black hat EUROPE 2018

DECEMBER 3-6, 2018

EXCEL LONDON / UNITED KINGDOM

DeepPhish Simulating Malicious AI

Alejandro Correa Bahnsen, PhD

VP, AI & Research – Cyxtera Technologies

About Me

- Industrial Engineer
- PhD in Machine Learning
- Passionate about open-source
- Scikit-Learn contributor
- Organizer of Data Science Meetups

Who I've worked with

Dow

uni.lu
UNIVERSITÉ DU LUXEMBOURG

Scotiabank

Cyxtera

SIX GROUP

toptal

Agenda

- Phishing URL Detection using Machine Learning
- Malicious Cert Detection using Deep Learning
- DeepPhish: Simulating Malicious AI
- Demo 😩

Typical Phishing Example


File Message ADOBE PDF Tell me what you want to do

Thu 6/8/2017 11:57 PM

SO Online Access <notify@payments.interac.ca>

ess Has Been Blocked

To

Dear Customer

Your Scotia Bank Online Access And The Available Balance In It Has Been Locked

Your Account Has Been Frozen For The Reasons Below ,

Your Security Questions Does Not Match Your Answers

You Will Not Be Able To Transfer , Withdraw Or Recieve Funds Until You Have Securely Click The Link Below To Confirm Your Security Questions And Answers As Registered With Scotiabank .

To Unlock Your Account , Click The Link Below , Login With Your Card Number , You Will Be Redirected To Your Account Area To Select Your Security Questions And Answers As You have Regisred With Scotiabank

Your Account Will Not Be Restored If You Do Not Provide The Correct Security Questions And Answer We Have On Our System

Your Account Will Be Locked Forever If You Do Not Provide Accurate Informations

CLICK HERE TO UNBLOCK YOUR ACCOUNT ACCESS

This is an automated message. Please do not reply to this Email, as your response will not be received.

Yours Sincerely,

THE BANK OF

Because your reply will not be transmitted via secure e-mail, the e-mail address that

Your Account Will Be Locked Forever If You Do Not Provide Accurate Informations

CLICK HERE TO UNBLOCK YOUR ACCOUNT ACCESS

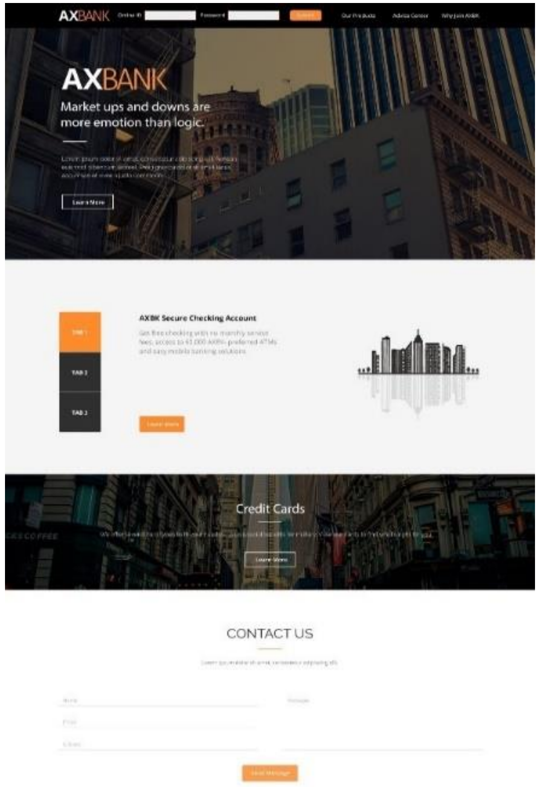
This is an automated message. Please do not reply to this Email, as your response will not be received.

91%

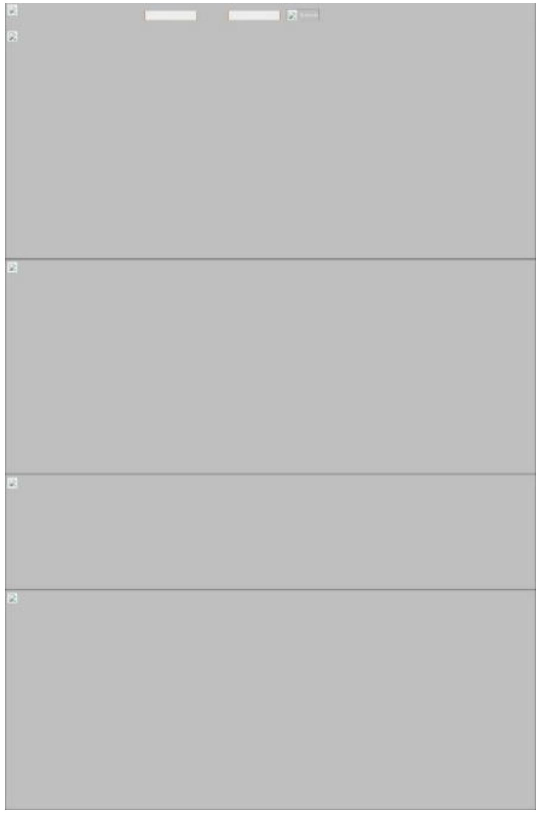
of cybercrimes
and attacks
**start with a
phishing email**

Why Phishing Detection is Hard?

