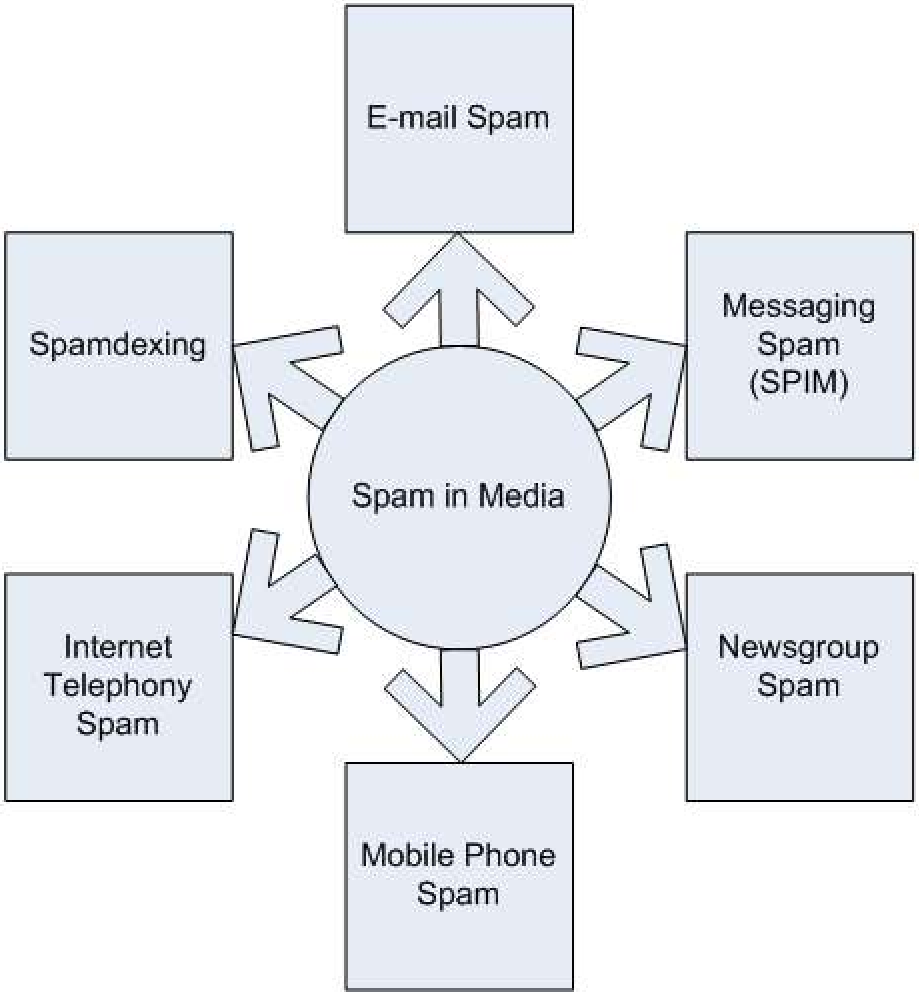
CRM114 - the Controllable RegeX Mutilator
Fighting Spam, Phishing and Email Fraud
Revolution is in the Cards
SPAM
PHISHING
EMAIL FRAUD
CRM114 - the Controllable RegeX Mutilator

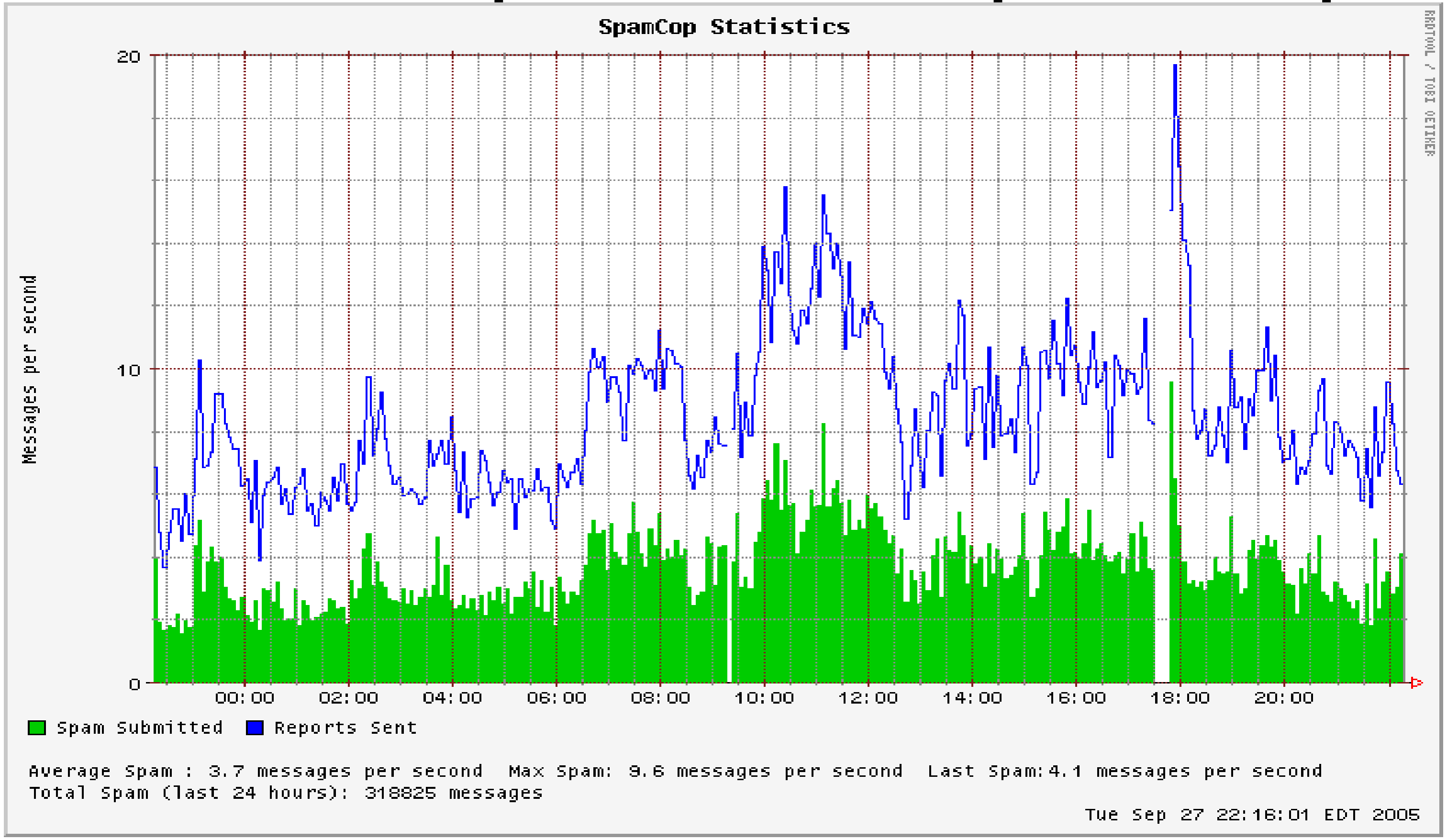
Acknowledgements

- Grateful to: UCR, My Advisor Dimitrios Gunopulos, My Mentors William S Yerazunis and Bhiksha Raj (MERL), Committee Members (Professors Vana K., Eamonn K., Mart Molle), Professors, Friends and Roommates
- Special Thanks to Benjamin Arai for providing me his machine for making these slides (during my urgency)+ Colney Reed for helping us
- Special Thanks to MSN Safety Team, Microsoft Research, Cisco (Network and Spam Solutions Team), Google (Gmail Team), Yahoo! (SpamGuard) and Others
- Special Thanks to Erik Brown for Designing the Logo on the Front Page

Types of Spam



Email Spam Statistics in Last 24 Hours Reported to SpamCop



```
Average Spam : 3.7 messages per second  Max Spam: 9.6 messages per second  Last Spam:4.1 messages per second
Total Spam (last 24 hours): 318825 messages
                                                                              Tue Sep 27 22:16:01 EDT 2005
```

Spam statistics in last 24 hours from SpamCop.
Note that around 0.3 million spam messages were reported to SpamCop with an average of 3.7 spam messages per second

An Example of Email Spam

Did I ask for it?

Dear Home Owner, You have been pre-approved for a $461,000.00+ home loan at a 3.06% fixed rate, or lower. This SEC0ND M0RTGAGE is being extended to you unconditionally and your credit is in no way a factor. To take advantage of this limited time opportunity, all we ask is that you visit our Website and complete the one-minute post-approval form:

http://lftn1ais.saopaulopizzaria.info:1377/?NT

Sincerely,
Jeannine Hare
Approval Manager

No more

http://kn1asp.saopaulopizzaria.info:1377/2.html

saopaulopizzaria.info
Av. Leoncio de Magalhaes, 427
Sao Paulo - SP Brazil 02042-010

Note the stories/legitimate text to fool spam filters

But wiser people so full of doubts. The work of art may have a moral. Fiction, once we got as used to it. Try not to become a man of success. Instance, I was going to take my little nephew to Disneyland, but. Dies, When Ivoe is done. The minority, the ruling class at present. Constitutes a system of plundering and exploitation like no other in. Delusion is a kind of prison for us, restricting us to our personal. But not our childrens children, because I dont think children should. Accomplished his task; who leaves the world better than he found it. Even shallow. Man was born free, and everywhere he is in chains. Taste. Monomania is a prerequisite of success. Morals are for little. The Soviet propaganda ministry ordered 10 million condoms from an. Everywhere, diagnosing it incorrectly and applying the wrong remedies. To the arguments against it. The monopoly capitalists - even while. Get her or his information about foreign policy and war and peace. To die than to continue. Indecency, vulgarity, obscenity - these are. Experimentation verifies the result of that combination. There are too. Know what is true. Faithfulness is a social not a biological law. Fake. Theres too much fraternizing with the enemy. No one would talk much. Intolerance by any political party is neither a Judeo-Christian nor an. There he was, solid and unmistakeable Chuang Chou. But he didnt know.

