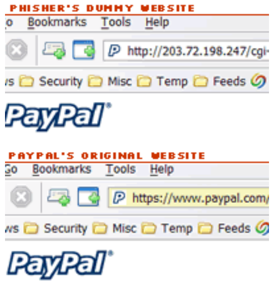


Its About You, Me and Every Netizen Because We've Got Spam and Phish!

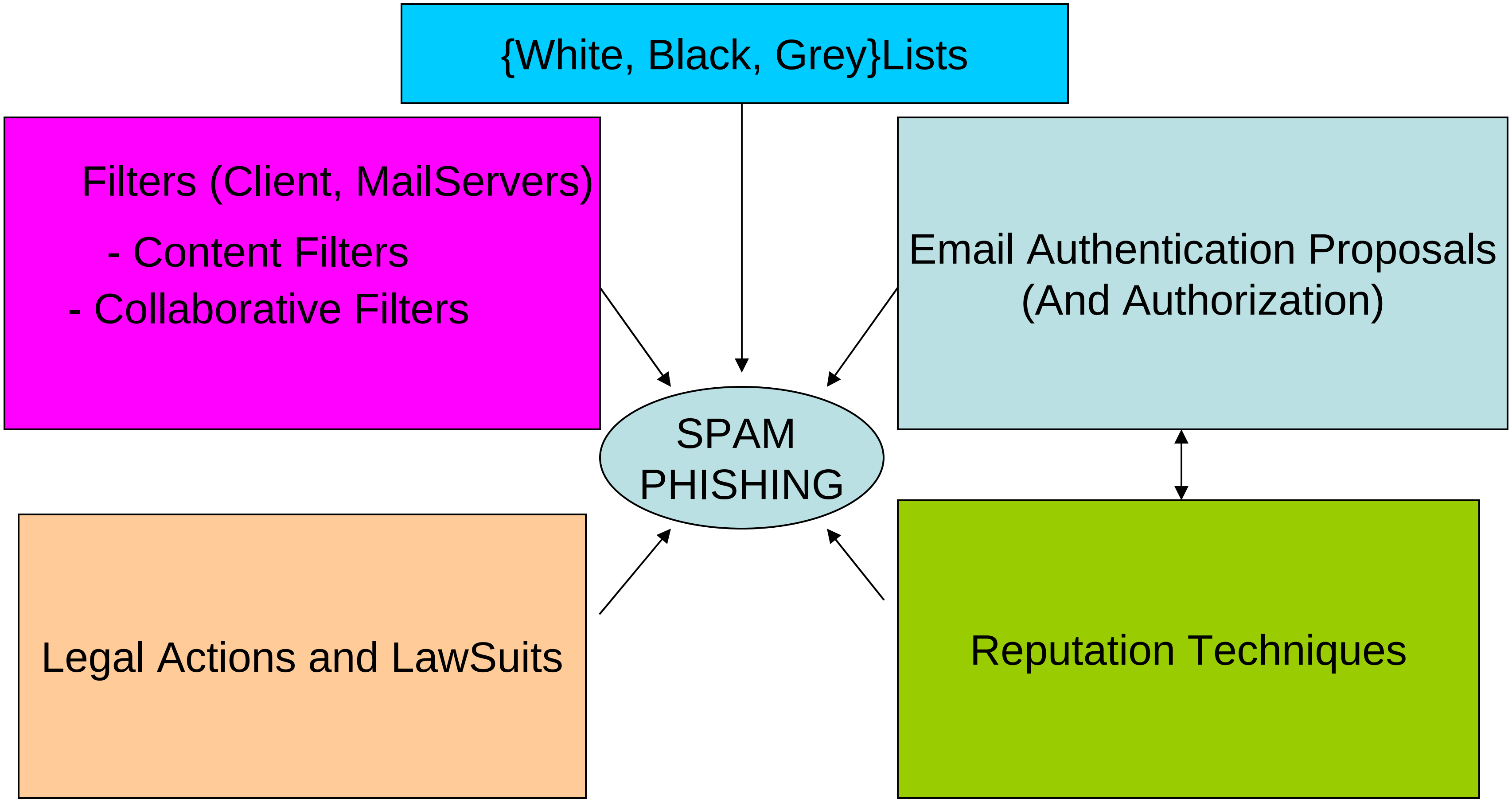
Shalendra Chhabra
University of California, Riverside
http://www.cs.ucr.edu/~schhabra
http://www.spam-research.com
schhabra@cs.ucr.edu

Venue: Cisco Systems

Its an Honor to Speak Here

- Thanks ☺ to Jim Fenton, Sanjay Pol, Shamim Pirzada and Jennifer Visaya for inviting me
- Regards to Cisco Anti Spam Team Members
- Congratulations to Cisco Systems for acquiring TopSpin

Tackling Spam and Phishing



{White, Black, Grey}Lists

Filters (Client, MailServers)