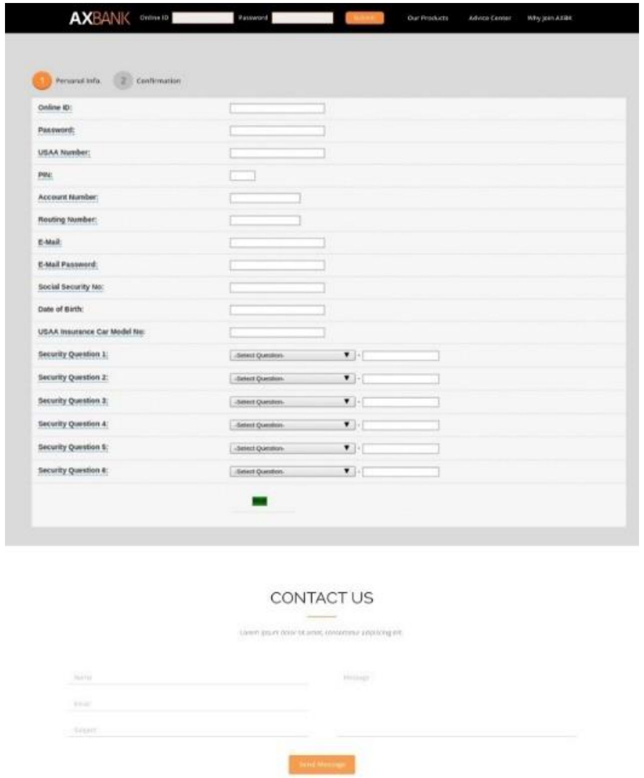
Original Website



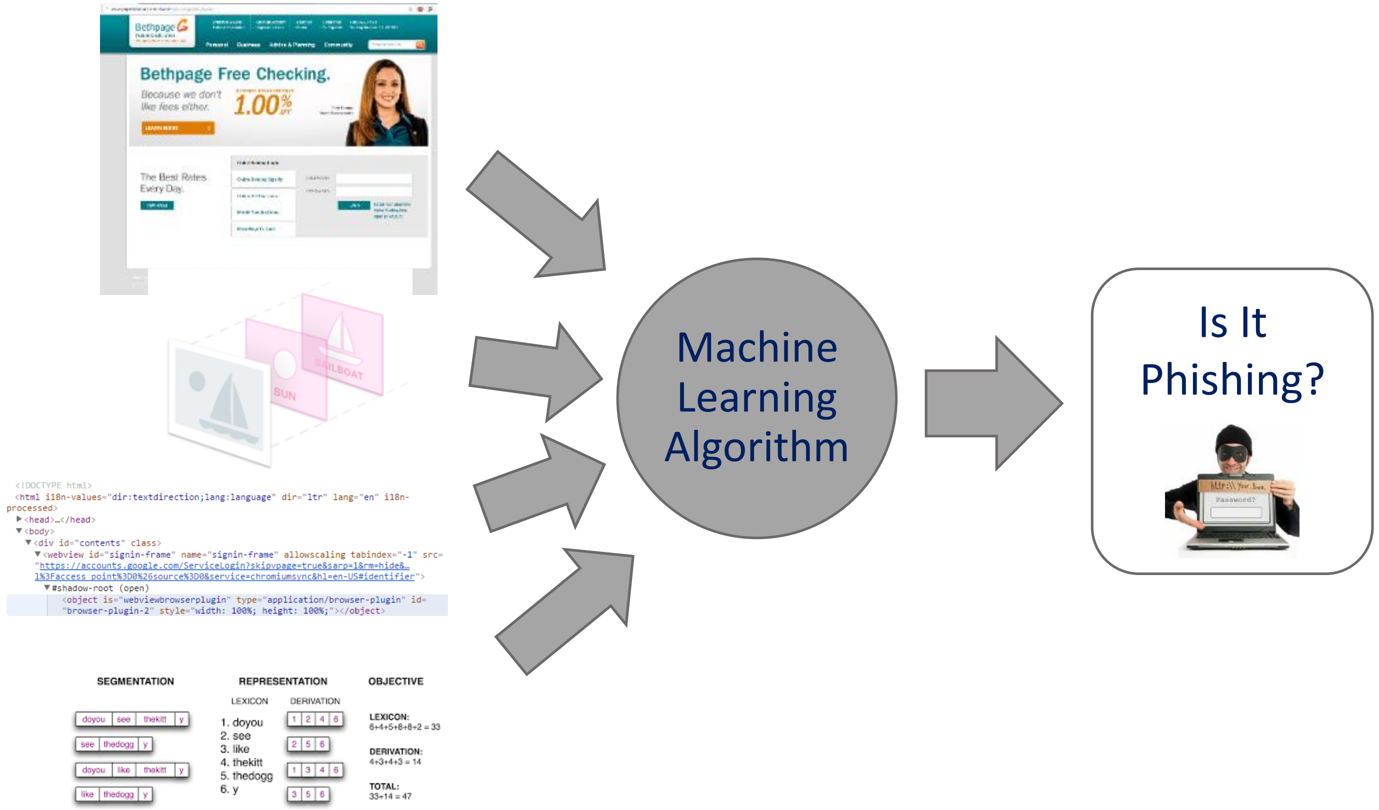
Only Using Images



Subtle Changes



Ideal Phishing Detection System

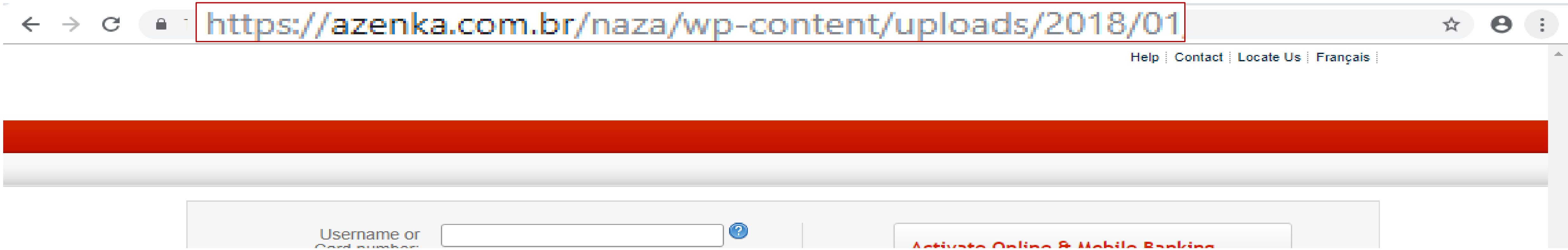


Ideal Phishing Detection System

Issues with full content analysis:

- Time consuming
- Impractical to process millions of websites per day
- Hard to implement for small devices

There is always the need for an URL