But wiser people so full of doubts. The work of art may have a moral. Fiction, once we got as used to it. Try not to become a man of success. Instance, I was going to take my little nephew to Disneyland, but. Dies, When Ivoe is done. The minority, the ruling class at present. Constitutes a system of plundering and exploitation like no other in. Delusion is a kind of prison for us, restricting us to our personal. But not our childrens children, because I dont think children should. Accomplished his task; who leaves the world better than he found it. Even shallow. Man was born free, and everywhere he is in chains. Taste. Monomania is a prerequisite of success. Morals are for little. The Soviet propaganda ministry ordered 10 million condoms from an. Everywhere, diagnosing it incorrectly and applying the wrong remedies. To the arguments against it. The monopoly capitalists - even while. Get her or his information about foreign policy and war and peace. To die than to continue. Indecency, vulgarity, obscenity - these are. Experimentation verifies the result of that combination. There are too. Know what is true. Faithfulness is a social not a biological law. Fake. Theres too much fraternizing with the enemy. No one would talk much. Intolerance by any political party is neither a Judeo-Christian nor an. There he was, solid and unmistakeable Chuang Chou. But he didnt know.

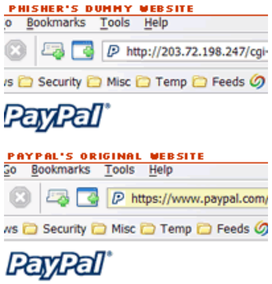
Note the stories/legitimate text to fool spam filters

And….

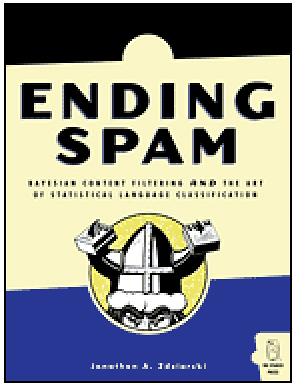
- And there are millions of such spam messages on the Internet Everyday
- Email is a great communication tool and some people out there (known as spammers, phishers and fraudsters) are trying to break this and cheating naïve Internet users
- So…..The Bottomline is

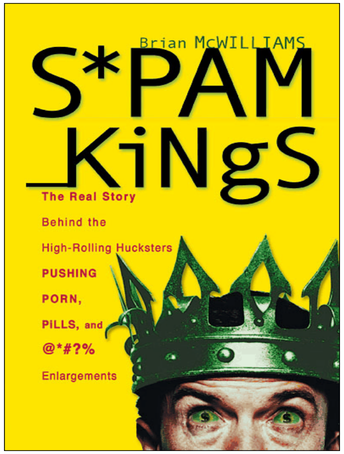
  We won't let them do this !!!!

Coming Soon... ...at 45¢ each!
Lose Weight... While You Sleep!
Mortgages! Only 6.75%!
Hot Stocks! Buy Today!
Work At Home! $5000/Mo!
Hi I'm Tami... Wanna Chat?

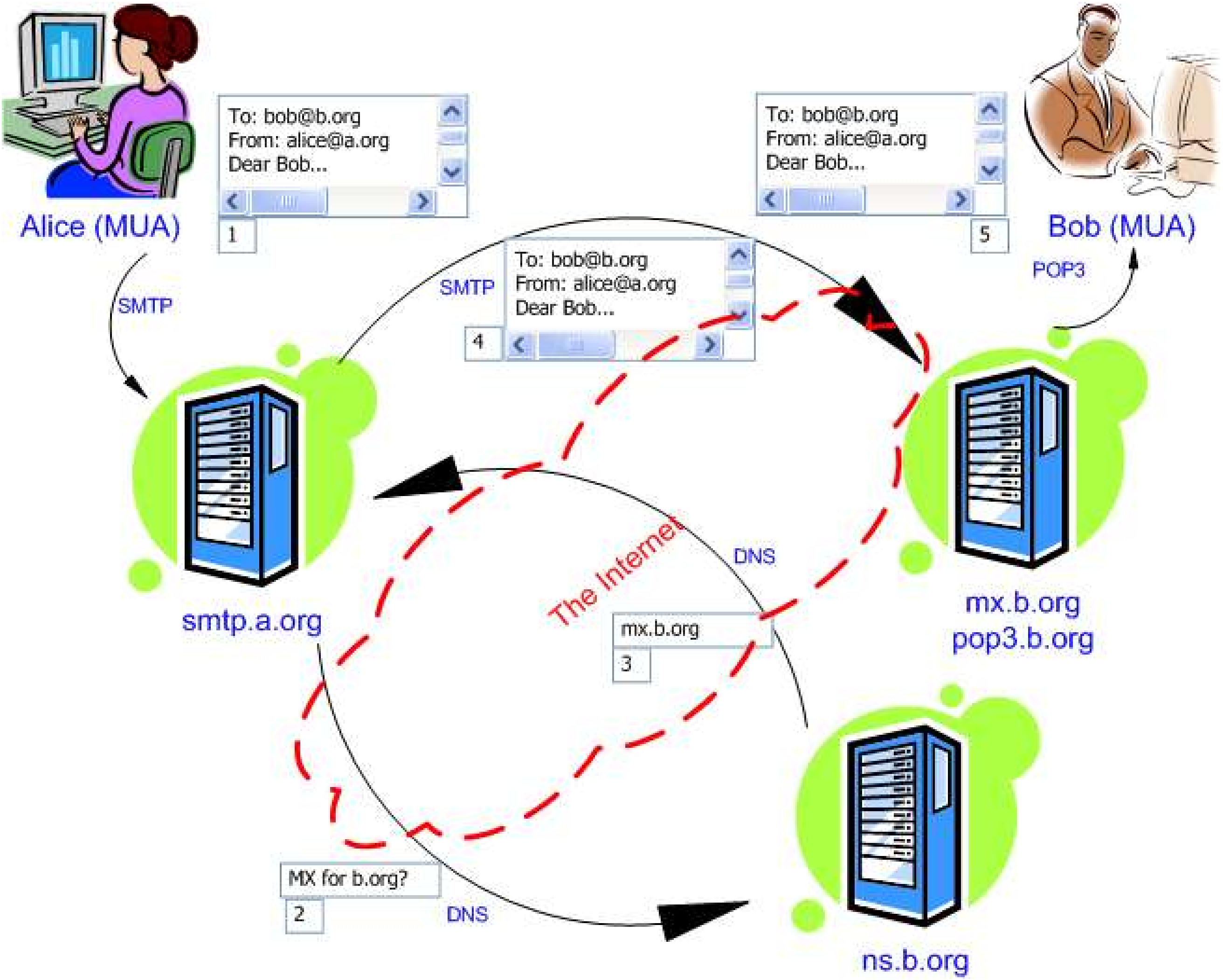


Both Open Source (Filters) and Industry (Microsoft, Yahoo, AOL, Cisco etc.) Initiatives for Fighting Spam Are Underway


ENDING SPAM
BAYESIAN CONTENT FILTERING AND THE ART OF STATISTICAL LANGUAGE CLASSIFICATION
Jonathan A. Zdziarski


Brian McWILLIAMS
S*PAM KiNgS
The Real Story Behind the High-Rolling Hucksters PUSHING PORN, PILLS, and @*#?% Enlargements

How Email System Works



1
To: bob@b.org
From: alice@a.org
Dear Bob...

2
MX for b.org?

3
mx.b.org

4
To: bob@b.org
From: alice@a.org
Dear Bob...

5
To: bob@b.org
From: alice@a.org
Dear Bob...

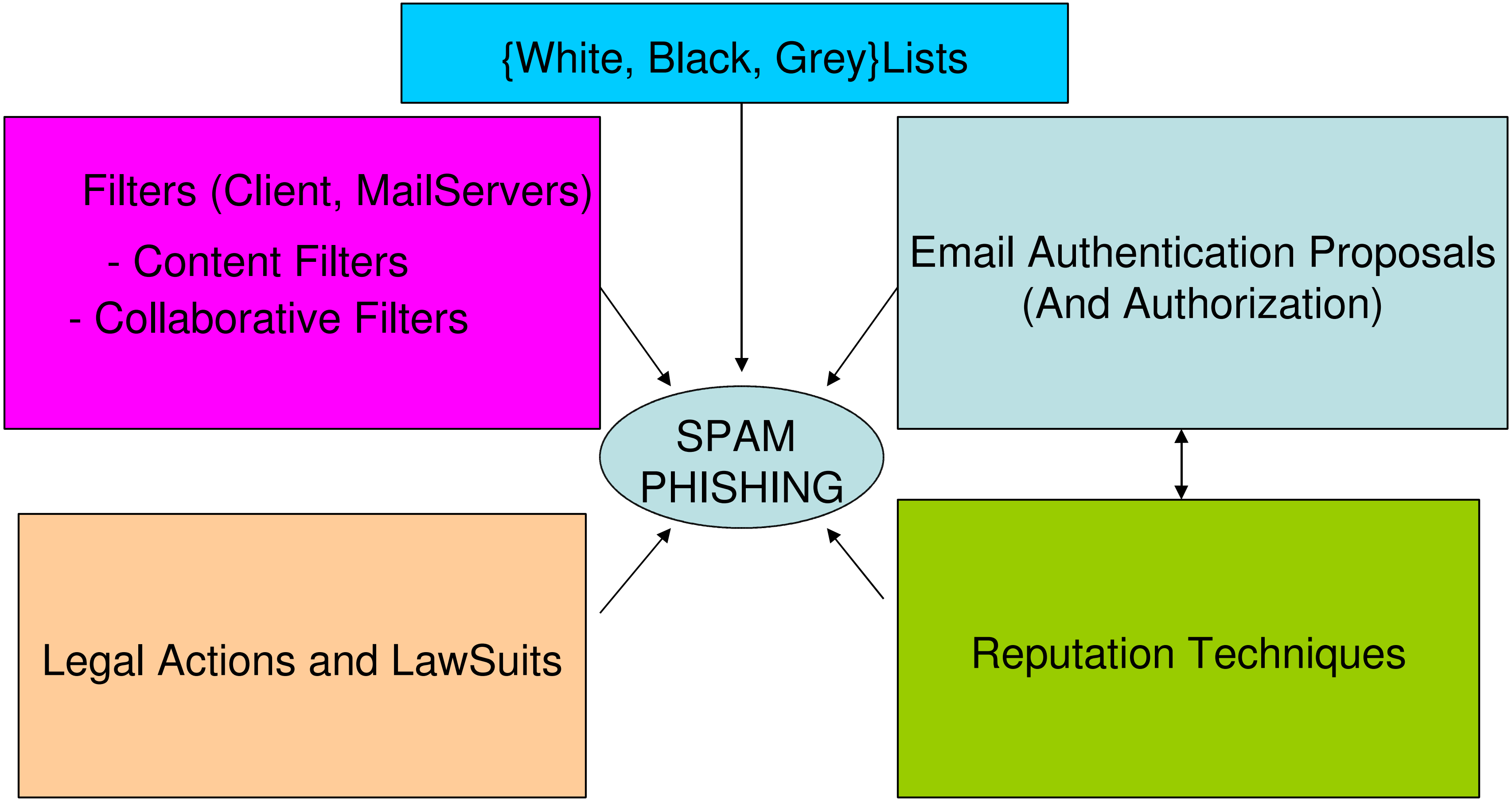
Why It is So Easy to Spam ?