- Content Filters
- Collaborative Filters

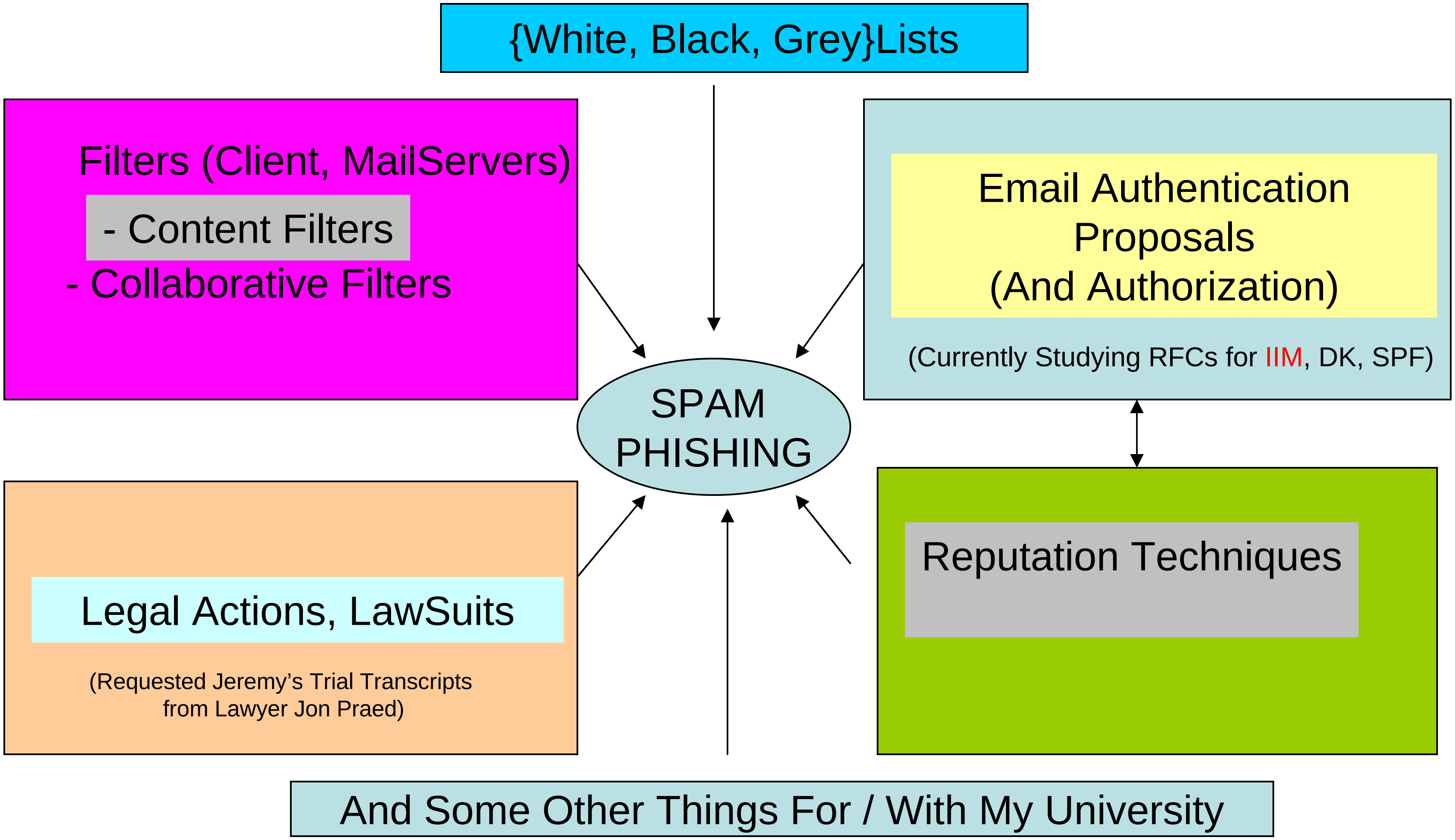
Email Authentication Proposals (And Authorization)

SPAM PHISHING

Legal Actions and LawSuits

Reputation Techniques

Masters Thesis On Tackling Spam and Phishing



{White, Black, Grey}Lists

Filters (Client, MailServers)

- Content Filters
- Collaborative Filters

Email Authentication Proposals (And Authorization)

(Currently Studying RFCs for IIM, DK, SPF)

SPAM PHISHING

Legal Actions, LawSuits

(Requested Jeremy's Trial Transcripts from Lawyer Jon Praed)

Reputation Techniques

And Some Other Things For / With My University

Motivation and How Did it All Start?

- September 2003 - Once was thinking for a Class Project and got spam, Clicked => Anti Spam
- Heard about MIT Spam Conference, January 2004
- January 2004 - Went up to attend MIT Spam Conference on my own, was a backseat audience
- Spam Conference 2004 - Found some errors in one presentation
- June 2004 - Proposed my Own Model and presented in UK
- 2005 spoke at MIT Spam Conference ☺ on a Unified Model of Spam Filtration

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

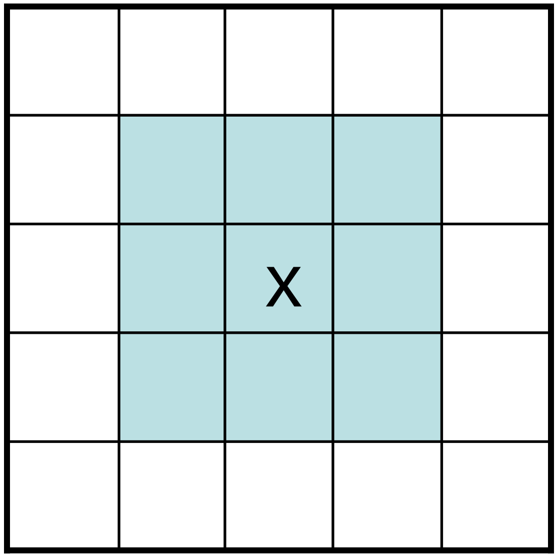
  The family F is called a <u>Random Field</u>
- $P(F = f) = P(F_1 = f_1, F_2 = f_2, F_3=f_3.\ldots,F_m=f_m)$ denotes a joint event

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:

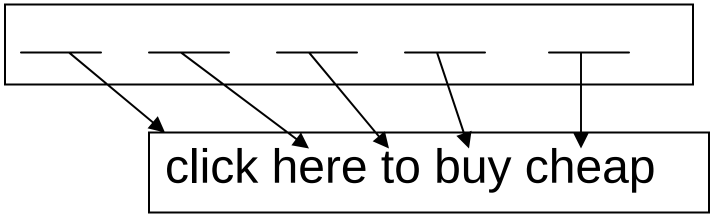
1. $P(f) > 0$; (positivity)

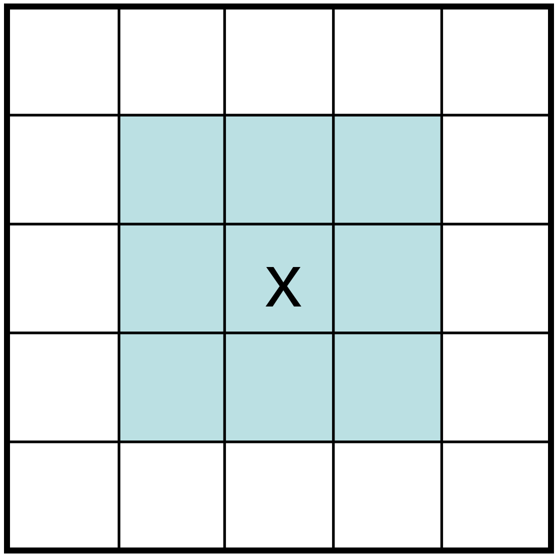
2. $P(f_i|f_{S-\{i\}}) = P(f_i|f_{N_i})$ (Markovianity)



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*





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:

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

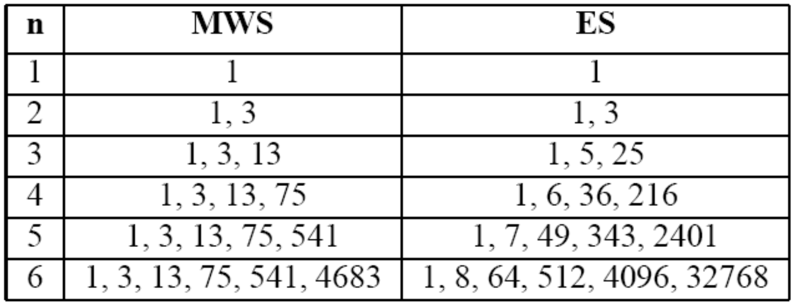
Weighting Schemes

Minimum Weighting Schemes

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

Exponential Scheme

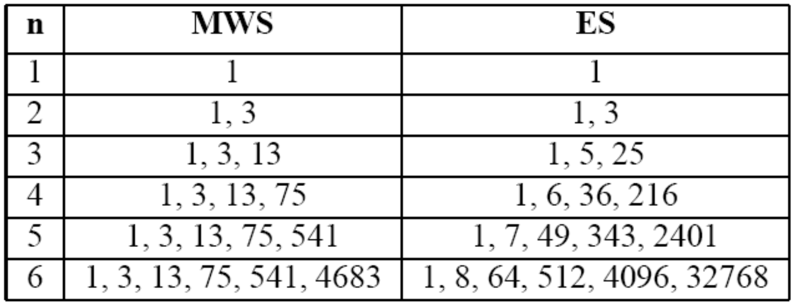
$$base^{n-1} > \sum_{k=1}^{n-1} \left(\binom{n}{k} \times base^{k-1}\right)$$



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

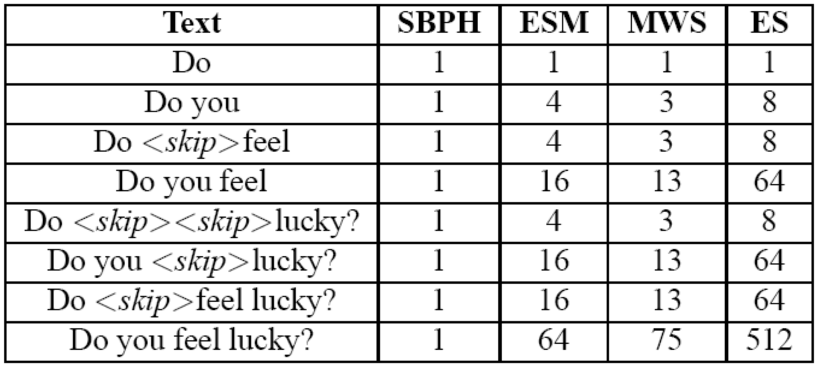
Table 1. Minimum & Exponential Weightings

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}*

- *$P(spam|F_i) = P(F_i|spam)P(spam)$*

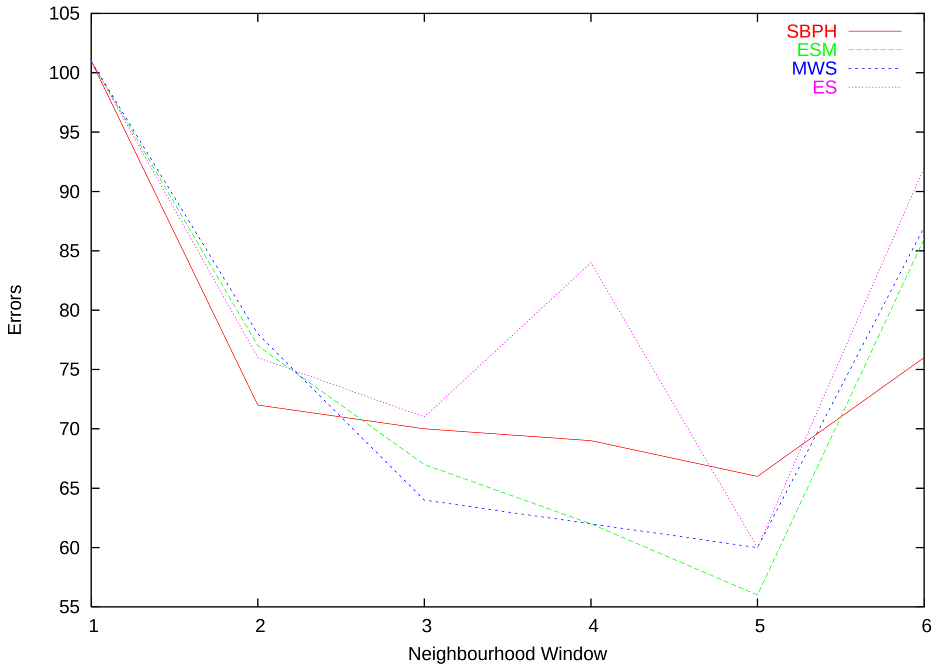
  $(P(F_i\ |spam)P(spam)+P(\ F_i|ham)P(ham)\ )$