Database of URLs

1,000,000 Phishing URLs from Phish Tank

http://moviesjingle.com/auto/163.com/index.php

http://**paypal.com**.update.account.**toughbook.cl**/8a30e847925afc5975161aeabe8930f1/?cmd=_home\&dispatch=d09b78f5812945a73610edf38

http://msystemtech.ru/components/com_users/Italy/zz/Login.php?run=_login-submit\&session=68bbd43c854147324d77872062349924

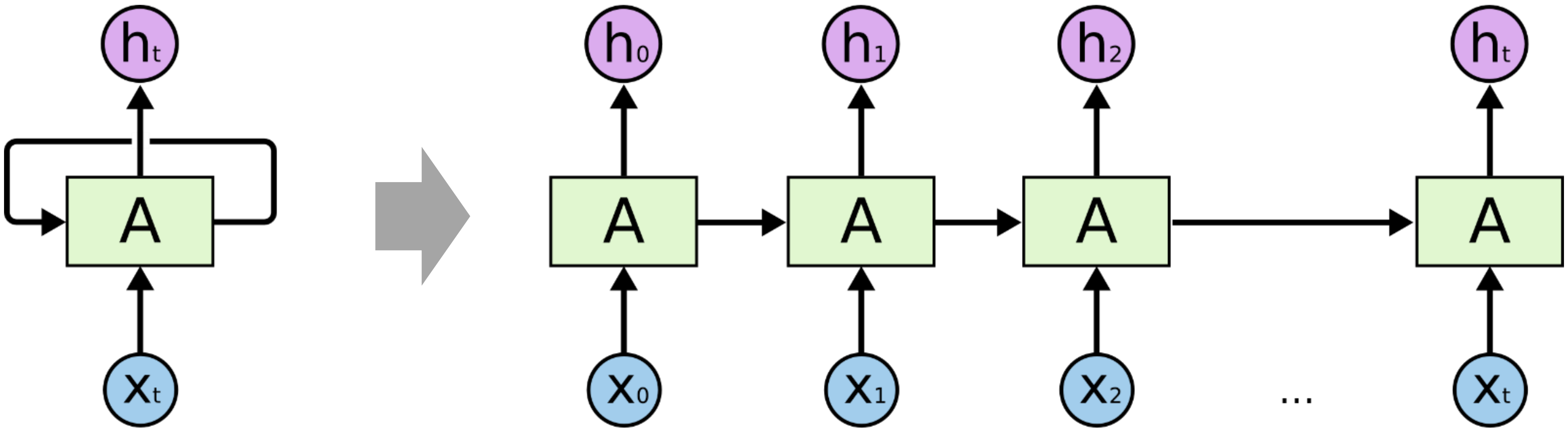
1,000,000 Legitimate URLs from Common Crawl

