CERIAS

The Center for Education and Research in Information Assurance and Security

SNIPE: Signature Generation for Phishing Emails

Jeff Avery, Christopher Gutierrez, Paul Wood, Raffaele Della Corte, Jon Fulkerson, Gaspar Modelo-Howard,
Brian Berndt, Keith McDermott, Saurabh Bagchi, Dan Goldwasser, Marcello Cinque

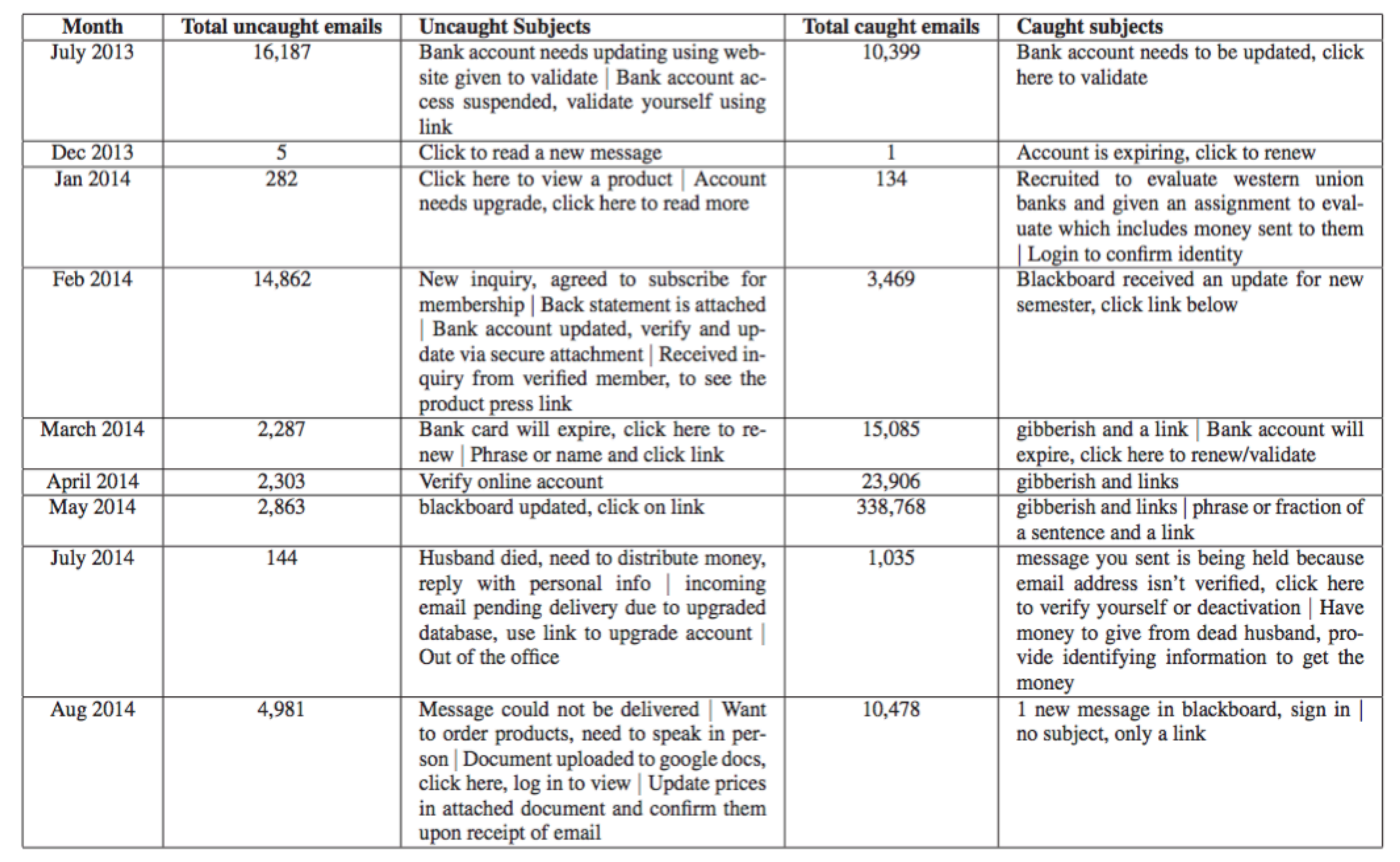
Problem Statement

- Misuse-based detection systems rely on attack signatures
- In practice, signature creation and maintenance is a manual process

Specific Goals

- Define process to ***automatically generate detection signatures***
- Create **generalized signatures**, matching for attacks and its variations