- *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

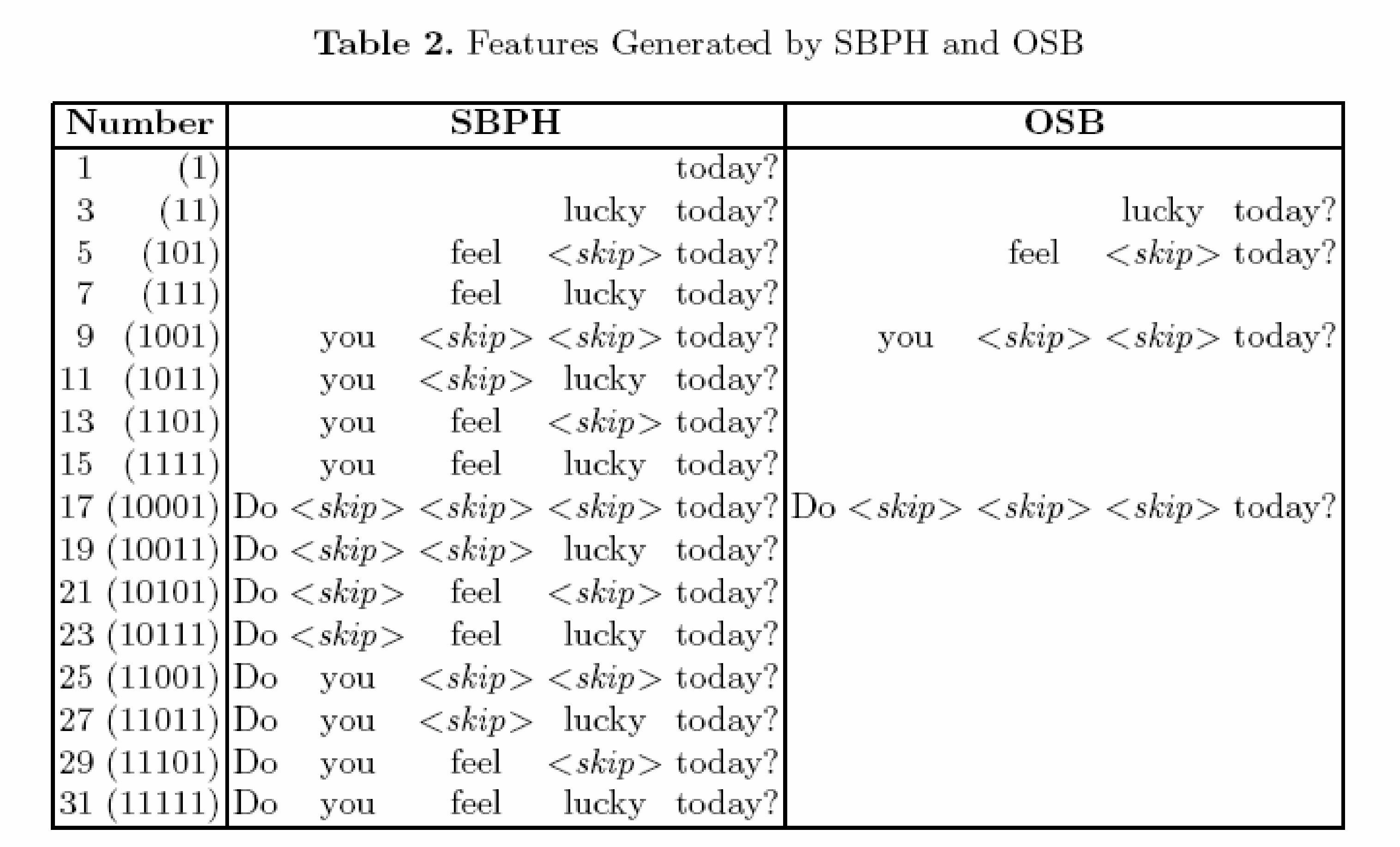
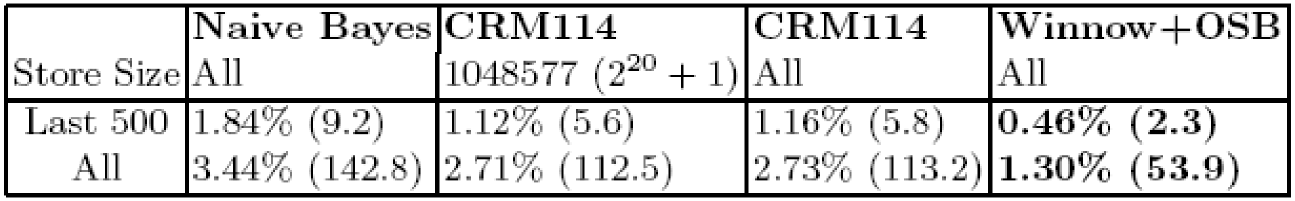


Table 2. Features Generated by SBPH and OSB

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?					

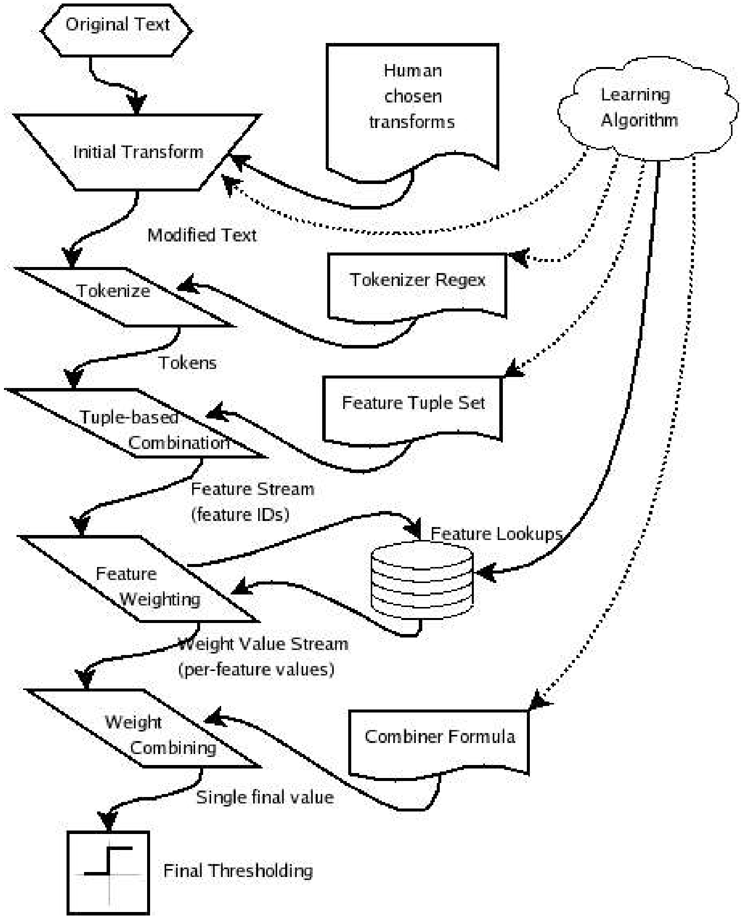
Comparison of Winnow, Naïve Bayes and CRM114 MRF Model



	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)

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

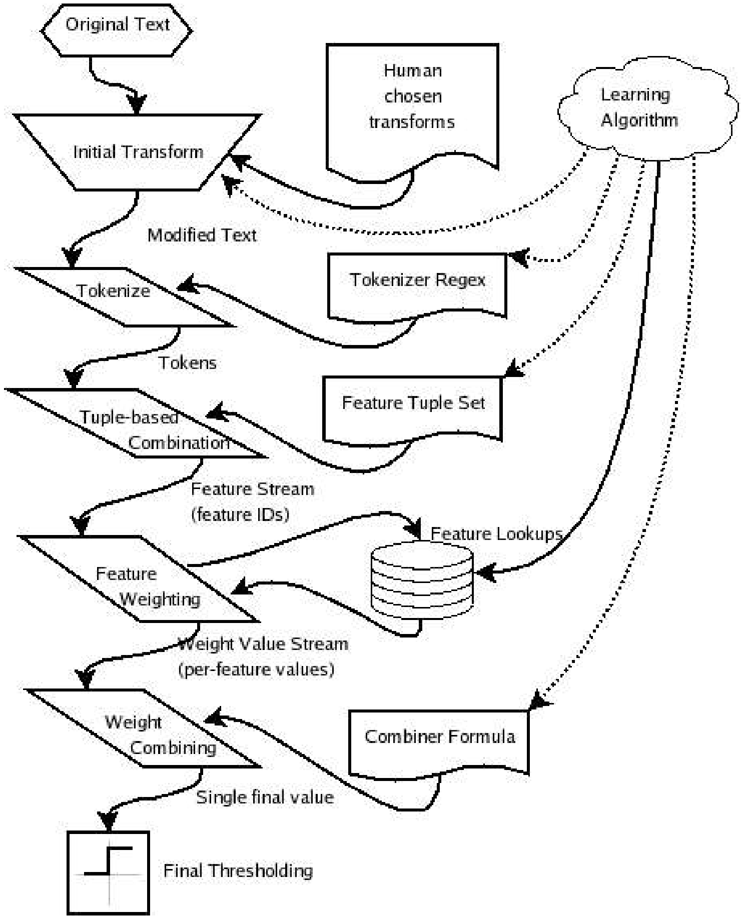
  Hypertextus Interruptus
- Heuristic Tagging