https://www.sanfordhealth.org/ChildrensHealth/Article/73980

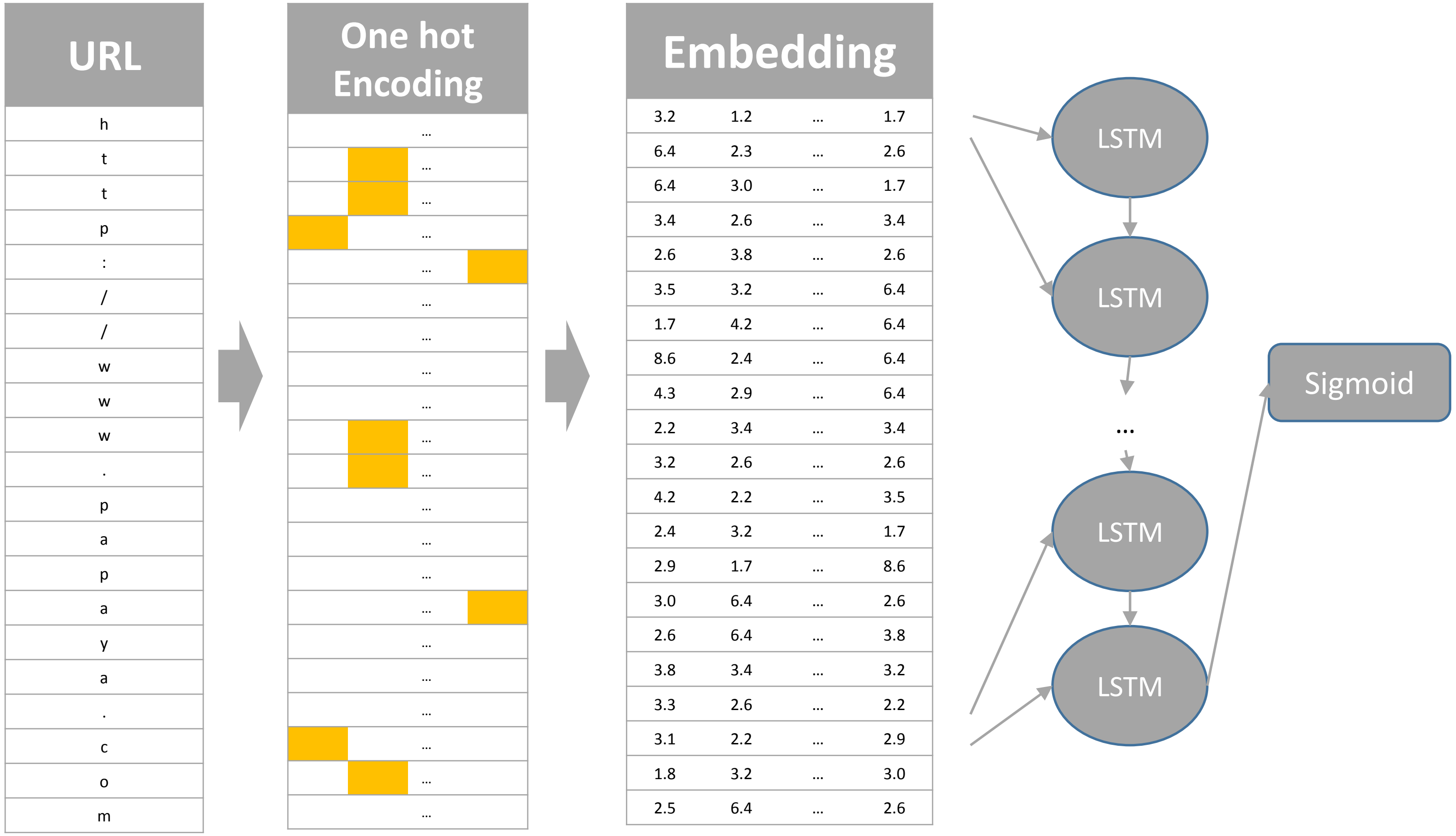
http://www.grahamleader.com/ci_25029538/these-are-5-worst-super-bowl-halftime-shows\&defid=1634182

http://www.carolinaguesthouse.co.uk/onlinebooking/?industrytype=1\&startdate=2013-09-05\&nights=2\&location\&productid=25d47a24-6b74