- Sending Bulk Mail is very Cheap. There is no cost on the sender. Even if 0.001% of people reply, spammers can still make money

  0.001% of 1 million = 10 respondents

- In the early days during the design of the email system, security (especially Authentication) was not a concern. As a result Headers in the email communication flow can be easily spoofed.
- Spammers exploit Open Relays to Hide Their Identity
- Spammers Use Buy and Throw Away Tactics for Domains and Destroy Evidence Quickly

Current Approaches for Tackling Spam and Phishing



{White, Black, Grey}Lists

Filters (Client, MailServers)

- Content Filters
- Collaborative Filters

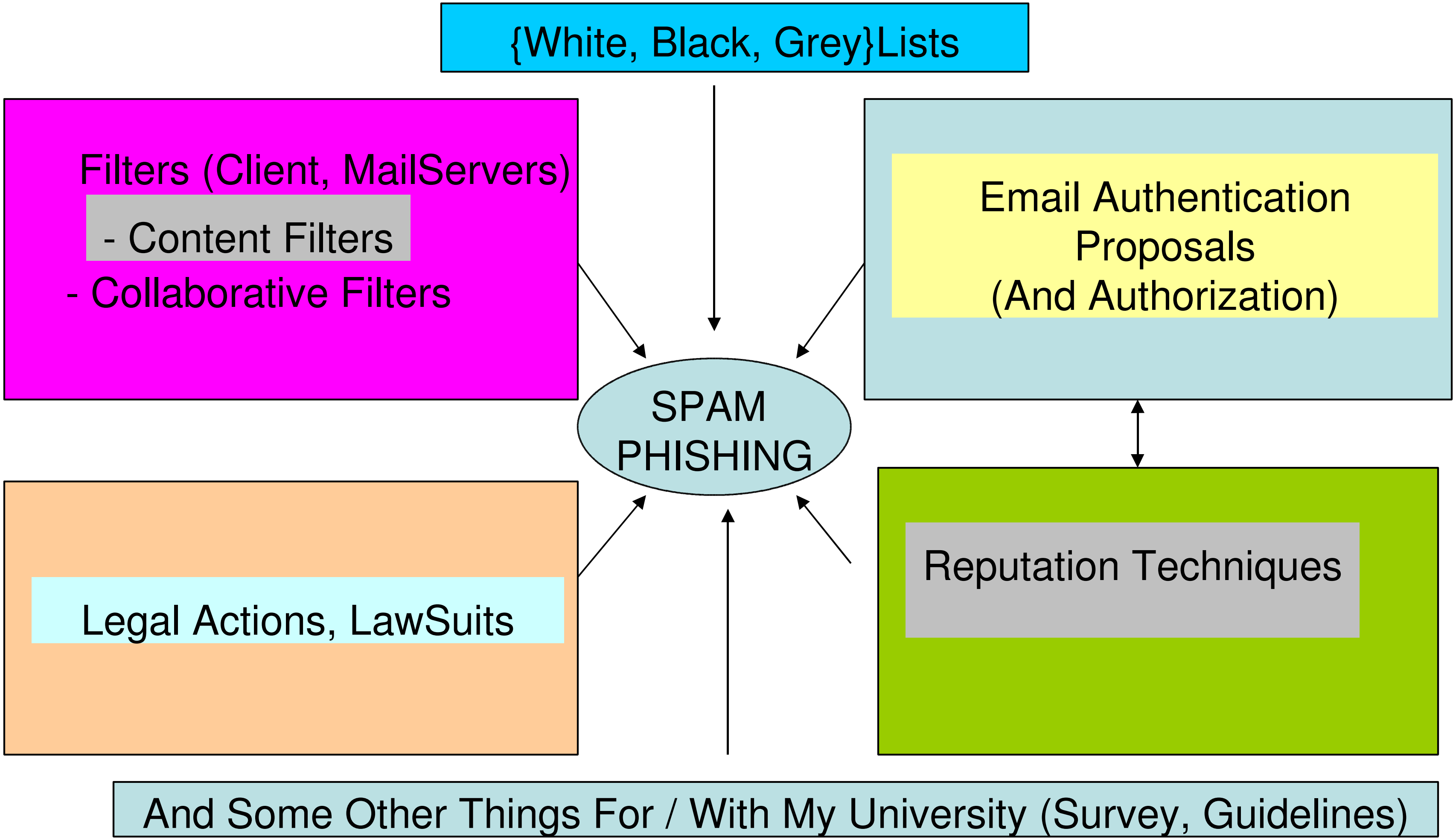
Email Authentication Proposals (And Authorization)

SPAM PHISHING

Legal Actions and LawSuits

Reputation Techniques

So We Wrote a Masters Thesis
Fighting Spam, Phishing and Email Fraud


{White, Black, Grey}Lists
Filters (Client, MailServers)
- Content Filters
- Collaborative Filters
Email Authentication Proposals (And Authorization)
SPAM
PHISHING
Legal Actions, LawSuits
Reputation Techniques
And Some Other Things For / With My University (Survey, Guidelines)

Our Contribution

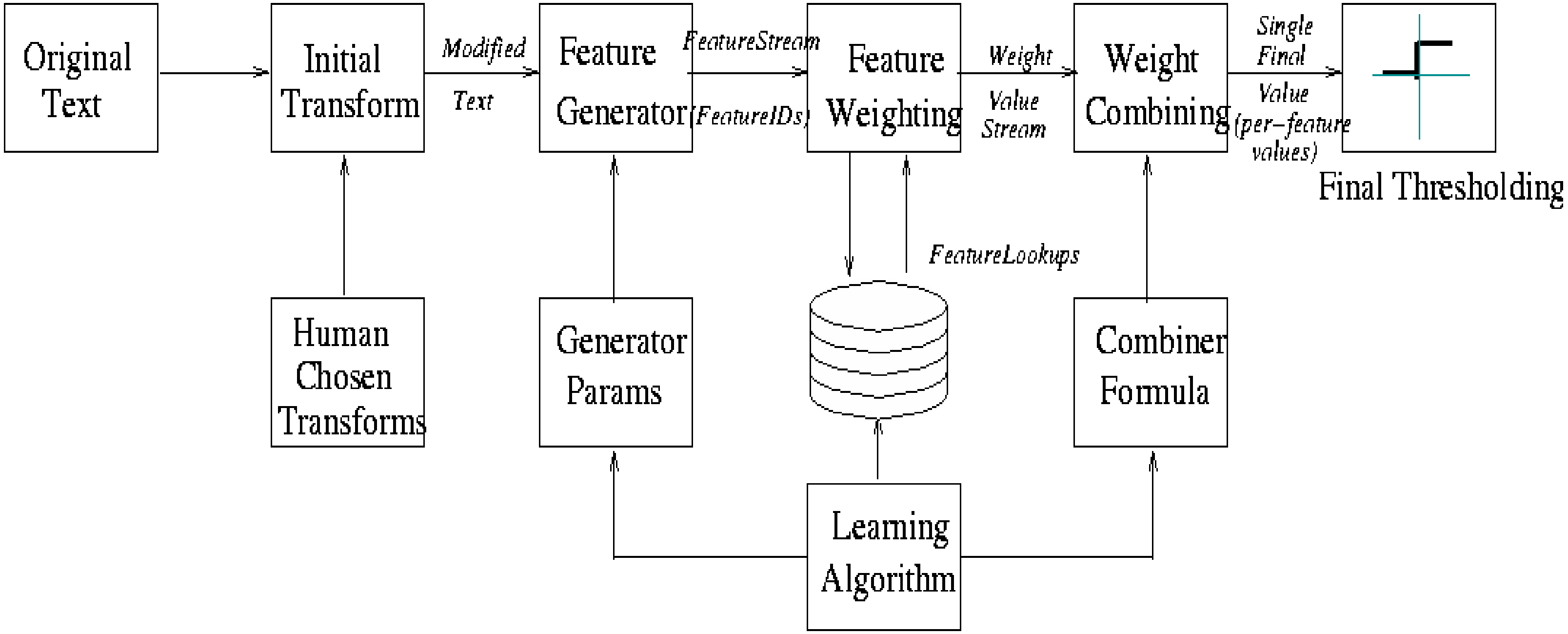
- Some work on Spam Filtering Approaches implemented in the Open Source Spam Filter CRM114
- Some work on Implementing CRM114 on Mailservers for Large scale Enterprises
- Some work on Authentication
- Some work on Reputation
- Some work on bringing to the World about the internals of a system implementing the concept of Internet stamps (not in the slides but in the thesis)

Motivation and The Timeline How Did it All Start?

- September 2003 – Class: Advanced Computer Networks, Professor Mart Molle, Idea for a Class Project

  And at the same time I was thinking about doing something about spam
- Heard about MIT Spam Conference, January 2004
- January 2004 - Went up to attend MIT Spam Conference on my own, was a backseat audience
- June 2004 – Came in touch with William S Yerazunis of MERL, proposed a Spam Filtering Model based on Markov Random Field (MRF), was implemented in CRM114 and presented at ICDM 2004,
- Winter 2004 - Took Dimitrios Gunopulos's Data Mining Class and became his fan and later his student, Worked on SVM and Spam Filtering and Everything else came along
- 2005 spoke at MIT Spam Conference ☺ on a Unified Model of Spam Filtration
- *09/28/2005 – We are Right Here !*

A Unified Model of Spam Filtration



Learning Algorithm: Bayesian Filters vs Our Model*

- Question: Why not Traditional Pattern Matching Algorithm (KMP) and Suffix Tries ?
- Almost all the filters at MIT Spam Conference Jan 2004, were Naïve Bayesian Filters
- Naïve Bayesian Filters have independence assumption for events for ex:

  "click here to buy cheap software " probability of occurrence of *"buy"* is assumed to be independent of probability of occurrence of *"click"* or *"cheap"*
- But probabilities of occurrence of these words together are highly related
- Proposed a Markov Random Field Model where occurrence of one word is dependent on the occurrence of other words in the vicinity, implemented and tested in CRM114
- Accuracy and Performance is higher than Paul Graham's Bayesian Filter Model

Shalendra Chhabra , William S. Yerazunis, and Christian Siefkes.* **"Spam Filtering using a Markov Random Field Model with Variable Weighting Schemas". *In Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM '04), Brighton UK, November 2004.*