  “FORGED_OUTLOOK_TAGS”
- Identifying Lookalike Transformations

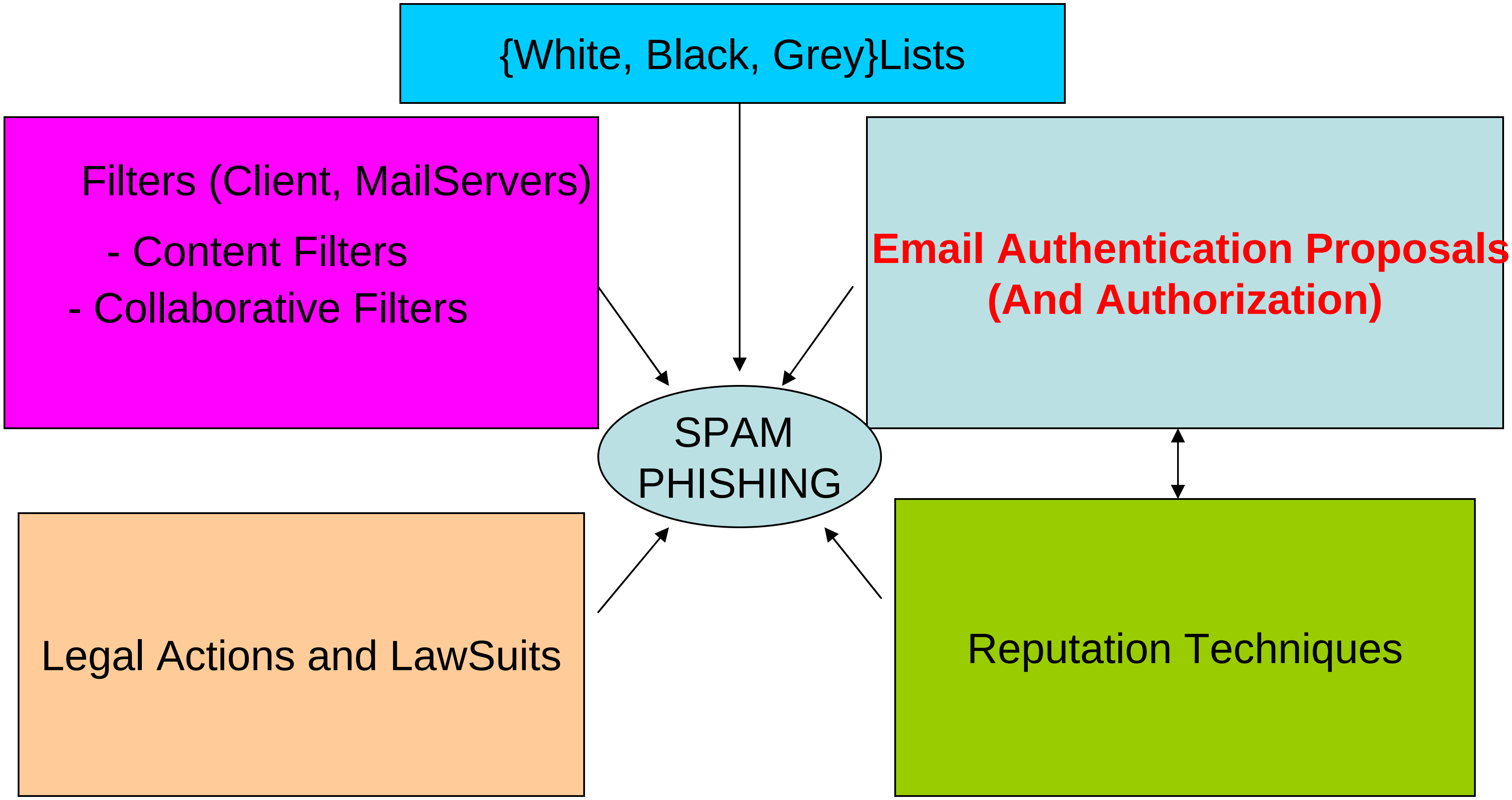
  ‘@’ instead of ‘a’, $ instead of S

  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

04/18/2005

Shalendra Chhabra
(Its About You, Me and Every Netizen -Limited Distribution)