Recurrent Neural Networks RNN



Recurrent Neural Networks RNN

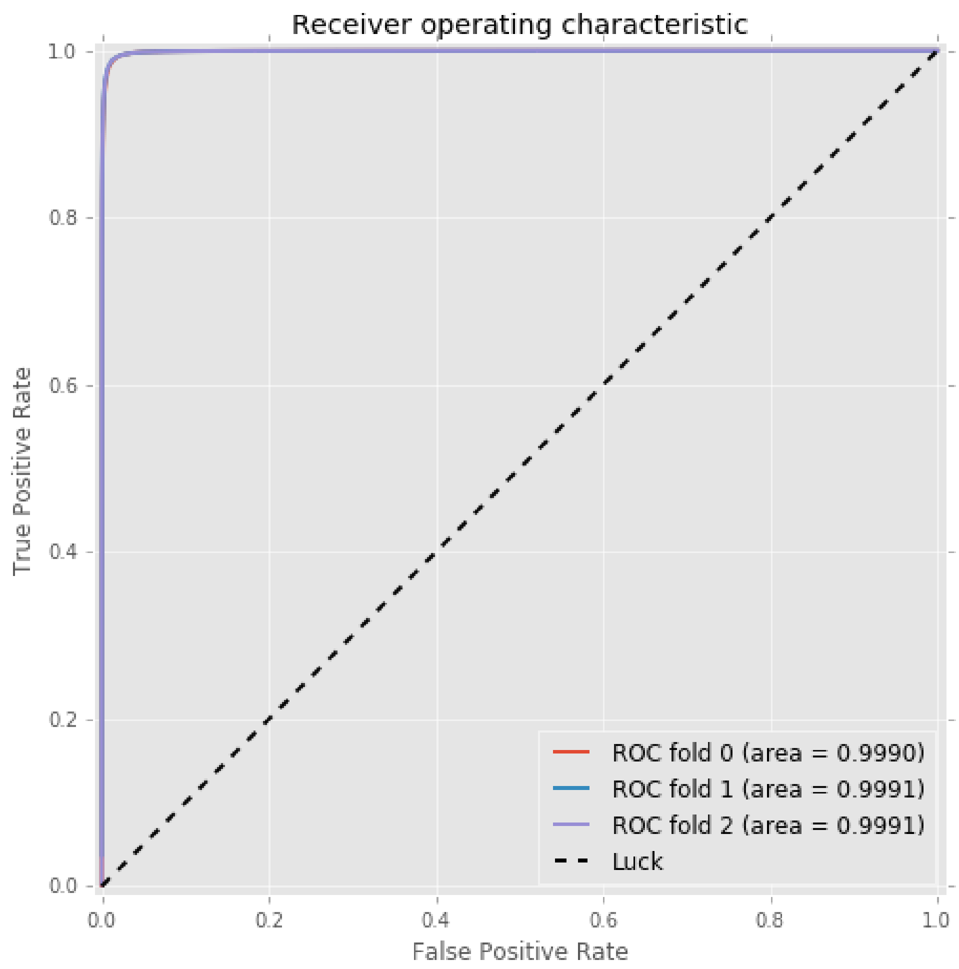


URL
h
t
t
p
:
/
/
w
w
w
.
p
a
p
a
y
a
.
c
o
m

Embedding			
3.2	1.2	...	1.7
6.4	2.3	...	2.6
6.4	3.0	...	1.7
3.4	2.6	...	3.4
2.6	3.8	...	2.6
3.5	3.2	...	6.4
1.7	4.2	...	6.4
8.6	2.4	...	6.4
4.3	2.9	...	6.4
2.2	3.4	...	3.4
3.2	2.6	...	2.6
4.2	2.2	...	3.5
2.4	3.2	...	1.7
2.9	1.7	...	8.6
3.0	6.4	...	2.6
2.6	6.4	...	3.8
3.8	3.4	...	3.2
3.3	2.6	...	2.2
3.1	2.2	...	2.9
1.8	3.2	...	3.0
2.5	6.4	...	2.6

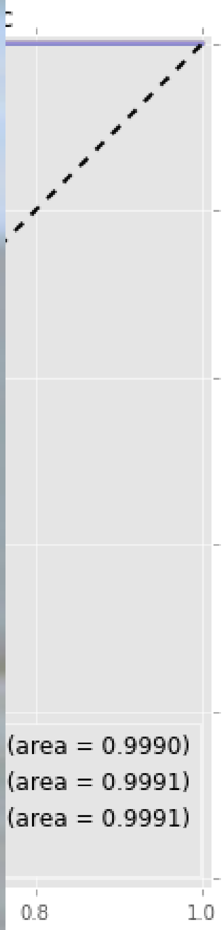
URL Classification Results

3-Fold CV	Accuracy	Recall	Precision
Average	98.76%	98.93%	98.60%
Deviation	0.04%	0.02%	0.02%



URL Clas

3-Fold CV
Average
Deviation



Detecting Malicious URLs Is Not Enough!!

WIRED | Phishing Schemes Are Using Encrypted Sites to Seem Legit | SIGN IN | SUBSCRIBE

