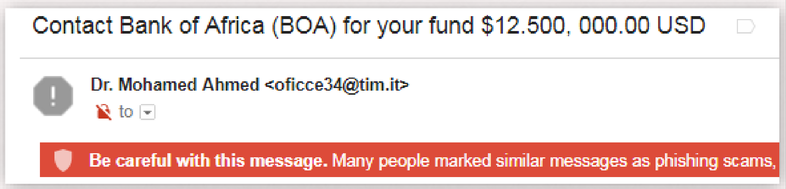
Received Field Analysis for Phishing Email Classification

Edward Filice, University of Houston
Dr. Rakesh Verma, University of Houston

Background



Phishing is a form of identity theft in which an attacker poses as a legitimate organization in order to steal personal details. Traditional methods of phishing email detection typically utilize a blacklist or examine features of the email such as its subject line or body. Utilizing features from the Received field in the header, however, has shown that this field provides significant data for phishing detection.

Objective

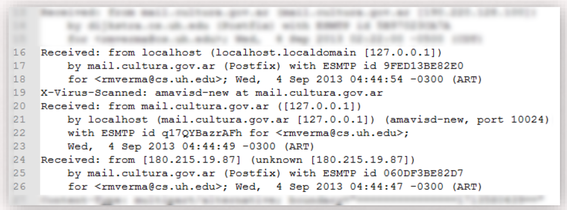
- Goal: Discover significant features to distinguish phishing and ham emails
  - Focused on Received field
- Used six email datasets
  - Ham-emfilice: 2000 from personal account
  - Ham-lftmoraes: 2000 from fellow researcher
  - Ham-SpamAssassin: 75 from SpamAssassin
  - Phishing-emfilice: 517 from personal account
  - Phishing-Nazario: 3388 from mentor
  - Phishing-RV: 196 from mentor's account

```
Received: from
va3outboundpool.messaging.microsoft.com
(10.7.14.245) by VA3EHSNDR001.bigfish.com
(10.7.40.50) with Microsoft SMTP Server (TLS)
id 14.1.225.23; Thu, 13 Sep 2012 07:32:59 +
0000
```

Example of a Received line. The "from" field is in red and the "by" field is in blue.

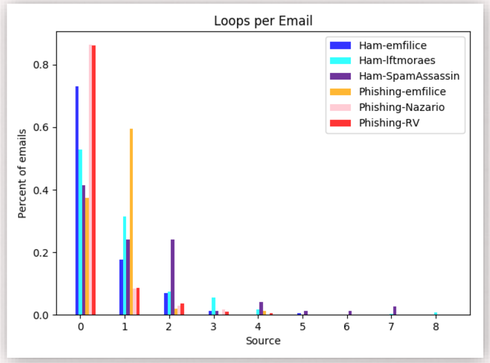
Methodology

Examine headers for testable, quantifiable features



```
Received: from localhost (localhost.localdomain [127.0.0.1])
    by mail.cultura.gov.ar (Postfix) with ESMTP id 9FED13BE82E0
    for <rmverma@cs.uh.edu>; Wed,  4 Sep 2013 04:44:54 -0300 (ART)
X-Virus-Scanned: amavisd-new at mail.cultura.gov.ar
Received: from mail.cultura.gov.ar ([127.0.0.1])
    by localhost (mail.cultura.gov.ar [127.0.0.1]) (amavisd-new, port 10024)
    with ESMTP id q17QYBazrAFh for <rmverma@cs.uh.edu>;
    Wed,  4 Sep 2013 04:44:49 -0300 (ART)
Received: from [180.215.19.87] (unknown [180.215.19.87])
    by mail.cultura.gov.ar (Postfix) with ESMTP id 060DF3BE82D7
    for <rmverma@cs.uh.edu>; Wed,  4 Sep 2013 04:44:47 -0300 (ART)
```

Parse in Python and graph for all datasets to visualize



Perform t-test to determine robustness

(>5: great, >2:good)

Table 1. Table comparing t-test results for loops.

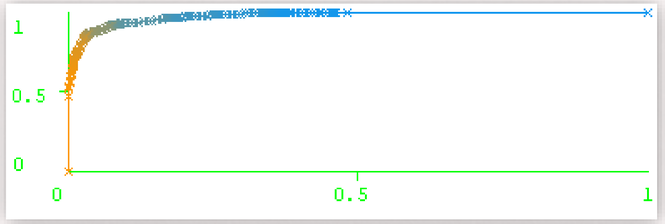
All 6 corpuses	5 corpuses
Levene score: 275.9 p-value: 5.916e-61 tvalue: 16.61 p-value: 5.916e-61	Ham-emfilice & Ham-lftmoraes to all phish: Levene score: 260.5 p-value: 1.073e-57 tvalue: 16.04 p-value: 7.885e-57
	Ham-emfilice & Ham-SpamAssassin to all phish: Levene score: 65.68 p-value: 6.357e-16 tvalue: 7.168 p-value: 9.481e-13
	Ham-lftmoraes & Ham-SpamAssassin to all phish: Levene score: 472.1 p-value: 6.258e-101 tvalue: 17.99 p-value: 2.216e-68
	Phishing-emfilice & Phishing-Nazario to all ham: Levene score: 259.9

Features Selected

- Loops: # of times an email was sent to same domain
- From gaps/IPs: # of times from field was empty/with IP
- Charlen/subdomlen: char length of from field/subdomain
- Reclines 1/2: # of rec. lines not from recipient/or sender
- UID: # of unique intermediary domains in received lines
- Dots: average # of dots in from field domains

Results

- Tested with multiple classifiers; Random Forest performed best
  - Precision: .919, Recall: .919, F-Measure: .919
  - 10-fold cross validation used for robustness
- Information gain used to rank features
  - Top feature: charlen with IG of .112
  - 7 of 9 features had IG of .05 or higher
  - Lowest feature: loops with IG of .031



ROC (false positive rate vs. true positive rate). Area under curve: .976.

Discussion

While 91.9% accuracy is not ideal it is definitely comparable to existing algorithms and features. There are still many features that could be tested with the Received field as well, and in combining our top features with others' features, we could likely attain upwards of 97%.

Acknowledgements

This research was made possible by a NSF grant to the University of Houston Computer Science Department (NSF IIS-1659755).