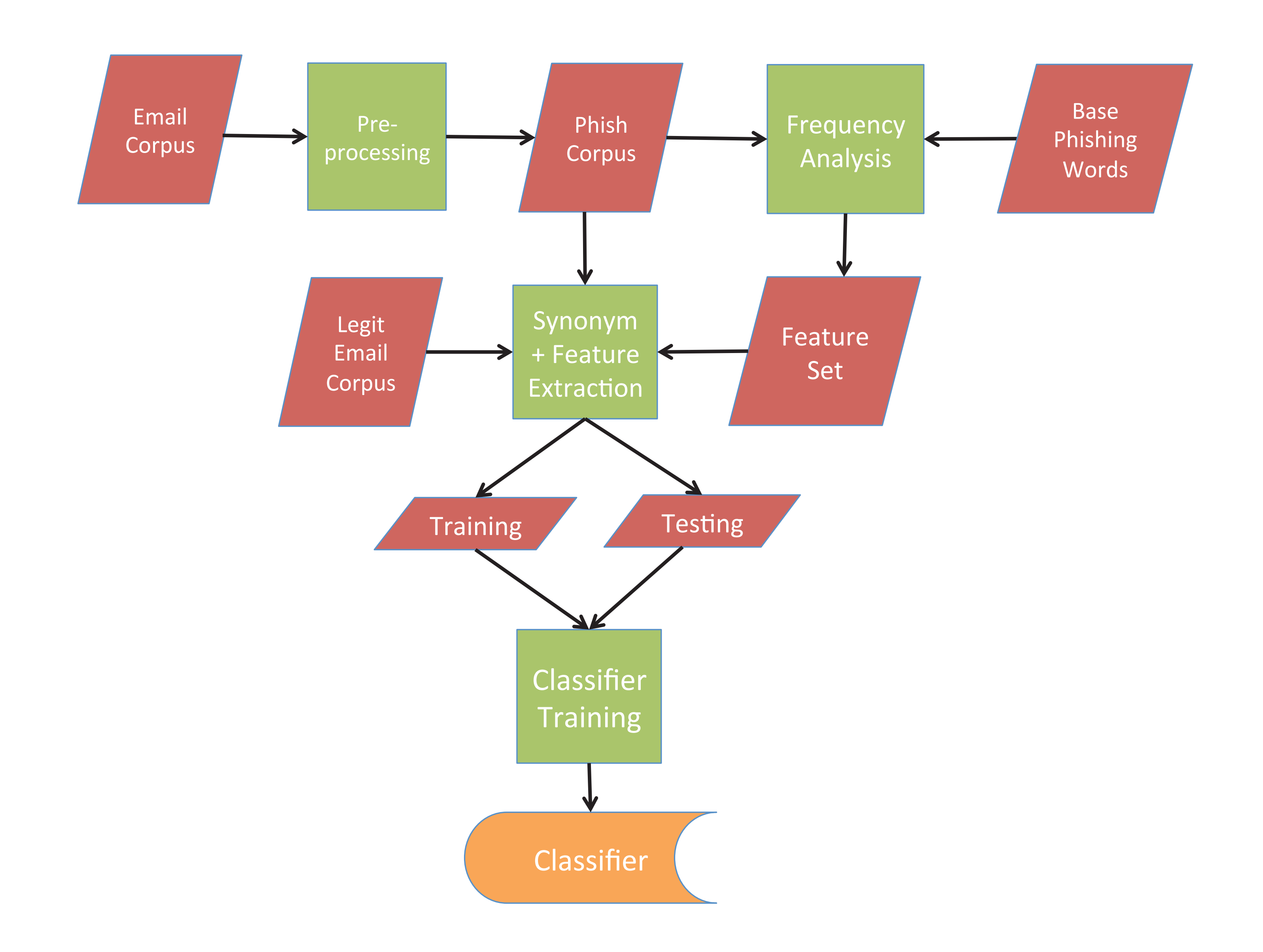
Phishing Campaign Topics



Month	Total uncaught emails	Uncaught Subjects	Total caught emails	Caught subjects
July 2013	16,187	Bank account needs updating using website given to validate | Bank account access suspended, validate yourself using link	10,399	Bank account needs to be updated, click here to validate
Dec 2013	5	Click to read a new message	1	Account is expiring, click to renew
Jan 2014	282	Click here to view a product | Account needs upgrade, click here to read more	134	Recruited to evaluate western union banks and given an assignment to evaluate which includes money sent to them | Login to confirm identity
Feb 2014	14,862	New inquiry, agreed to subscribe for membership | Back statement is attached | Bank account updated, verify and update via secure attachment | Received inquiry from verified member, to see the product press link	3,469	Blackboard received an update for new semester, click link below
March 2014	2,287	Bank card will expire, click here to renew | Phrase or name and click link	15,085	gibberish and a link | Bank account will expire, click here to renew/validate
April 2014	2,303	Verify online account	23,906	gibberish and links
May 2014	2,863	blackboard updated, click on link	338,768	gibberish and links | phrase or fraction of a sentence and a link
July 2014	144	Husband died, need to distribute money, reply with personal info | incoming email pending delivery due to upgraded database, use link to upgrade account | Out of the office	1,035	message you sent is being held because email address isn't verified, click here to verify yourself or deactivation | Have money to give from dead husband, provide identifying information to get the money
Aug 2014	4,981	Message could not be delivered | Want to order products, need to speak in person | Document uploaded to google docs, click here, log in to view | Update prices in attached document and confirm them upon receipt of email	10,478	1 new message in blackboard, sign in | no subject, only a link