LILY HAY NEWMAN SECURITY 12.05.17 02:32 PM

PHISHING SCHEMES ARE USING ENCRYPTED SITES TO SEEM LEGIT

SHARE

SHARE 750

TWEET

COMMENT

EMAIL

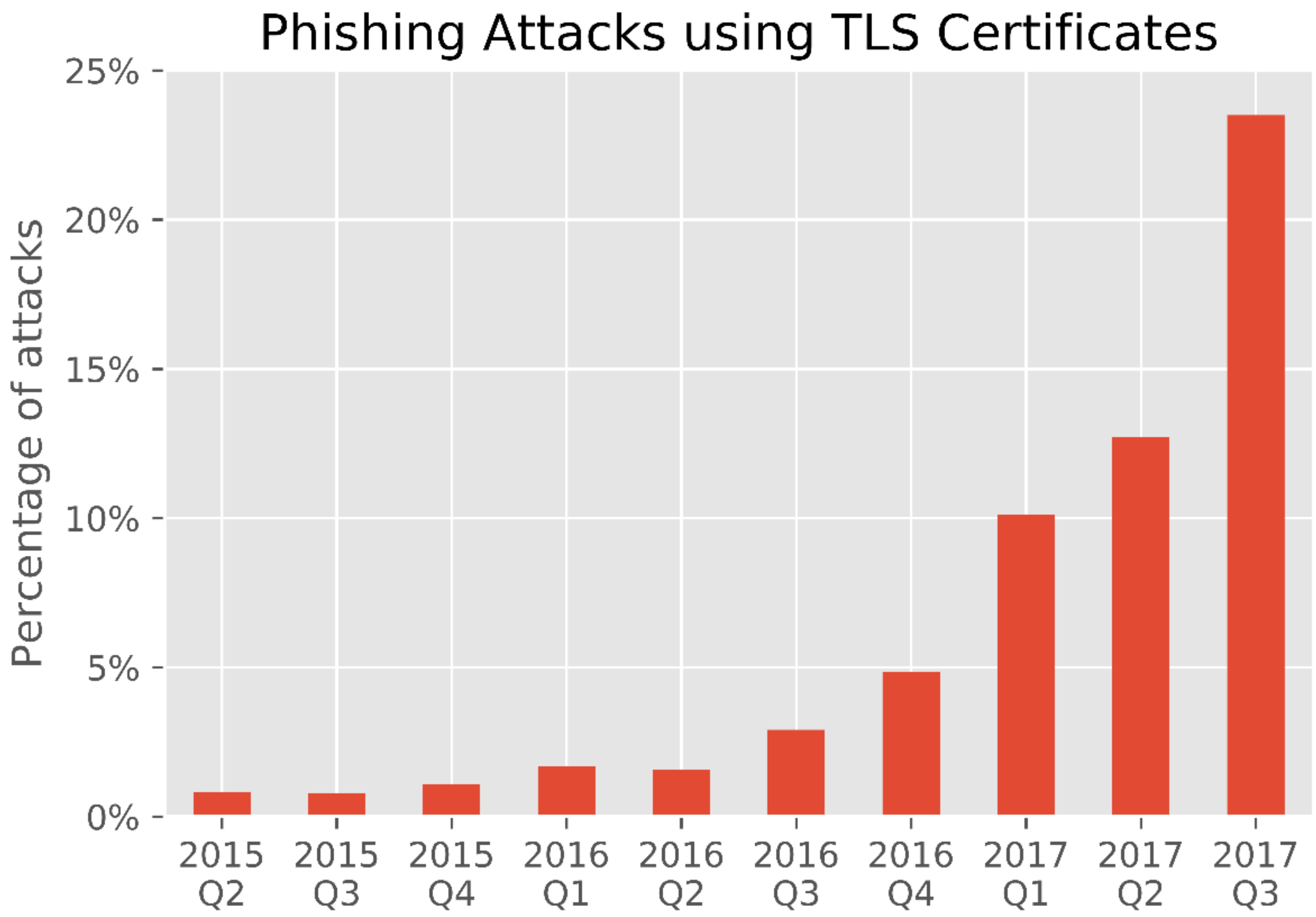
GETTY IMAGES

MOST POPULAR

SCIENCE
The Peculiar Math That Could Underlie the Laws of Nature
NATALIE WOLCHOVER

CULTURE
A Deadly Hunt for Hidden Treasure Spawns an Online Mystery
DAVID KUSHNER

CULTURE
MoviePass Raises Prices, Limits First-Run Availability as Pressures...
BRIAN BARRETT



Images from: https://info.phishlabs.com/blog/quarter-phishing-attacks-hosted-https-domains

What is a Web Certificate?

http://ultrabank.com

Secure | https://ultrabank.com

Forrester survey asked users: “Some websites receive the following browser user interface security indicator in the browser. What do you think the security indicator is intended to tell users?”

Secure | https://ultrabank.com

The website is safe: 82%
The website is encrypted: 75%
The website is trustworthy: 66%
The website is private: 32%

Forrester survey asked u… …ollowing browser user interface security ir… …ou think the security indicator is inter…