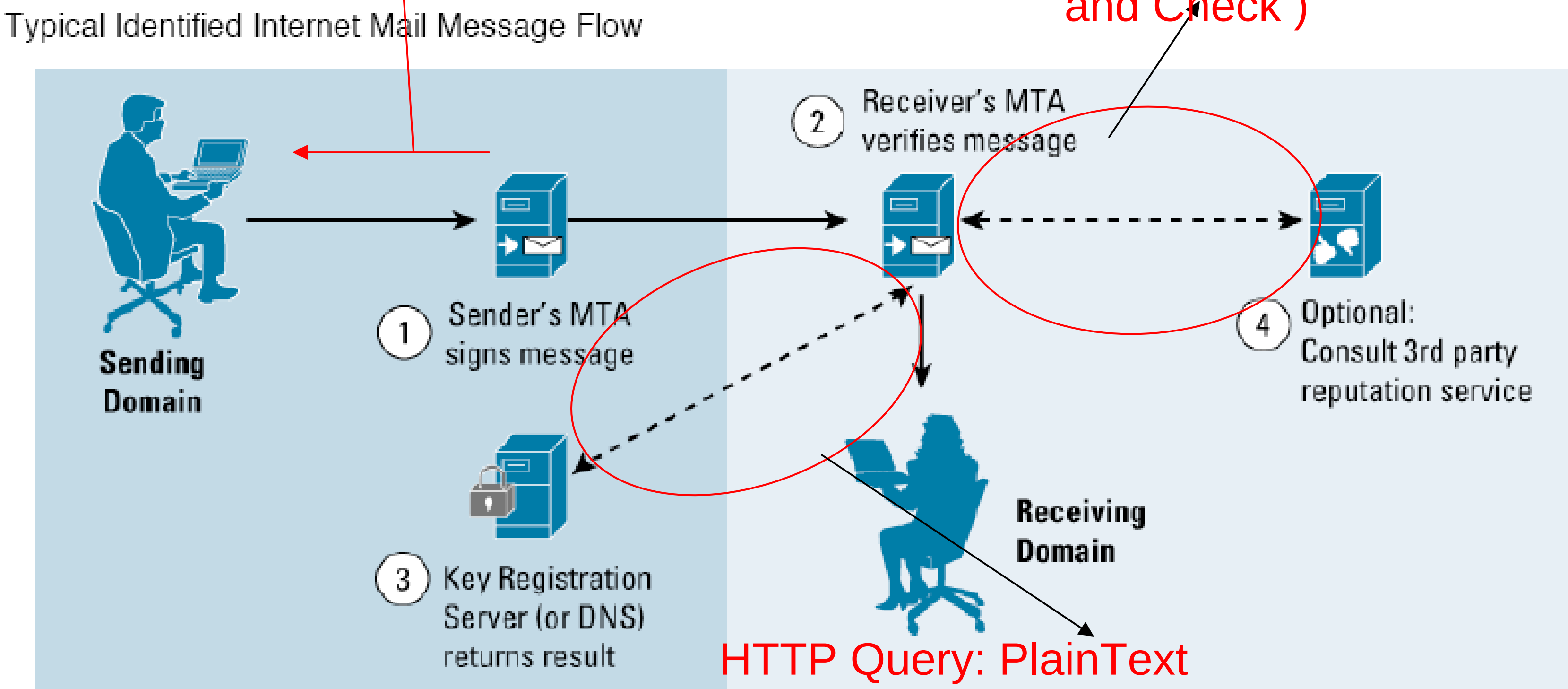
Authentication and Authorization

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

Cisco's IIM

Sending Domain checks if the Source is allowed to send Mail using its Domain

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


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

HTTP Query: PlainText

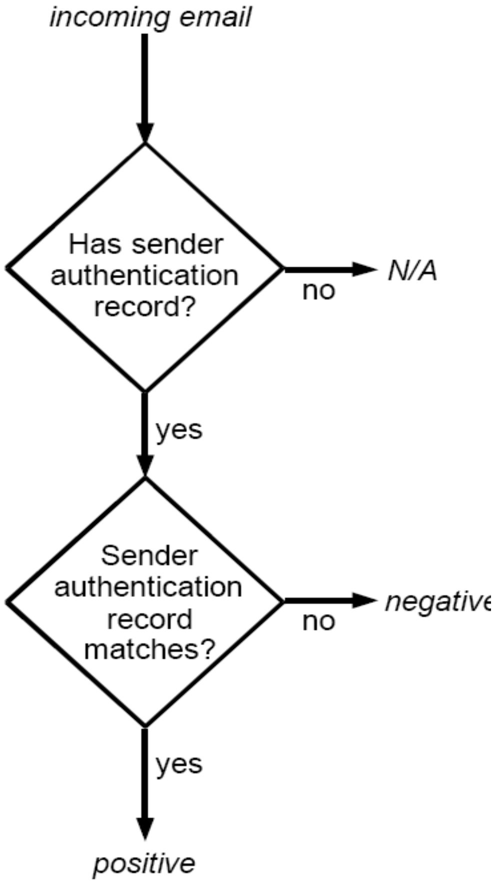
HTTPS – SSL, TLS?

Response Format with values not mentioned in RFC

(Locally Sensitive Hash) ex: Nilsimsa Hash?

State with Email Authentication Systems *

(John Graham Cumming)


incoming email
Has sender authentication record?
no
N/A
yes
Sender authentication record matches?
no
negative
yes
positive