Borrowed Idea from Computer Vision

- A Site represents a point or region in Euclidean space
- A Label is an event that may happen to a site for ex: In edge detection, the label set is

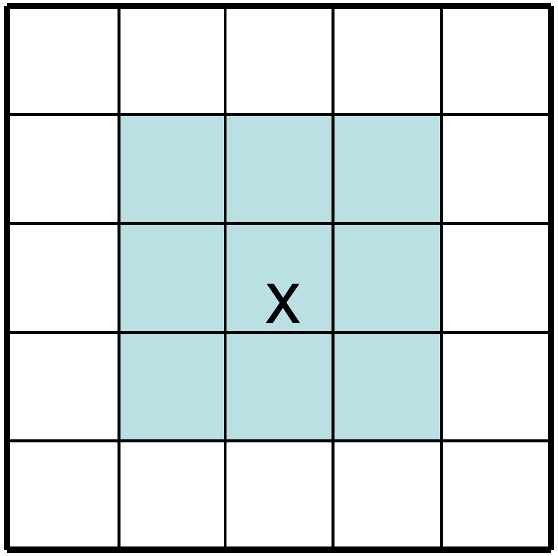
  L = {edge,non-edge}
- Let F = {F_1, F_2, …F_m} be a family of random variables on the discrete set of sites S, in which each random variable F_i takes the value f_i in the discrete label set L

  The family F is called a <u>Random Field</u>
- $P(F = f) = P(F_1 = f_1, F_2 = f_2, F_3=f_3 \ldots, F_m=f_m)$ denotes a joint event

Neighborhood System

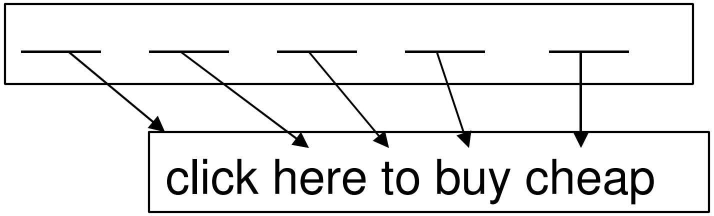
- The Sites in S are related to one another via a Neighborhood System. A Neighborhood System for a site X denotes the set of sites surrounding X
- Any F is said to be a MRF on S with respect to a neighborhood N iff:

1. $P(f) > 0$; *(positivity)*
2. $P(f_i | f_{S-\{i\}}) = P(f_i | f_{N_i})$ *(Markovianity)*

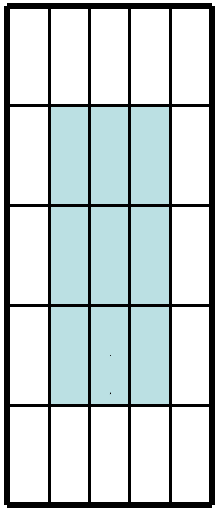


Analogy with Spam Text

A Site in the context of spam classification refers to *relative position* of word in a sequence
And a Label maps to *word values*



click here to buy cheap



Assigning Weights to These Features

- Sequence ABC has 8 subsequences including empty sequence and itself:

  {A, B, C, A_C, BC, AB, ABC, 0}.

- Idea: Weight of Feature with n terms in the sequence should be greater than combined weight of all Features of length less than n:

$$W(n) > \sum_{k=1}^{n-1} \left(\binom{n}{k} \times W(k) \right)$$

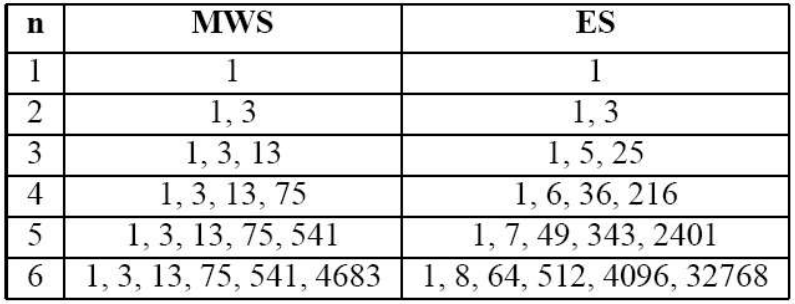
Weighting Schemes

Minimum Weighting Schemes

$$W(n) = \sum_{k=1}^{n-1} \left(\binom{n}{k} \times W(k) \right) + 1.$$

Exponential Scheme

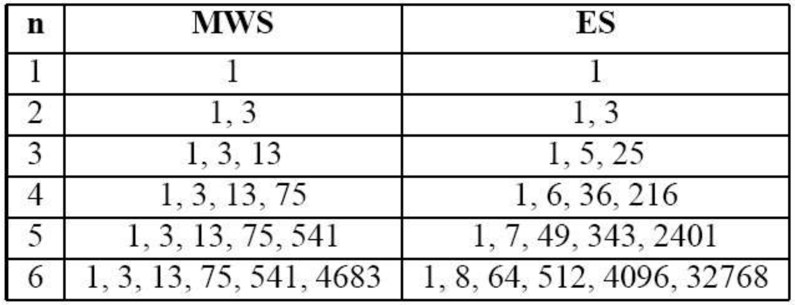
$$base^{n-1} > \sum_{k=1}^{n-1} \left(\binom{n}{k} \times base^{k-1} \right)$$



n	MWS	ES
1	1	1
2	1, 3	1, 3
3	1, 3, 13	1, 5, 25
4	1, 3, 13, 75	1, 6, 36, 216
5	1, 3, 13, 75, 541	1, 7, 49, 343, 2401
6	1, 3, 13, 75, 541, 4683	1, 8, 64, 512, 4096, 32768

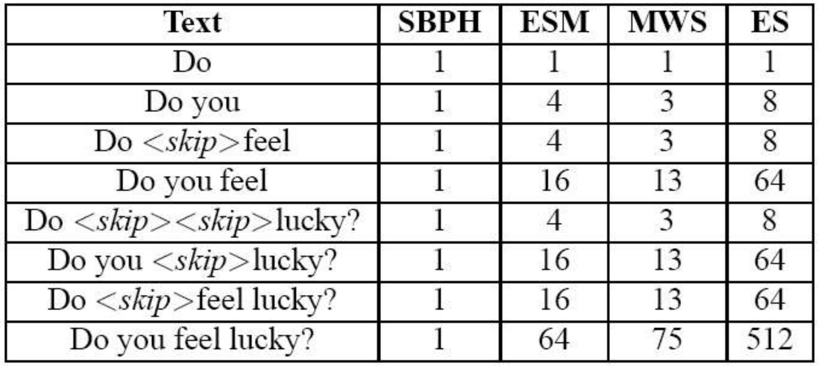
Table 1. Minimum & Exponential Weightings

Example Subphrases and Models Tested



n	MWS	ES
1	1	1
2	1, 3	1, 3
3	1, 3, 13	1, 5, 25
4	1, 3, 13, 75	1, 6, 36, 216
5	1, 3, 13, 75, 541	1, 7, 49, 343, 2401
6	1, 3, 13, 75, 541, 4683	1, 8, 64, 512, 4096, 32768

Table 1. Minimum & Exponential Weightings



Text	SBPH	ESM	MWS	ES
Do	1	1	1	1
Do you	1	4	3	8
Do <*skip*>feel	1	4	3	8
Do you feel	1	16	13	64
Do <*skip*><*skip*>lucky?	1	4	3	8
Do you <*skip*>lucky?	1	16	13	64
Do <*skip*>feel lucky?	1	16	13	64
Do you feel lucky?	1	64	75	512

SBPH: 1,1,1,1,1

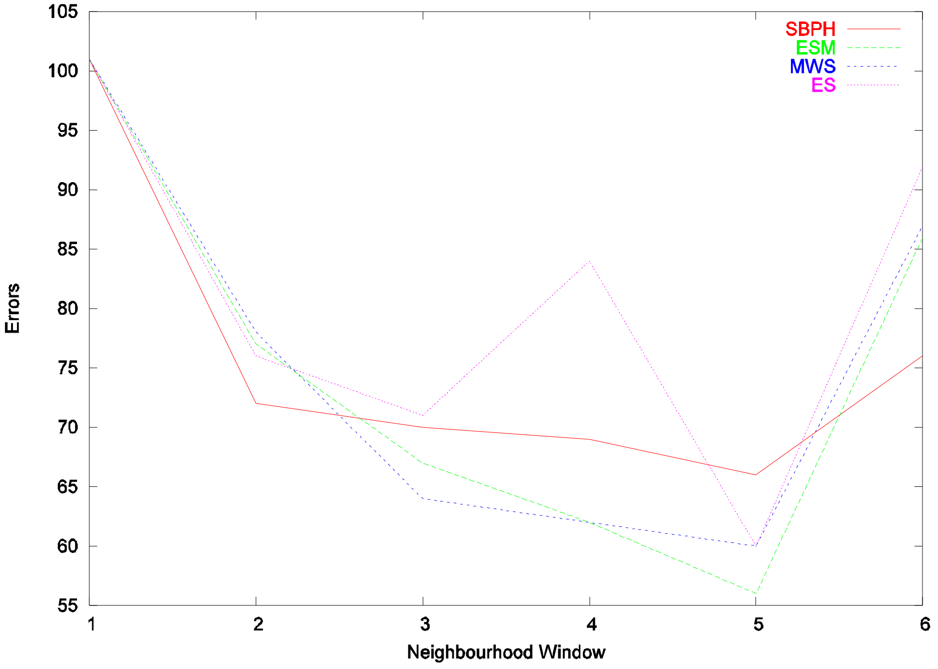
ESM ($2^{2(n-1)}$): 1,4,16,64

MRF Model for Spam

- All incoming email is broken in features
- *A random class function C is defined C:Omega -> {spam,nonspam}*
- $$P(spam|F_i) = \frac{P(F_i|spam)P(spam)}{(P(F_i\ |spam)P(spam)+P(\ F_i|ham)P(ham)\ )}$$
- *Local Formula for* $P(F_i|spam)$ ***
- *The output* $P(spam|F_i)$ *becomes* $P(spam)$ *for the feature* F_{i+1}

If $P(spam|F_n)$ is higher than $P(ham|F_n)$, email is tagged as “spam”