Se…

The w…
The w…
The w…
The w…

Hunting Malicious TLS Certificates with Deep Neural Networks

Database of TLS Certificates

1,000,000 Legitimate Certificates from Common Crawl

CN = *.stackexchange.com, O = Stack Exchange, Inc., L = New York, S = NY, C = US

CN = slack.com, O = Slack Technologies, Inc., L = San Francisco, S = CA, C = US

CN = *.trello.com, O = Trello Inc., L = New York, S = New York, C = US

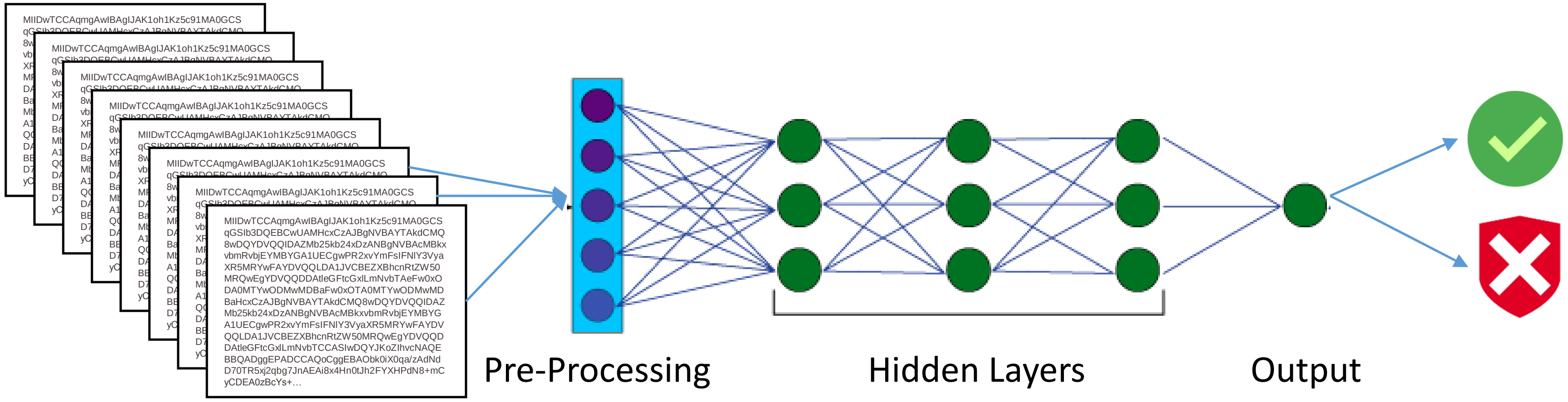