Forged Message or False Negative

Use Bayesian Filter to Train (State, Output) ☺

Only sure when its positive: like whitelists

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

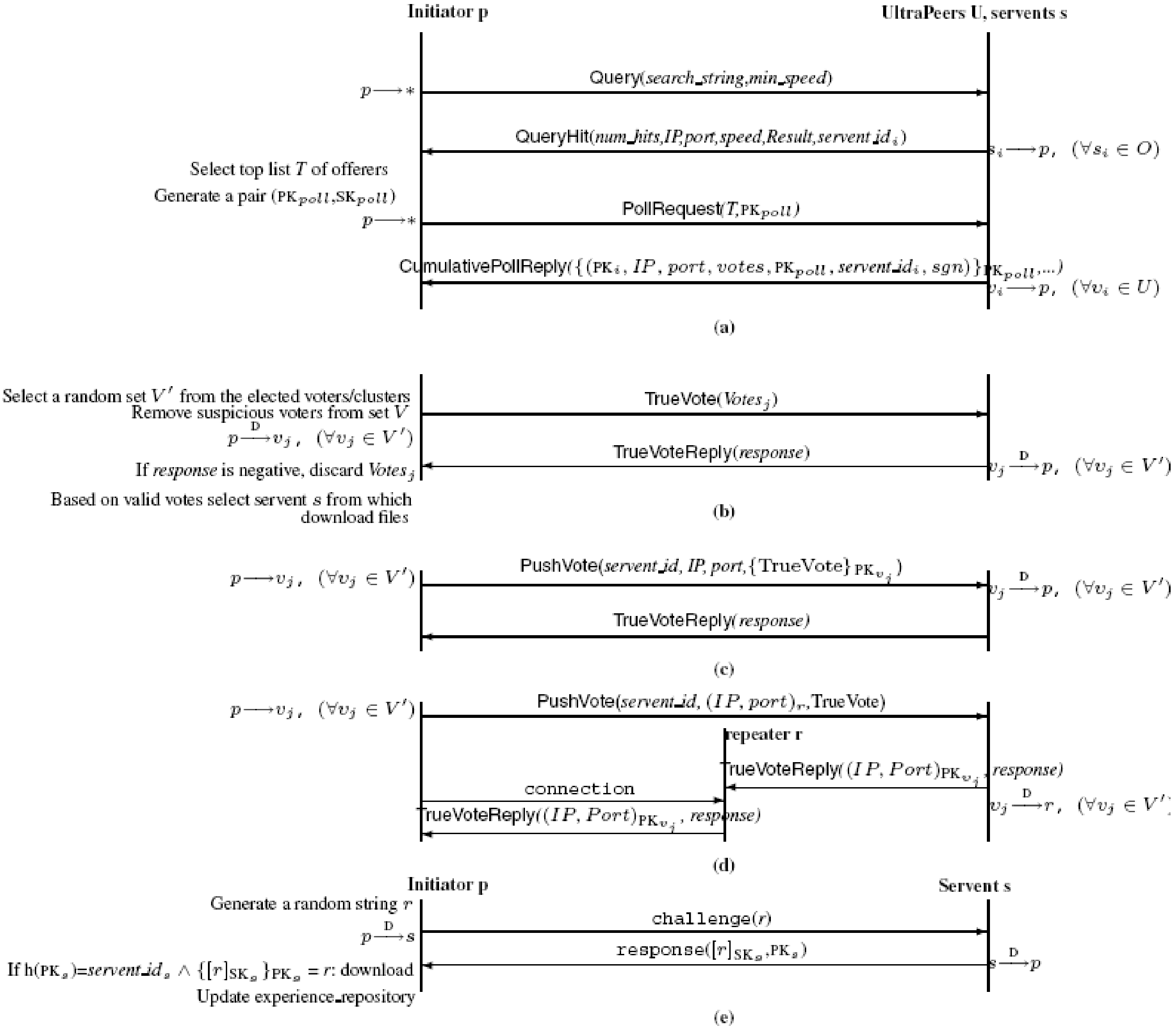

Initiator p
UltraPeers U, servents s
$p \longrightarrow *$
Query(search_string,min_speed)
QueryHit(num_hits,IP,port,speed,Result,servent_id$_i$)
$s_i \longrightarrow p, \ (\forall s_i \in O)$
Select top list T of offerers
Generate a pair (PK_{poll},SK_{poll})
$p \longrightarrow *$
PollRequest(T,PK_{poll})
CumulativePollReply($\{(\mathrm{PK}_i, IP, port, votes, \mathrm{PK}_{poll}, servent\_id_i, sgn)\}_{\mathrm{PK}_{poll}}$,...)
$v_i \longrightarrow p, \ (\forall v_i \in U)$
(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, $\{\mathrm{TrueVote}\}_{\mathrm{PK}_{v_j}}$)
$v_j \xrightarrow{D} p, \ (\forall v_j \in V')$
TrueVoteReply(response)
(c)
$p \longrightarrow v_j, \ (\forall v_j \in V')$
PushVote(servent_id, $(IP, port)_r$,TrueVote)
repeater r
TrueVoteReply($(IP, Port)_{\mathrm{PK}_{v_j}}$, response)
connection
TrueVoteReply($(IP, Port)_{\mathrm{PK}_{v_j}}$, response)
$v_j \xrightarrow{D} r, \ (\forall v_j \in V')$
(d)
Initiator p
Servent s
Generate a random string r
$p \xrightarrow{D} s$
challenge(r)
response($[r]_{\mathrm{SK}_s}$,PK_s)
$s \xrightarrow{D} p$
If h(PK_s)=$servent\_id_s$ $\wedge$ $\{[r]_{\mathrm{SK}_s}\}_{\mathrm{PK}_s} = r$: 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 great number of votes from a single node, giving them different IP addresses, and multiple IDs (TrueVoteConnection detects this)
- *Shilling*: Clique / Control over many servents affecting reputation (Scalability in Gnutella and repeaters for servents behind firewalls takes care of this)
- *ID Stealth*: Malicious Servent replies with QueryReplie's as if generated from genuine servents (Challenge Response detects this)

Lessons from the Past

- Always think about the possibility of DNS Poisoning in Caches (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)

