Fighting Spam, Phishing and Email Fraud

Revolution is in the Cards

SPAM PHISHING EMAIL FRAUD
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

- E-mail Spam
- Messing Spam (SPIM)
- Newsgroup Spam
- Mobile Phone Spam
- Internet Telephony Spam
- Spamdexing
Email Spam Statistics in Last 24 Hours Reported to SpamCop

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 SECOND MORTGAGE 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:13772.html
saopaulopizzaria.info
Av. Leoncio de Magalhaes, 427
Sao Paulo - SP Brazil 02042-010

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 love 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 unmistakable Chuang Chou. But he didnt know.

Note the stories/legitimate text to fool spam filters

09/28/2005
University of California, Riverside

MS Thesis Defense - Fighting Spam, Phishing and Email Fraud
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 !!!!
Both Open Source (Filters) and Industry (Microsoft, Yahoo, AOL, Cisco etc.) Initiatives for Fighting Spam Are Underway
How Email System Works

Alice (MUA)

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

SMTP

smtp.a.org

mx.b.org
pop3.b.org

DNS

The Internet

mx.b.org

ns.b.org

Bob (MUA)

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

POP3
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

\[0.001\% \text{ of } 1 \text{ million } = 10 \text{ respondents}\]

• 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

- Filters (Client, MailServers)
  - Content Filters
  - Collaborative Filters
- Email Authentication Proposals (And Authorization)
- Reputation Techniques
- {White, Black, Grey}Lists
- SPAM PHISHING
- Legal Actions and LawSuits
So We Wrote a Masters Thesis
Fighting Spam, Phishing and Email Fraud

Filters (Client, MailServers)
- Content Filters
- Collaborative Filters

{White, Black, Grey}Lists

Email Authentication Proposals (And Authorization)

Legal Actions, LawSuits

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 Gunopoulos’ 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

Original Text

Initial Transform

Modified Text

Feature Generator

Feature Stream

Feature Lookups

Generator Params

Learning Algorithm

Weight Stream

Weight Combing

Weight Value Stream

Single Final Value (per-feature values)

Combiner Formula

Final Thresholding
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

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
  \[ L = \{\text{edge, non-edge}\} \]
- Let \( F = \{F_1, F_2, \ldots F_m\} \) 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 Random Field
- \( P(F = f) = P(F_1 = f_1, F_2 = f_2, F_3 = f_3, \ldots, F_m = f_m) \) 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. $P(f_i | f_{S \setminus \{i\}}) = P(f_i | f_N)$ (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:

\[ W(n) > \sum_{k=1}^{n-1} \binom{n}{k} \times W(k) \]
Weighting Schemes

Minimum Weighting Schemes

\[ W(n) = \sum_{k=1}^{n-1} \binom{n}{k} \times W(k) + 1. \]

Exponential Scheme

\[ \text{base}^{n-1} > \sum_{k=1}^{n-1} \binom{n}{k} \times \text{base}^{k-1} \]

<table>
<thead>
<tr>
<th>n</th>
<th>MWS</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1, 3</td>
<td>1, 3</td>
</tr>
<tr>
<td>3</td>
<td>1, 3, 13</td>
<td>1, 5, 25</td>
</tr>
<tr>
<td>4</td>
<td>1, 3, 13, 75</td>
<td>1, 6, 36, 216</td>
</tr>
<tr>
<td>5</td>
<td>1, 3, 13, 75, 541</td>
<td>1, 7, 49, 343, 2401</td>
</tr>
<tr>
<td>6</td>
<td>1, 3, 13, 75, 541, 4683</td>
<td>1, 8, 64, 512, 4096, 32768</td>
</tr>
</tbody>
</table>

Table 1. Minimum & Exponential Weightings
Example Subphrases and Models Tested

<table>
<thead>
<tr>
<th>n</th>
<th>MWS</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1, 3</td>
<td>1, 3</td>
</tr>
<tr>
<td>3</td>
<td>1, 3, 13</td>
<td>1, 5, 25</td>
</tr>
<tr>
<td>4</td>
<td>1, 3, 13, 75</td>
<td>1, 6, 36, 216</td>
</tr>
<tr>
<td>5</td>
<td>1, 3, 13, 75, 541</td>
<td>1, 7, 49, 343, 2401</td>
</tr>
<tr>
<td>6</td>
<td>1, 3, 13, 75, 541, 4683</td>
<td>1, 8, 64, 512, 4096, 32768</td>
</tr>
</tbody>
</table>

Table 1. Minimum & Exponential Weightings

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 $C: \Omega \rightarrow \{\text{spam, nonspam}\}$

\[
P(\text{spam}|F_i) = P(F_i|\text{spam})P(\text{spam})
\]

\[
\frac{P(F_i|\text{spam})P(\text{spam})}{(P(F_i|\text{spam})P(\text{spam}) + P(F_i|\text{ham})P(\text{ham})}
\]

- Local Formula for $P(F_i|\text{spam})$

- The output $P(\text{spam}|F_i)$ becomes $P(\text{spam})$ for the feature $F_{i+1}$

