Using Online Activity as Digital Fingerprints to Create a Better Spear Phisher

Joaquim Espinhara & Ulisses Albuquerque
JEspinhara@trustwave.com  UAlbuquerque@trustwave.com
Agenda

- Introduction
- Motivation
- Background
- HowStuffWorks
  - Our Approach
- \(\mu\)phisher
- Demo
- Future Work
- Conclusion
About us

• Joaquim Espinhara
  – From Aracaju, Brazil
  – Security Consultant at Trustwave Spiderlabs

• Ulisses Albuquerque
  – Coder for offense & defense... as long as it’s fun!
  – Lab Manager at Trustwave Spiderlabs
INTRODUCTION
OUR MOTIVATION
Our Motivation

• Why?
• Tools available
BACKGROUND
Background

- Social Networks
- Social Engineering
- Data Mining
- Natural Language Processing - NLP
Background

- Social Networks

[Diagram showing Facebook, Twitter, LinkedIn, and Others with overlapping circles]
Background

• Social Networks
  – Communication channel for keeping in touch with someone (Facebook, Twitter)
  – Media sharing (Instagram)
  – Specialized networks (GetGlue, TripIt, LastFM)
Background

- Social Engineering
  - Phishing

http://www.d00med.net/uploads/0d832c77559a2070a766f899e7efb783.png
Background

• Data Mining
  – What is it?
  – What do you need know about it?
  – How do we use it?
Background

• Data Mining

Raw data set

"Had lunch with @urma and @jespinhara today #tgif #lunch"

Data cleaning

"Had lunch with @urma and @jespinhara today"

Data integration

"Had lunch with @urma and @jespinhara today"

Data normalization

"Had lunch with @urma and @jespinhara today (2013-06-05)"

black hat
USA 2013
Background

• Natural Language Processing – NLP
  – What is it?
  – What do you need know about it?
  – How do we use it?
  – Text analysis
Background

- Natural Language Processing - NLP

http://webu2.upmf-grenoble.fr/sciedu/nlpsl/nlpsl.jpg
Our Approach

Identifying the subject to profile
Collecting social network data
Analyzing and building the profile
• The Unknown Subject (*Unsub*)

- Joaquim Espinhara
- @jespinhara (Twitter)
- joaquim.espinhara (Facebook)
- uid=12345 (LinkedIn)
Our Approach

- Data Collection
  - Social Network IDs
  - Official APIs
  - Web Scraping
  - OAuth
Our Approach

• Data Collection - Twitter

- Application ID (μphisher)
- User ID (@jespinhara)
- Twitter @urma @effffn @SpiderLabs
μphisher

• Reference implementation
• Goals
  – Validate potential unsub content
  – Assisted textual content input
μphisher

- Web Application
- Twitter only (for now)
- Open Source (GPLv3)
μphisher

- MongoDB, Mongoid
- DelayedJob
- Ruby on Rails
- OAuth
µphisher

automated social engineering profiling

Word Frequency:  
Sentence Length:
Sentence Count:  
Phrase Count:
User References:  
Pronoun Usage:

Overall Score:

Code hackaton with @jespinhara for the latest µphisher rele

release  rele
DEMO
(FINGERS CROSSED)
Future Work

- Support for additional data sources
- Machine learning
- More metrics and feedback for assisted input
CONCLUSION
THANK YOU!