Visual spoofing in content-based spam detection

Mark Sokolov, Kehinde Olufowobi, and Nic Herndon East Carolina University

13th International Conference on Security of Information and Networks November 4 – 7, 2020 Istanbul, Turkey

Overview

Introduction

- Unchanged vs. changed email
- 2 Experimental design
 - Data source and format
 - Workflow

3 Results

- Accuracy
- Microsoft mail
- Google mail
- Other domains



BECU

Introduction

Subject: Please send money

Body: I am so distraught. I thought i could reach out to you to help me out. I came down to United Kingdom for a short vacation unfortunately i was mugged at the part of the hotel i stayed, all cash, credit card and cell phone was stolen from me but luckily for me i still have my passport with me. I've been to the embassy and to the police here but they're not helping issues at all end, my flight leaves in few hours time from now but. I am having problems settling the hotel bills and the hotel manager won't let me leave until i settle my hotel bills. I am freaked out at the moment.



Unchanged vs. changed email



BECU

Sokolov, Olufowobi, and Herndon

Data source

• Emails from the Enron1 dataset



BECU

Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

æ

イロト イポト イヨト イヨト

Workflow



BECU

Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

3

୬ < ୍ (4)



Analysis details

- Machine learning algorithms used
 - Naïve Bayes (NB)
 - Support vector machine (SVM)
 - Decision tree (DT)
 - Random forest (RF)
- Evaluation metrics
 - Accuracy
 - Precision
 - Recall
 - F-1 score
 - Confusion matrix
 - Production testing (Microsoft)

æ

・ロト ・ 日 ト ・ 日 ト ・ 日 ト

Accuracy

Results and discussion



Control experiment involving entirely unmodified datasets – the default encoding of characters in both the training and testing sets is preserved



Encoding of the training set is preserved and the encoding of certain characters in the testing set uses corresponding confusables from the Cyrillic alphabet



Confusables introduced in training and testing sets so that each model is trained and evaluated with data from a single, mixed-script that contains confusables

・ロト ・ 日 ト ・ 日 ト ・ 日 ト

3

Sokolov, Olufowobi, and Herndon

₿ECU

Junk mail vs. inbox - Microsoft mail

This email originated from outside ECU.

I am so distraught. I thought i could reach out to you to heip me out. I came down to United Kingdom for a short vacation unfortunately i vas mugged at the park of the hotel i staved, all cash, credit card and cell phone was stolen from me but luckly for me i still have my passport with me. I've bene to the embassy and to the police here but therive not helping issues at all and, my flight leaves in few hours time from now but. I am having problems settling the hotel bills and the hotel manager won't let me leave until settle my hotel bills. I'm freaked out at the moment.





Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

Junk mail vs. inbox - Gmail

	Money for Viagra D Inbex X		ē	Z			
*	Sokolov, Mark to me ~	Fri, Oct 30, 1:30 PM (1 day ago) 🕺	4	:			
	I am so distraught. I thought I could reach out to you to help me out. I came down to Uni unfortunately I was mugged at the part of the hotel I stayed, all cash, credit card and ce me i still have my passport with me. I've been to the embassy and to the police here but th flight leaves in few hours time from now but. I am having problems settling the hotel bills until I settle my hotel bills. I am freaked out at the moment Who are you? Thanks, flicheck them out. Very nicet	I phone was stolen from me but lu ey're not helping issues at all end,	my				
	Money for Viagra D Beam K		ē	Z			
0	Sokolov, Mark to me ~	Oct 30, 2020, 1:31 PM (1 day ago) 🛛 😭	*	:			
	Why is this message in spam? It is similar to messages that were identified as spam in the par Report not spam	it.	Ċ				
I am so distraught. I thought i could reach out to you to help me out. I came down to United Kingdom for a short vacation unfortunately i was mugged at the part of the hotel i stayed, all cash, credit card and cell phone was stolen from me but luckily for me i still have my passport with me. I ve been to the embassy and to the police here but they're not helping issues at all end, my flight leaves in few hours time from now but. I am having problems settling the hotel bills and the hotel manager won't let me leave until i settle my hotel bills. I am freaked out at the moment							

Sokolov, Olufowobi, and Herndon

BECU

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

E ∽ Q (9)