Results with MRF Model for Spam Filtering



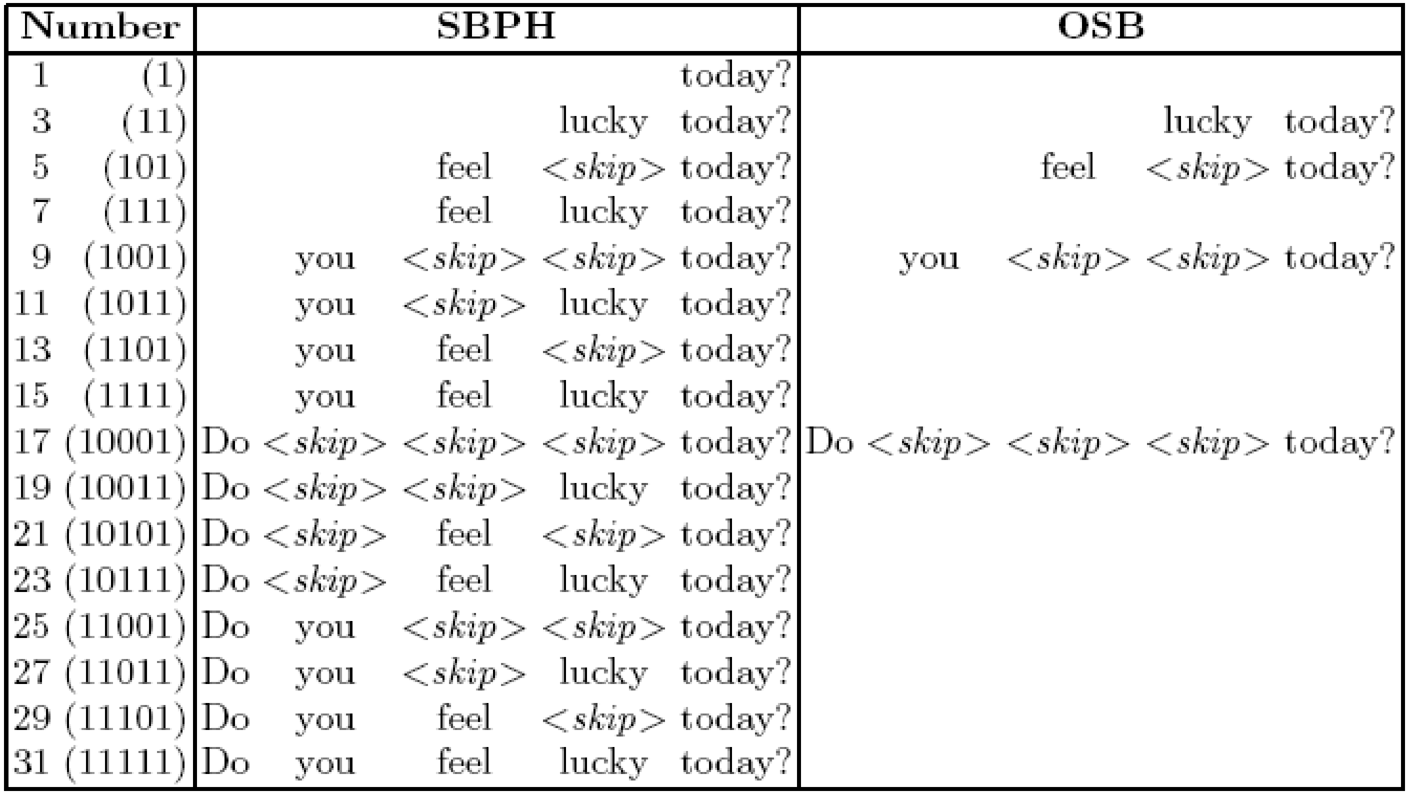
Winnow Algorithm and Orthogonal Sparse Bigrams**

- Winnow is a statistical but non probabilistic algorithm i.e. it computes score and not probability
- It keeps n dimensional weight vector for each class c, i.e. $w^c=(w^c_{1,}\ w^c_2 ,\ldots w^c_m)$, where w^c_i is the weight of the i^{th} feature for class c
- The algorithm returns 1 for a class iff the summed weights for all active features surpass a predefined threshold

*** Christian Siefkes, Fidelis Assis, Shalendra Chhabra and William S. Yerazunis.* **Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering.** *Lecture Notes in Computer Science. Springer, 2004*, Springer Verlag

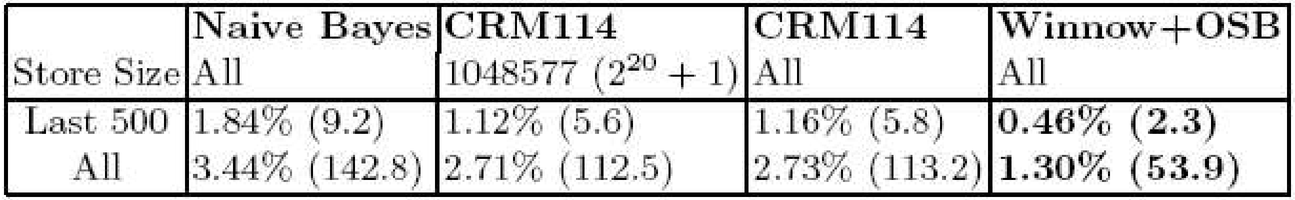
Expressivity of Features

Table 2. Features Generated by SBPH and OSB



Number		SBPH					OSB				
1	(1)					today?					
3	(11)				lucky	today?				lucky	today?
5	(101)			feel	<*skip*>	today?			feel	<*skip*>	today?
7	(111)			feel	lucky	today?					
9	(1001)		you	<*skip*>	<*skip*>	today?		you	<*skip*>	<*skip*>	today?
11	(1011)		you	<*skip*>	lucky	today?					
13	(1101)		you	feel	<*skip*>	today?					
15	(1111)		you	feel	lucky	today?					
17	(10001)	Do	<*skip*>	<*skip*>	<*skip*>	today?	Do	<*skip*>	<*skip*>	<*skip*>	today?
19	(10011)	Do	<*skip*>	<*skip*>	lucky	today?					
21	(10101)	Do	<*skip*>	feel	<*skip*>	today?					
23	(10111)	Do	<*skip*>	feel	lucky	today?					
25	(11001)	Do	you	<*skip*>	<*skip*>	today?					
27	(11011)	Do	you	<*skip*>	lucky	today?					
29	(11101)	Do	you	feel	<*skip*>	today?					
31	(11111)	Do	you	feel	lucky	today?					

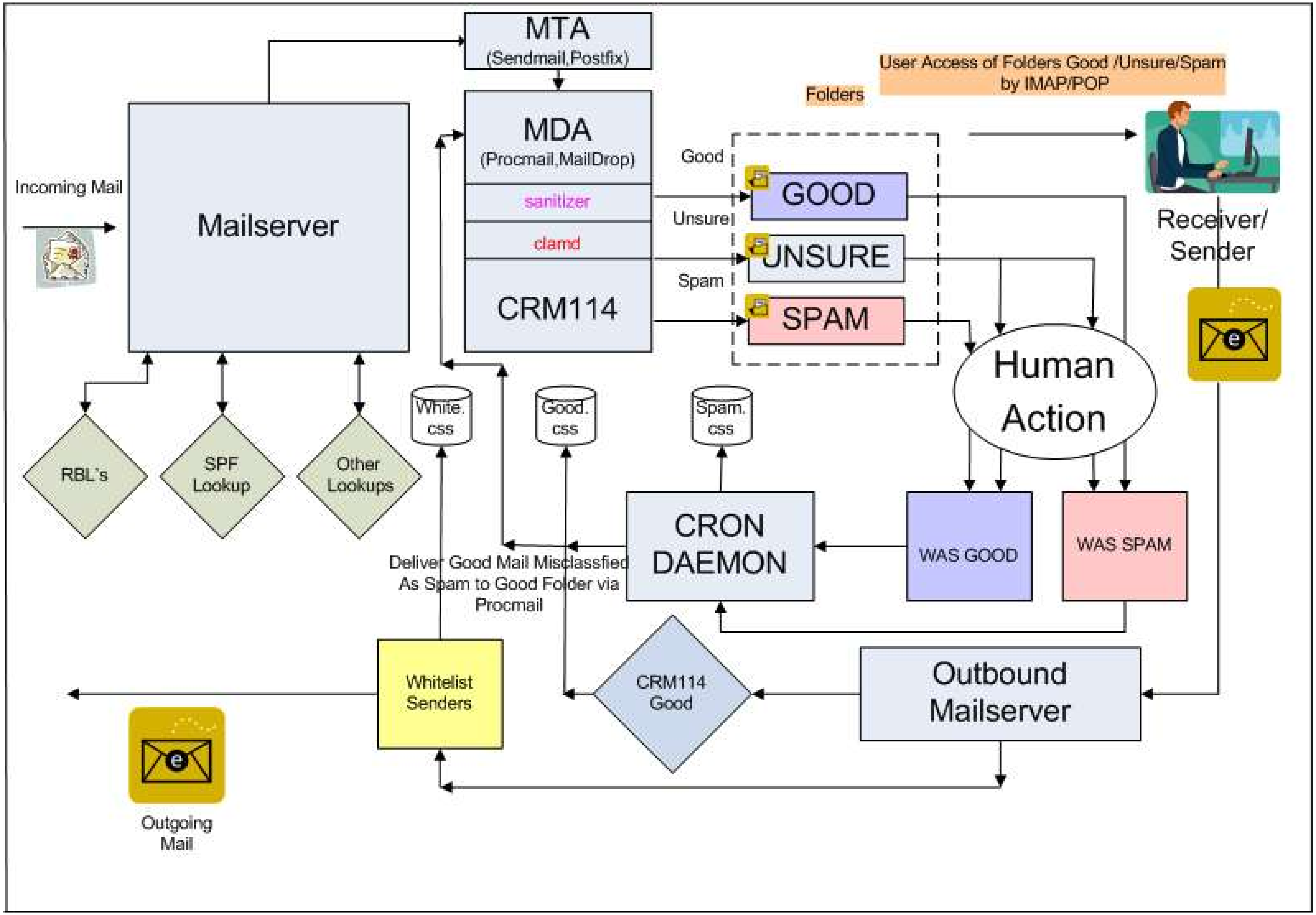
Comparison of Winnow, Naïve Bayes and CRM114 MRF Model