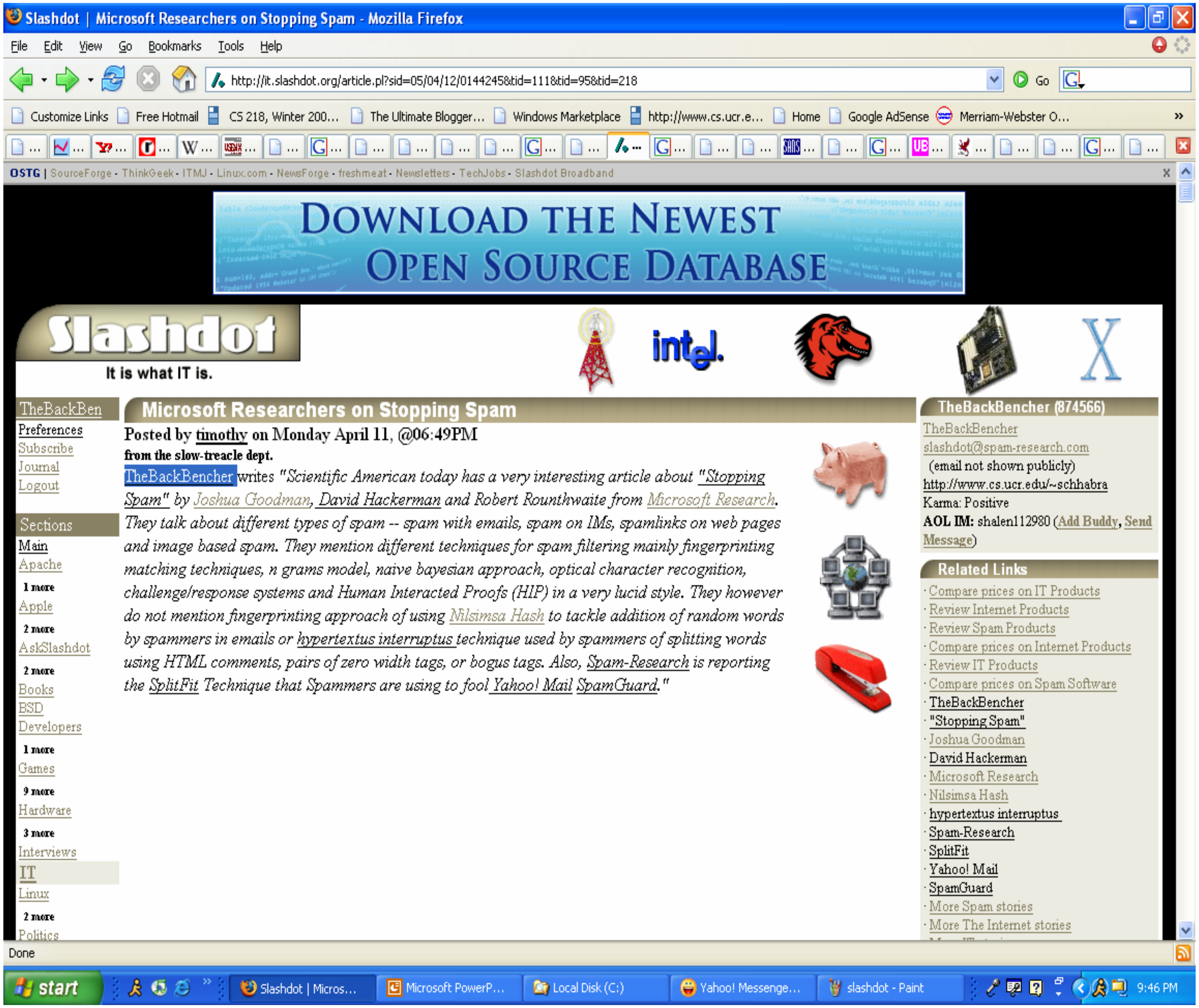
Other stuff I am doing

- Conducting a survey at UCR (population > 10000) – 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
- Yahoo Mail SpamGuard SplitFit Yaho޽o޴!

  (with Miles Libbey)
- A comment on Microsoft's Article on Slashdot

  (On Nilsimsha Hash and "Cmabirgde Uinersvtiy Sapm". , It was on Slashdot)

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.
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."*

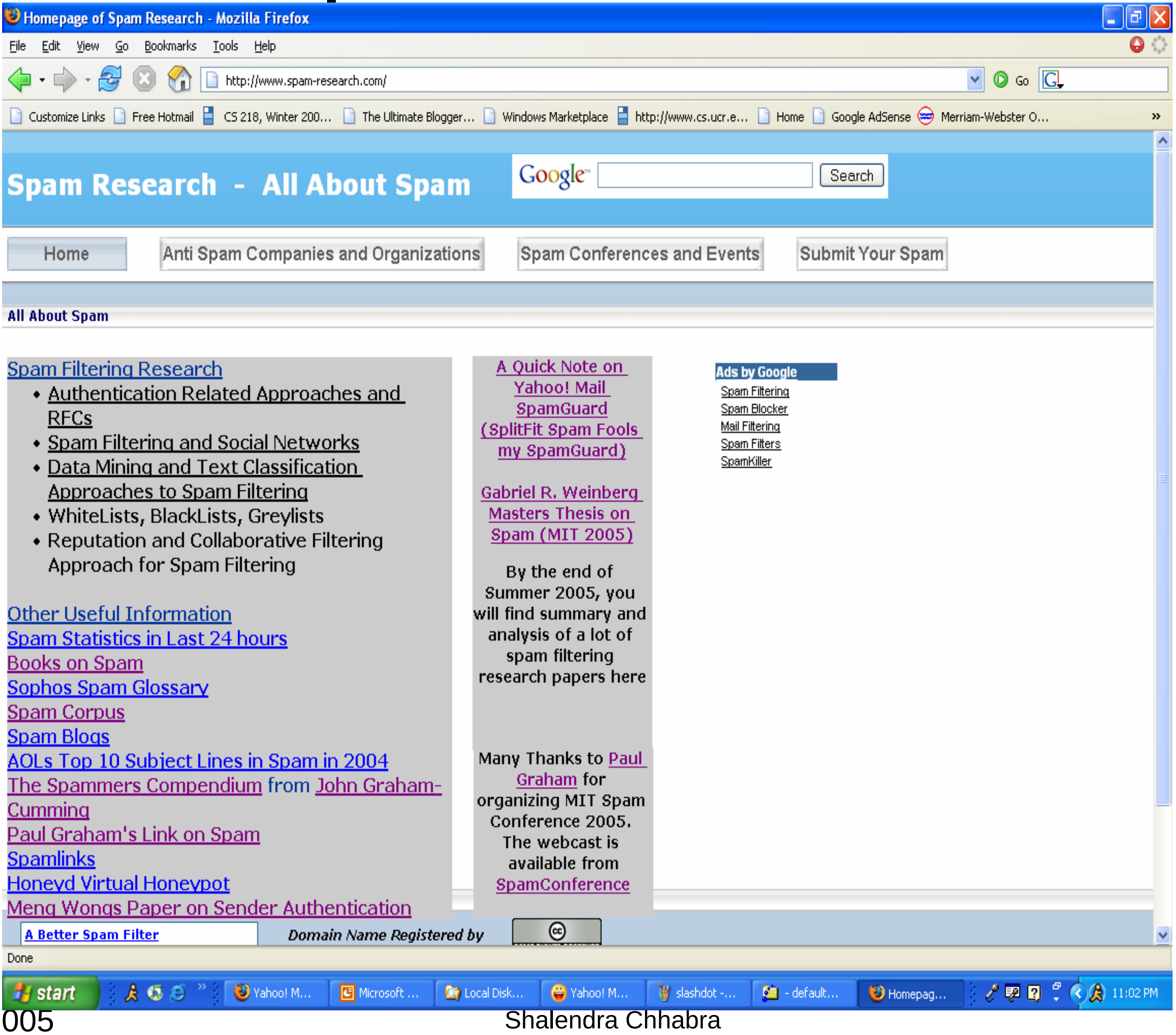
TheBackBencher (874566)

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

Spam-Research



Spam Research - All About Spam

Home | Anti Spam Companies and Organizations | Spam Conferences and Events | Submit Your Spam

All About Spam

Spam Filtering Research

- Authentication Related Approaches and RFCs
- Spam Filtering and Social Networks
- Data Mining and Text Classification Approaches to Spam Filtering
- WhiteLists, BlackLists, Greylists
- Reputation and Collaborative Filtering Approach for Spam Filtering

Other Useful Information
Spam Statistics in Last 24 hours
Books on Spam
Sophos Spam Glossary
Spam Corpus
Spam Blogs
AOLs Top 10 Subject Lines in Spam in 2004
The Spammers Compendium from John Graham-Cumming
Paul Graham's Link on Spam
Spamlinks
Honeyd Virtual Honeypot
Meng Wongs Paper on Sender Authentication

A Quick Note on Yahoo! Mail SpamGuard (SplitFit Spam Fools my SpamGuard)

Gabriel R. Weinberg Masters Thesis on Spam (MIT 2005)

By the end of Summer 2005, you will find summary and analysis of a lot of spam filtering research papers here

Many Thanks to Paul Graham for organizing MIT Spam Conference 2005. The webcast is available from SpamConference

Ads by Google
Spam Filtering
Spam Blocker
Mail Filtering
Spam Filters
SpamKiller

A Better Spam Filter

Domain Name Registered by

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 an Anti Spam Leader within next 5 years

Spam Free World?