SNIPE Architecture

1. PREPROCESSING: Collect emails, remove duplicates, remove spam
2. FEATURE CREATION: Develop rich set of features from phishing emails
3. CREATE VECTORS: Evaluate emails for features and create vector representation
4. MACHINE LEARNING: Use GentleBoost algorithm for imbalanced data classification



Ongoing Work – Phishing

Goal

- Automatically generate phishing signatures targeted towards universities

Data and Features

- ~350k phishing samples from ITaP
- ~60k benign samples
- ~270 features

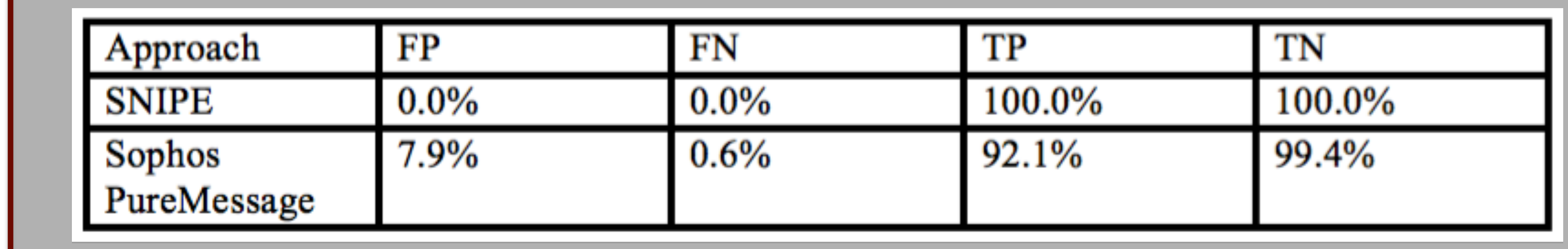
Intermediate Conclusions

- Banks/financial institutions still primary target
- Urgency seen throughout phishing emails
- Attackers follow semester schedule

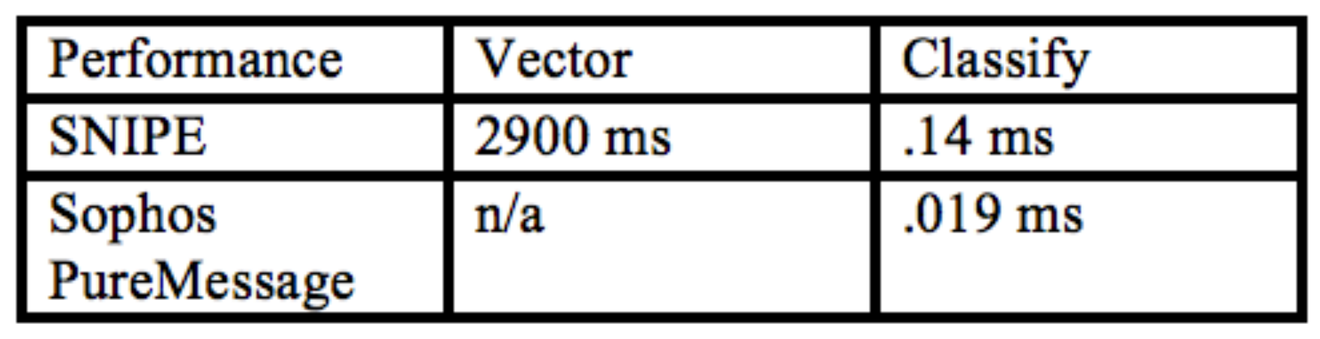
Evaluation

Training Set 22770 unique benign and 230 unique malicious emails

Test Set 22770 unique benign and 230 unique malicious emails



Approach	FP	FN	TP	TN
SNIPE	0.0%	0.0%	100.0%	100.0%
Sophos PureMessage	7.9%	0.6%	92.1%	99.4%



Performance	Vector	Classify
SNIPE	2900 ms	.14 ms
Sophos PureMessage	n/a	.019 ms

Future Work

- Analyze different classification techniques
- Develop more efficient email analysis technique
- Incorporate more NLP techniques to analyze sentence structure of phishing emails