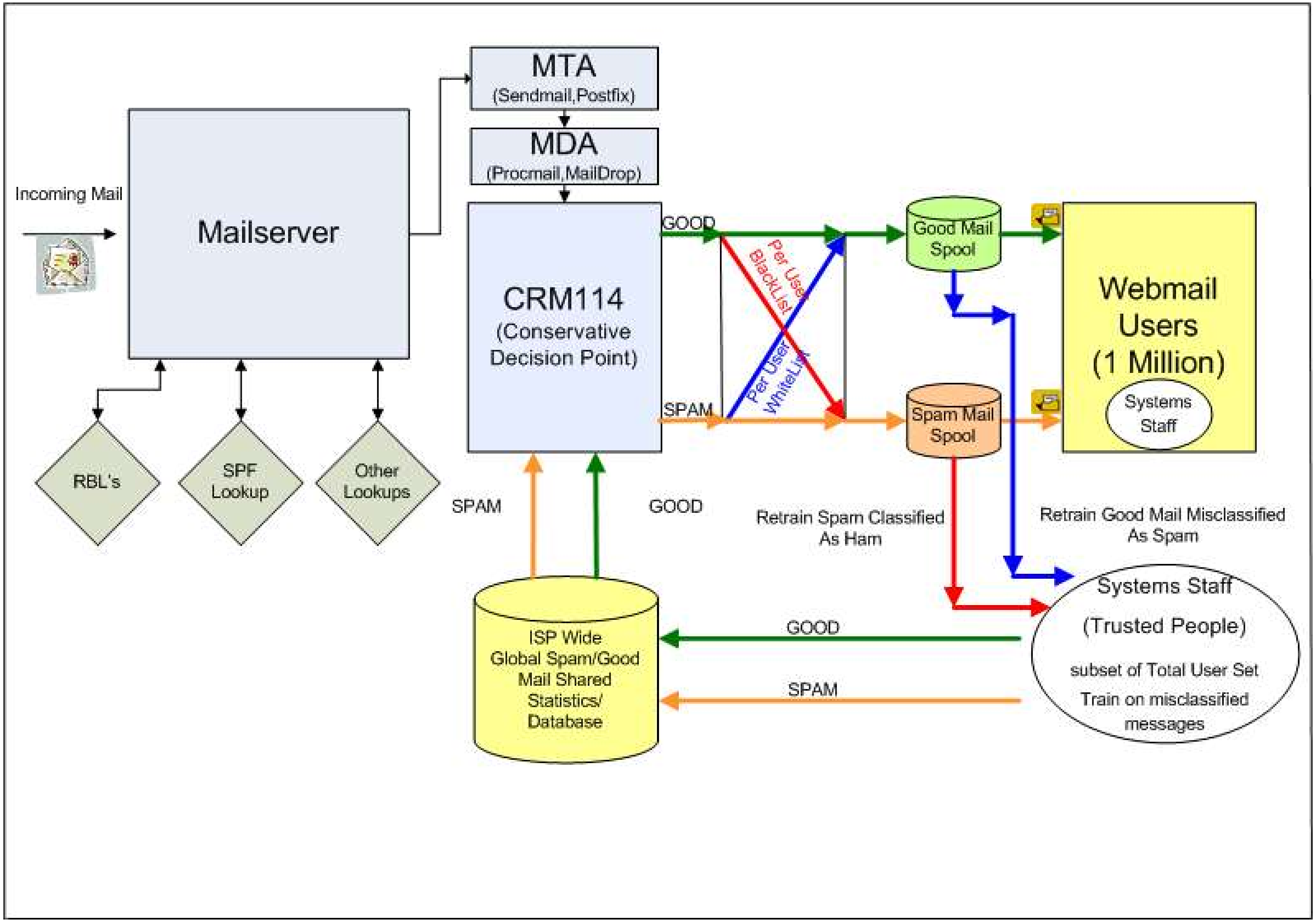
	Naive Bayes	CRM114	CRM114	Winnow+OSB
Store Size	All	1048577 ($2^{20}+1$)	All	All
Last 500	1.84% (9.2)	1.12% (5.6)	1.16% (5.8)	**0.46% (2.3)**
All	3.44% (142.8)	2.71% (112.5)	2.73% (113.2)	**1.30% (53.9)**

Note that Error Rate is Halved and Computational Overhead is also reduced (retaining the expressivity)

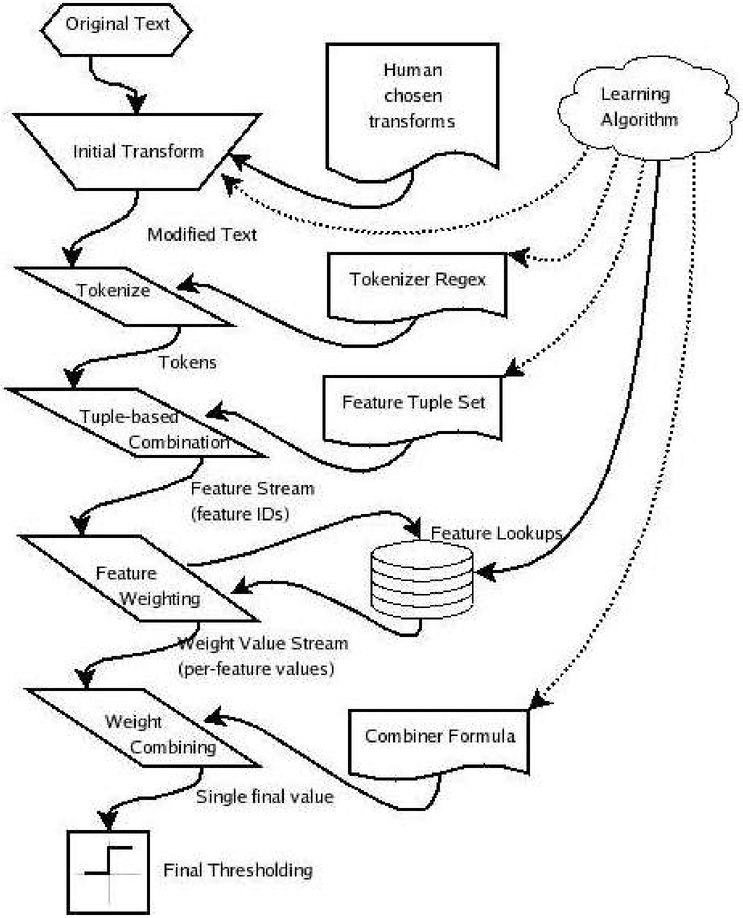
Implementing CRM114 At MailServers



CRM114 As Used By a Large ISP For Filtering more than 1 Million Client Email Accounts



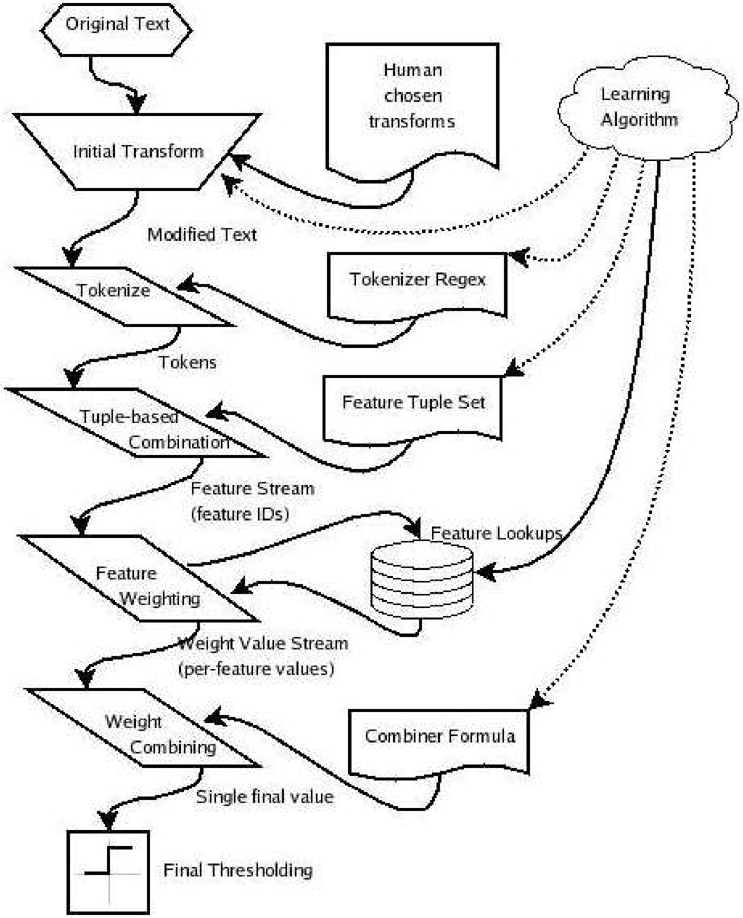
A Unified Model of Spam Filtration MIT Spam Conference, 2005