If $P(\text{spam}|F_n)$ is higher than $P(\text{ham}|F_n)$, 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

### Expressivity of Features

**Table 2. Features Generated by SBPH and OSB**

<table>
<thead>
<tr>
<th>Number</th>
<th>SBPH</th>
<th>OSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (1)</td>
<td>today?</td>
<td></td>
</tr>
<tr>
<td>3 (11)</td>
<td>lucky today?</td>
<td>lucky today?</td>
</tr>
<tr>
<td>5 (101)</td>
<td>feel &lt;\textit{skip}&gt; today?</td>
<td>feel &lt;\textit{skip}&gt; today?</td>
</tr>
<tr>
<td>7 (111)</td>
<td>feel lucky today?</td>
<td></td>
</tr>
<tr>
<td>9 (1001)</td>
<td>you &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; today?</td>
<td>you &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; today?</td>
</tr>
<tr>
<td>11 (1011)</td>
<td>you &lt;\textit{skip}&gt; lucky today?</td>
<td></td>
</tr>
<tr>
<td>13 (1101)</td>
<td>you feel &lt;\textit{skip}&gt; today?</td>
<td></td>
</tr>
<tr>
<td>15 (1111)</td>
<td>you feel lucky today?</td>
<td></td>
</tr>
<tr>
<td>17 (10001)</td>
<td>Do &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; today?</td>
<td>Do &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; today?</td>
</tr>
<tr>
<td>19 (10011)</td>
<td>Do &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; lucky today?</td>
<td></td>
</tr>
<tr>
<td>21 (10101)</td>
<td>Do &lt;\textit{skip}&gt; feel &lt;\textit{skip}&gt; today?</td>
<td></td>
</tr>
<tr>
<td>23 (10111)</td>
<td>Do &lt;\textit{skip}&gt; feel lucky today?</td>
<td></td>
</tr>
<tr>
<td>25 (11001)</td>
<td>Do you &lt;\textit{skip}&gt; &lt;\textit{skip}&gt; today?</td>
<td></td>
</tr>
<tr>
<td>27 (11011)</td>
<td>Do you &lt;\textit{skip}&gt; lucky today?</td>
<td></td>
</tr>
<tr>
<td>29 (11101)</td>
<td>Do you feel &lt;\textit{skip}&gt; today?</td>
<td></td>
</tr>
<tr>
<td>31 (11111)</td>
<td>Do you feel lucky today?</td>
<td></td>
</tr>
</tbody>
</table>
Comparison of Winnow, Naïve Bayes and CRM114 MRF Model

<table>
<thead>
<tr>
<th>Store Size</th>
<th>Naïve Bayes</th>
<th>CRM114</th>
<th>CRM114</th>
<th>Winnow + OSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1.84% (9.2)</td>
<td>1.12% (5.6)</td>
<td>1.16% (5.8)</td>
<td>0.46% (2.3)</td>
</tr>
<tr>
<td>Last 500</td>
<td>3.44% (142.8)</td>
<td>2.71% (112.5)</td>
<td>2.73% (113.2)</td>
<td>1.30% (53.9)</td>
</tr>
</tbody>
</table>

Note that Error Rate is Halved and Computational Overhead is also reduced (retaining the expressivity)
Implementing CRM114 At MailServers

Diagram:

- MTA (Sendmail, Postfix)
- MDA (Procmail, MailDrop)
  - sanitizer
  - clamd

CRM114

Folders:
- Good
- Unsure
- Spam

Receiver/Sender

Human Action:
- Good
- Unsure
- Spam

Outbound Mailserver

CRON Daemon
- Deliver Good Mail Misclassified As Spam to Good Folder via Procmail

Whitelist Senders

Outgoing Mail

User Access of Folders:
- Good
- Unsure
- Spam

by IMAP/POP
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

09/28/2005
University of California, Riverside

Shalendra Chhabra
MS Thesis Defense - Fighting Spam, Phishing and Email Fraud
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

Filters (Client, MailServers)
- Content Filters
- Collaborative Filters

{White, Black, Grey} Lists

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

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

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

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

Use HTTPS – SSL, TLS
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

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 servants affecting reputation (Scalability in Gnutella fixes this and repeaters for servants behind firewalls)
- **ID Stealth**: Malicious Servent replies with QueryReplies as if generated from genuine servants (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

Microsoft Researchers on Stopping Spam

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

writes "Scientific American today has a very interesting article about "Stopping Spam" by Joshua Goodman, David Hawkeman and Robert Routhewalt from Microsoft Research. They talk about different types of spam — spam with emails, spam on IM, spamtricks 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 Interface Proofs (HIP) in a very lucid style. They however do not mention fingerprinting approach of using "Honey Hole" to tackle addition of random words by spammers in emails or "hyperstink interstink technique used by spammers of splitting words using HTML comments, pairs of zero width tags, or bonsai tags. Also, "Spam Research" is reporting the SplitPin Technique that spammers are using to fool Yahoo! Mail SpamGuard."
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?

Thank You!