Google search

11/30/2019	I am so distratight. I thought i could reach out to you to help me out. I came down to United Kingd	om for a short vacation unfortunately i was mu	11/30/2019	I am so doitriaight. I thought i could reach out to you to help me out. I came down to United Kingdom for a short vacation infortunately i was mu				
Google	I am so distraught. I thought i could reach out to you to help me out. I came down Q	III 🔞	Google	I am so distraught. I thought i could reach out to you to help me out. I came dowr Q III 🔞				
	Q. All [] Images 면 News 단 Videos 양 Maps I More Settings Tools About 205,000 results (1.48 seconds)			C, All (1) Volues (2) Maps (1) Images (2) News I More Settings Tools Advant 5 results (0.89 seconds)				
	"for" (and any subsequent words) was ignored because we limit queries to 32 words.			"all" (and any subsequent words) was ignored because we limit queries to 32 words.				
	For the Love of Vlagra Spann and the 419 Email Scam HuffPost http://www.huffpost.com.setu; for the love of vlagra an, b, 76650 * 00 11, 2010 - Ina distangli. Ithough call and what with you hold have not. (can eleven to bink Bidgeten for a short vascelinic undiristantisty ices magged at the park of the heat) steps, all calas, decide and or displayees was ables from we had hubbly for me i still have my passport with me. Twe laver to the			DBI your mean: I am so distanced by T. Bengahi roadi roads or to you to holp me cut. Learner down to builde for fighting the salet roadson understroad and wave mogged at the gave of the hold i stayset, all cash, control card or coll and the same of the same road of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same build wave of the same of the same of the same of the same of the same build wave of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the sa				
	People also ask			Trump lashes out at Iran for shutting down internet - KEYT				
	What to do if you get robbed in a foreign country? $\qquad \lor$			https://keyt.com > news > national-world > 2019/11/21 > trump-lashes-out ▼ Published Neversher 21, 2019 10.21 am. Trump-lashes out at iran for shutting dawn internet.				
	What should I do if I get robbed?			WACHINGTON (AP) President Donald Trurp says Iran is so "unstable" that the Iranian government has shut down the internet so Iranians cannot				
	Does travel insurance cover being robbed?			Trump lashes out at Iran for shutting down internet - KTVZ				
	My Morm Got Robbed In Our Hotell (One Mile at a Time Mpc), Unreministrating core, may more approached to our hotel * to entry of home ymour approached in an hotel (at the Marines, to assist only (one server and prophore) and the out of prophore and the Marines and the Marines and the approached and the Marines and the Marines and the Marines and the Marines and the approached and the Marines and the Marines and the Marines and the Marines and the approached and the Marines and the Marines and the Marines and the Marines and the approached and the Marines and the Marines and the Marines and the Marines and the American and the Marines and American and the Marines and the Marines and the Marines and the Marines and the Marines and American and the Marines and the Marines and the Marines and the Marines and the Marines and American and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines and the Marines an			Http://black.com/news-national-and/si.2019/11/21 http://plackase.com/news/ http://blackas/blackase/index/site/site/site/site/site/site/site/site				
Inpulvers gospk.com/search/scofeACYBOR83TBVID_domIN122698885680283A1575147692364aci-Q43X7_BELdqL4r4Qdap-Lamosondinc 14 http://www.gospk.com/search/scofeACYBOR83TBVID_domIN122698885680283A1575147692364aci-Q43X7_BELdqL4r4Qdap-Lamosondinc 14								



Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

2

ク^{へへ} (10)

・ロン ・四 と ・ ヨ と ・ ヨ と

Google translate

 11/202019
 Geogle Translate

 ENGLISH - DETECTED
 +⁻⁺
 ENGLISH

I am so distraught. I thought i could reach out to you to help me out. I came down to United Kingdom for a short vacation unfortunately i was mugged at the park of the hotel istayed, all cash, credit card end cell phone was stolen from me but luckily for me i still have my passport with me. I've been to the embassy and to the police here but they're not helping issues at all and, my flight leaves in few hours time from now but. I am having problems settling the hotel bills and the hotel manager won't let me leave until i settle my hotel bills. I am freaked out at the moment.

I am so distraught. I thought i could reach out to you to help me out. I came down to United Kingdom for a short vacation unfortunately i was mugged at the park of the hotel istayed, all cash, credit card cnd cell phone was stolen from me but luckily for me i still have my passport with me. I've been to the embassy and to the police here but they're not helping issues at all and, my flight leaves in few hours time from now but. I am having problems settling the hotel bills and the hotel manager won't let me leave until i settle my hotel bills. I am freaked out at the moment.

https://translate.google.com/#view=home&op=translate&sl=auto&tl=en&text=I am so distraught. I thought i could reach out to you to help me out. I came down... 1/1

ECU

Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey

◆□▶ ◆圖▶ ◆厘▶ ◆厘▶ ○

Summary

Introduction

- Unchanged vs. changed email
- 2 Experimental design
 - Data source and format
 - Workflow

3 Results

- Accuracy
- Microsoft mail
- Google mail
- Other domains



BECU

Thank you

Questions

Mark Sokolov sokolovm19@students.ecu.edu

https://github.com/sokolovm19/Visual_Spoofing



Sokolov, Olufowobi, and Herndon

Visual spoofing in content-based spam detection

SINCONF 2020, Istanbul, Turkey