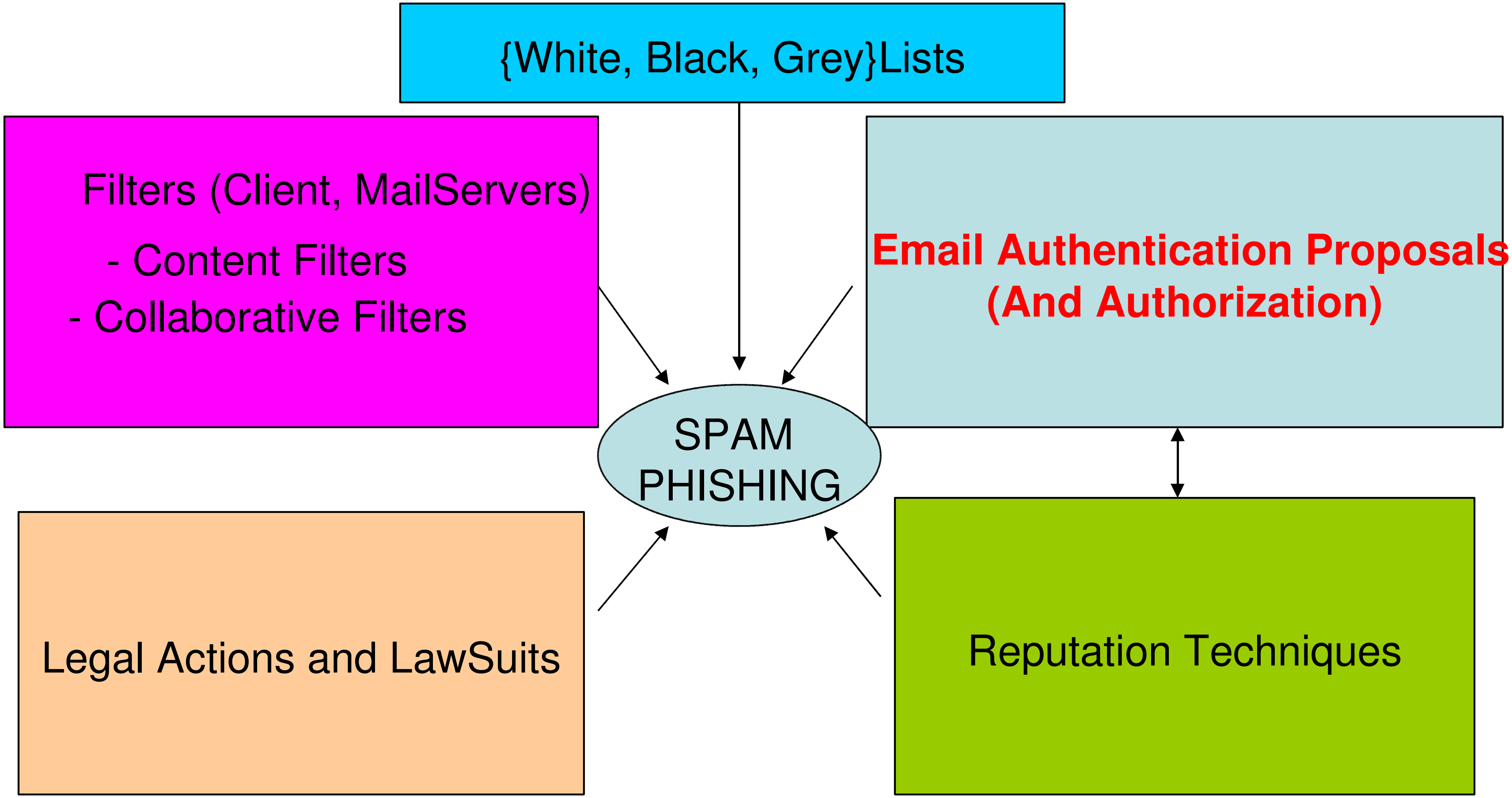
Pre Processing: Arbitrary Text to Text Transformation

- Character Set Folding / Case Folding
- Stopword Removal
- MIME Normalization / Base64 Decoding
- HTML Decommenting
  - Hypertextus Interruptus
- Heuristic Tagging
  - “FORGED_OUTLOOK_TAGS”
- Identifying Lookalike Transformations
  - ‘@’ instead of ‘a’, $ instead of S
  - Ex: V1agra

A Unified Model of Spam Filtration MIT Spam Conference, 2005



Tackling Spam and Phishing



{White, Black, Grey}Lists

Filters (Client, MailServers)

- Content Filters
- Collaborative Filters

Email Authentication Proposals (And Authorization)

SPAM PHISHING

Legal Actions and LawSuits

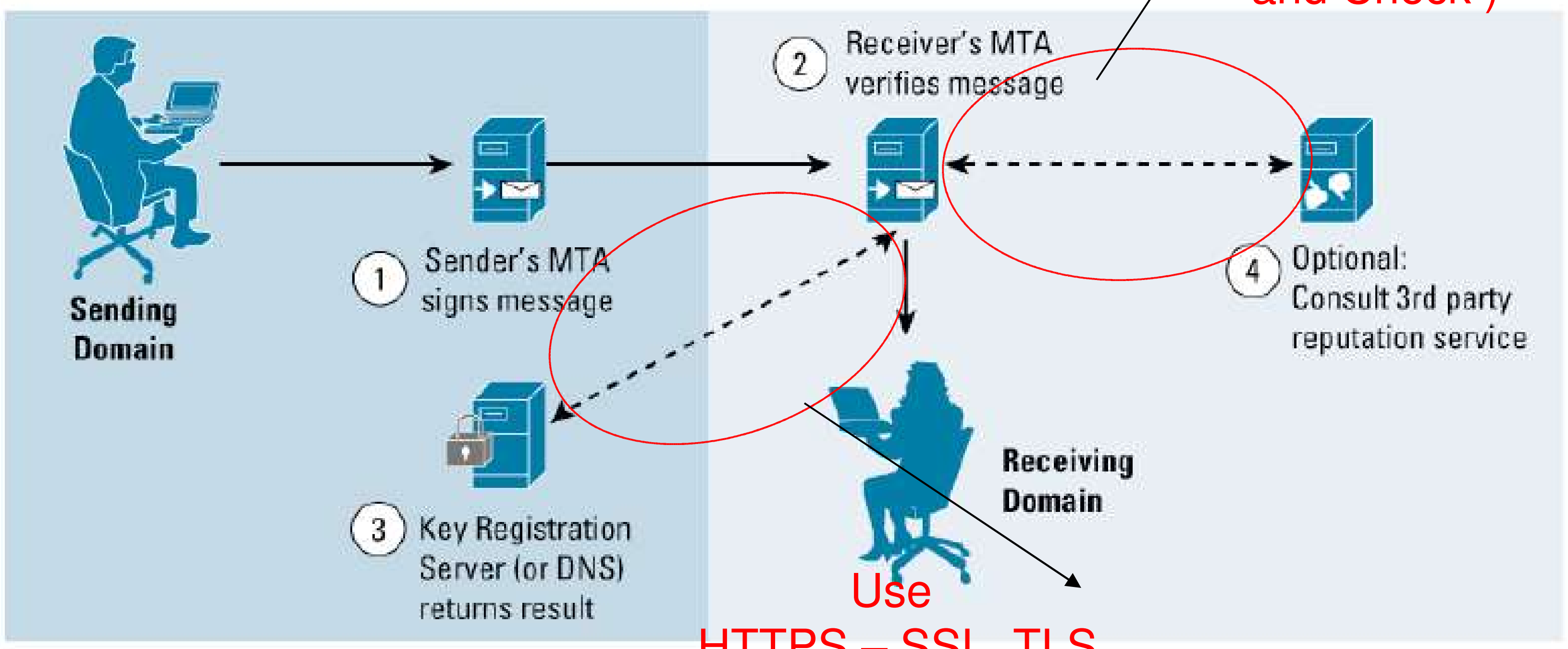
Reputation Techniques

Authentication and Authorization

- Authentication is the process of checking or verifying an entity using some form of integrity information such as an authorization policy.

Cisco's IIM
(now Merged with DK to form DKIM)

Typical Identified Internet Mail Message Flow


Sending Domain
1 Sender's MTA signs message
2 Receiver's MTA verifies message
3 Key Registration Server (or DNS) returns result
4 Optional: Consult 3rd party reputation service
Receiving Domain

Analysis of Reputation Attacks (Adapt *IDStealth, Shilling, PseudoSpoofing* and Check)

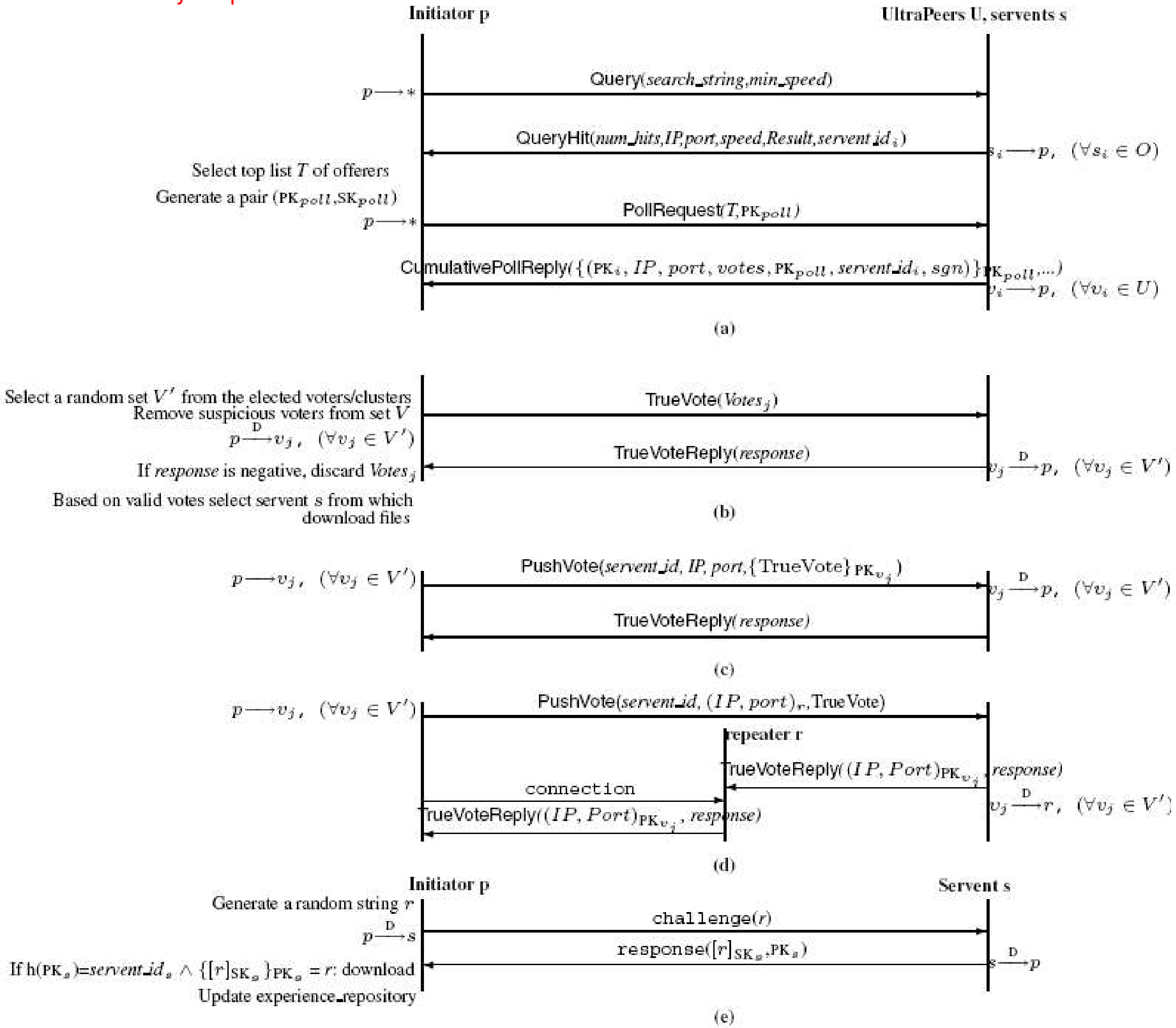
Use HTTPS – SSL, TLS

When using Third Party Reputation services we have to take countermeasures to probable attacks

With Email Authentication Systems What's Going to Happen Next?

- Spammers are adept at deploying sender authentication technologies for domains they are not forging
- Timeliness /reputation of domain in place
- Spammers will send from non-forged addresses (Blacklisting is the solution)

SupRup: Shalendra Chhabra etal. Italy – Spain Summer 2004



Initiator p
UltraPeers U, servents s
$p \longrightarrow *$
Query(search_string,min_speed)
QueryHit(num_hits,IP,port,speed,Result,servent_id$_i$)
$s_i \longrightarrow p, \ (\forall s_i \in O)$
Select top list T of offerers
Generate a pair (PK_{poll},SK_{poll})
$p \longrightarrow *$
PollRequest(T,PK_{poll})
CumulativePollReply($\{(PK_i, IP, port, votes, PK_{poll}, servent\_id_i, sgn)\}_{PK_{poll}}, \ldots$)
$v_i \longrightarrow p, \ (\forall v_i \in U)$
(a)
Select a random set V' from the elected voters/clusters
Remove suspicious voters from set V
$p \xrightarrow{D} v_j, \ (\forall v_j \in V')$
TrueVote(Votes$_j$)
TrueVoteReply(response)
If response is negative, discard Votes$_j$
$v_j \xrightarrow{D} p, \ (\forall v_j \in V')$
Based on valid votes select servent s from which download files
(b)
$p \longrightarrow v_j, \ (\forall v_j \in V')$
PushVote(servent_id, IP, port,$\{\mathrm{TrueVote}\}_{PK_{v_j}}$)
$v_j \xrightarrow{D} p, \ (\forall v_j \in V')$
TrueVoteReply(response)
(c)
$p \longrightarrow v_j, \ (\forall v_j \in V')$
PushVote(servent_id, $(IP, port)_r$, TrueVote)
repeater r
TrueVoteReply($(IP, Port)_{PK_{v_j}}$, response)
connection
$v_j \xrightarrow{D} r, \ (\forall v_j \in V')$
TrueVoteReply($(IP, Port)_{PK_{v_j}}$, response)
(d)
Initiator p
Servent s
Generate a random string r
$p \xrightarrow{D} s$
challenge(r)
response($[r]_{SK_s}$,PK_s)
$s \xrightarrow{D} p$
If h(PK_s)=servent_id$_s$ $\wedge$ $\{[r]_{SK_s}\}_{PK_s} = r$: download
Update experience_repository
(e)

Figure 1. SupRep protocol: query and poll (a), vote verification (b)-(d), and resource download (e)

Check Possibility of These Attacks when using Third Party Reputation Services with Email Authentication Systems

- *PseduoSpoofing*: Forging a great number of votes from a single node and giving them different IP addresses, and multiple IDs (TrueVoteConnection detects this)
- *Shilling*: Clique / Control over many servents affecting reputation (Scalability in Gnutella fixes this and repeaters for servents behind firewalls)
- *ID Stealth*: Malicious Servent replies with QueryReplies as if generated from genuine servents (Challenge Response detects this)

Lessons from the Past

- Always think about the possibility of DNS Poisoning in Caches, DNS Record Hijacking

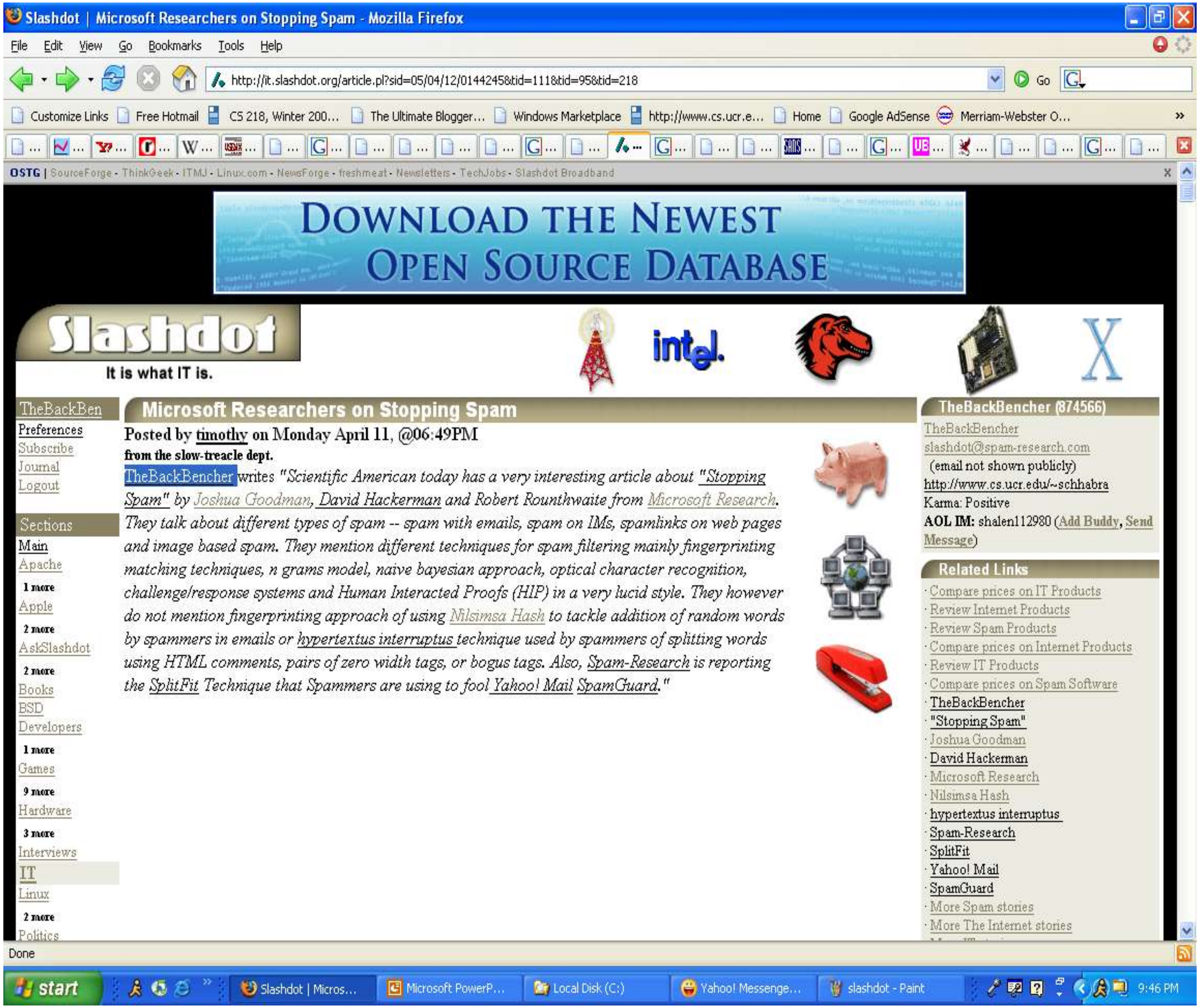
  (Refer ***Using*** *the* ***Domain Name System*** *for* ***System*** *Break-ins - Bellovin)*

- IP Spoofing Attacks
- DoS Attacks on Blacklists
- Some other Ideas ex: LOC record in DNS (Zombie Zones)

Other stuff We Are Doing

- Conducting a survey at UCR (population > 12000) – This will give us an idea how students and professors react to spam (will publish in *Nature*)
- Implementing Spam Filters at UCR MailServers in cooperation with the author of these filters and write effective guidelines for system administrators
- antispam.ucr.edu , antispam.cs.ucr.edu

On Slashdot


Slashdot | Microsoft Researchers on Stopping Spam - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://it.slashdot.org/article.pl?sid=05/04/12/0144245&tid=111&tid=95&tid=218
Go
Customize Links
Free Hotmail
CS 218, Winter 200...
The Ultimate Blogger...
Windows Marketplace
http://www.cs.ucr.e...
Home
Google AdSense
Merriam-Webster O...
OSTG | SourceForge - ThinkGeek - ITMJ - Linux.com - NewsForge - freshmeat - Newsletters - TechJobs - Slashdot Broadband
DOWNLOAD THE NEWEST OPEN SOURCE DATABASE
Slashdot
It is what IT is.
intel.
X
Done
start
Slashdot | Micros...
Microsoft PowerP...
Local Disk (C:)
Yahoo! Messenge...
slashdot - Paint
9:46 PM

TheBackBen
Preferences
Subscribe
Journal
Logout

Sections
Main
Apache
1 more
Apple
2 more
AskSlashdot
2 more
Books
BSD
Developers
1 more
Games
9 more
Hardware
3 more
Interviews
IT
Linux
2 more
Politics

Microsoft Researchers on Stopping Spam

Posted by timothy on Monday April 11, @06:49PM
from the slow-treacle dept.

TheBackBencher writes *"Scientific American today has a very interesting article about "Stopping Spam" by Joshua Goodman, David Hackerman and Robert Rounthwaite from Microsoft Research. They talk about different types of spam -- spam with emails, spam on IMs, spamlinks on web pages and image based spam. They mention different techniques for spam filtering mainly fingerprinting matching techniques, n grams model, naive bayesian approach, optical character recognition, challenge/response systems and Human Interacted Proofs (HIP) in a very lucid style. They however do not mention fingerprinting approach of using Nilsimsa Hash to tackle addition of random words by spammers in emails or hypertextus interruptus technique used by spammers of splitting words using HTML comments, pairs of zero width tags, or bogus tags. Also, Spam-Research is reporting the SplitFit Technique that Spammers are using to fool Yahoo! Mail SpamGuard."*

TheBackBencher (874566)

TheBackBencher
slashdot@spam-research.com
(email not shown publicly)
http://www.cs.ucr.edu/~schhabra
Karma: Positive
AOL IM: shalen112980 (Add Buddy, Send Message)

Related Links

- Compare prices on IT Products
- Review Internet Products
- Review Spam Products
- Compare prices on Internet Products
- Review IT Products
- Compare prices on Spam Software
- TheBackBencher
- "Stopping Spam"
- Joshua Goodman
- David Hackerman
- Microsoft Research
- Nilsimsa Hash
- hypertextus interruptus
- Spam-Research
- SplitFit
- Yahoo! Mail
- SpamGuard
- More Spam stories
- More The Internet stories

Finishing My Thesis

- Want to make my thesis a very important resource for Anti Spam Industry
- And Miles to go before I sleep….

  In order to contribute have to learn a lot with disciplined and ambitious instincts

Seek Your Blessings, Guidance, Comments and Criticism for becoming a Leader within next 5 years

Spam Free World?