5,000 Phishing Certificates

CN = localhost, L = Springfield

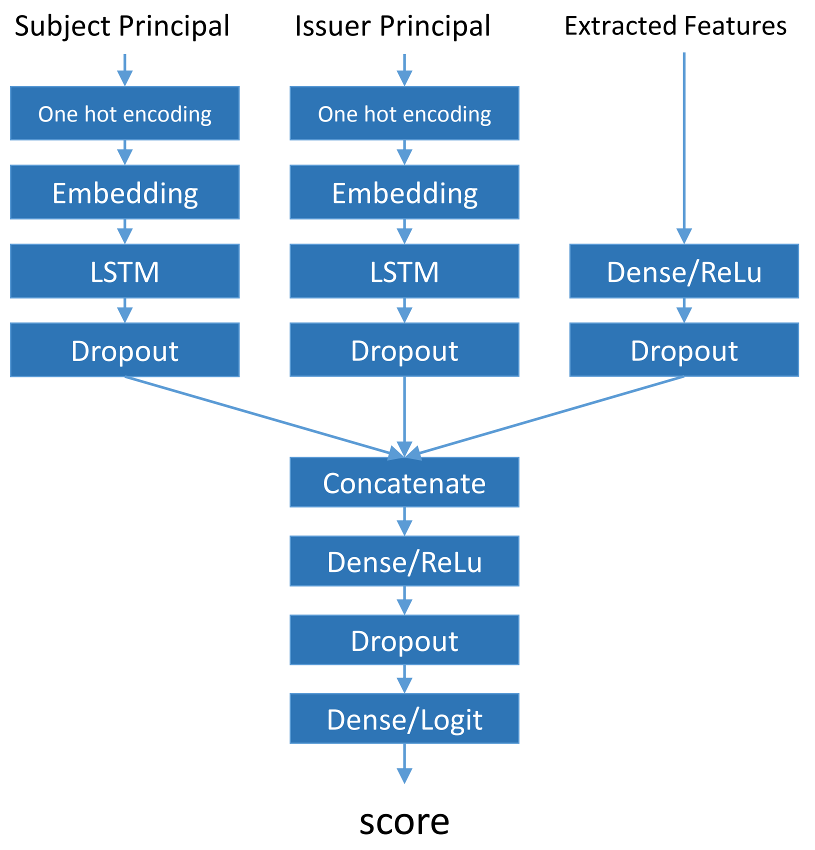
CN = localhost.localdomain

CN = example.com, L = Springfield

Deep Learning Algorithm

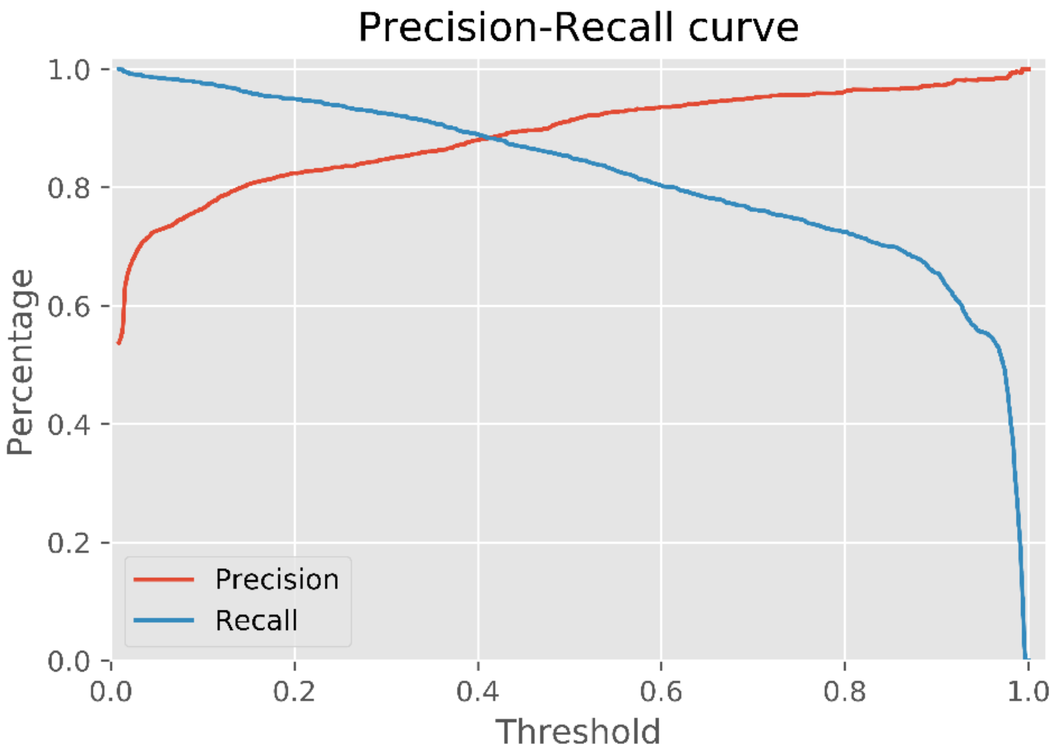


Deep Learning Algorithm

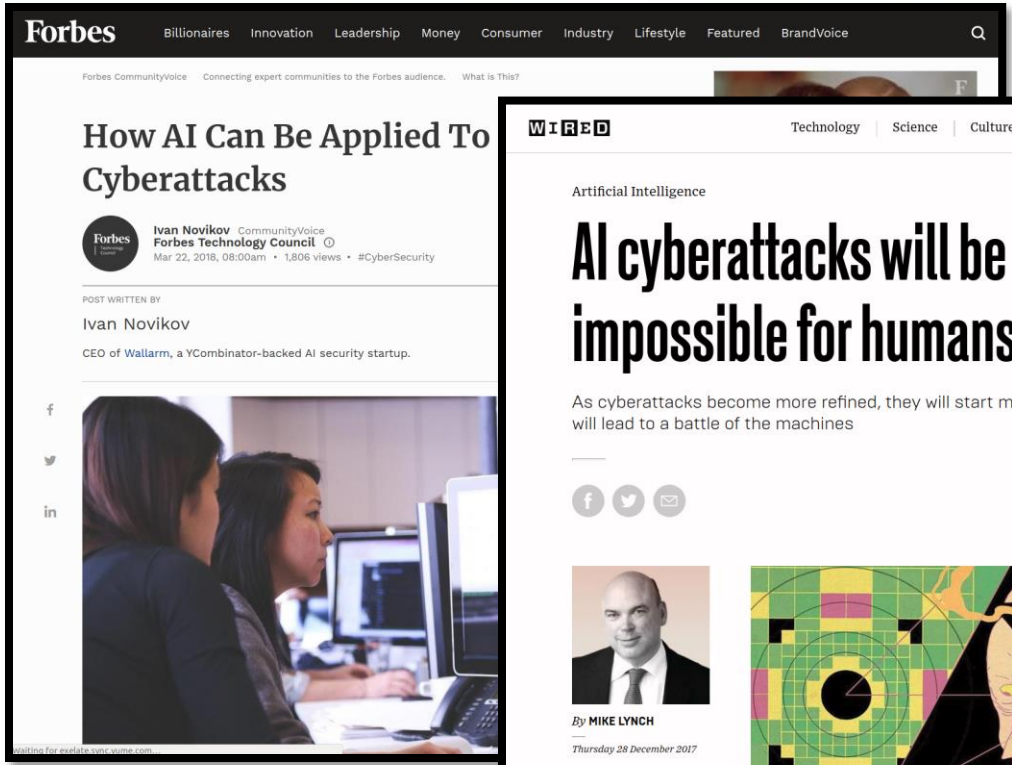


Malicious Cert Classification Results

5-Fold CV	Accuracy	Recall	Precision
Average	86.41%	83.20%	88.86%
Deviation	1.22%	3.29%	1.04%




EVERY
ACTION
HAS A
FRAUDSTER
REACTION



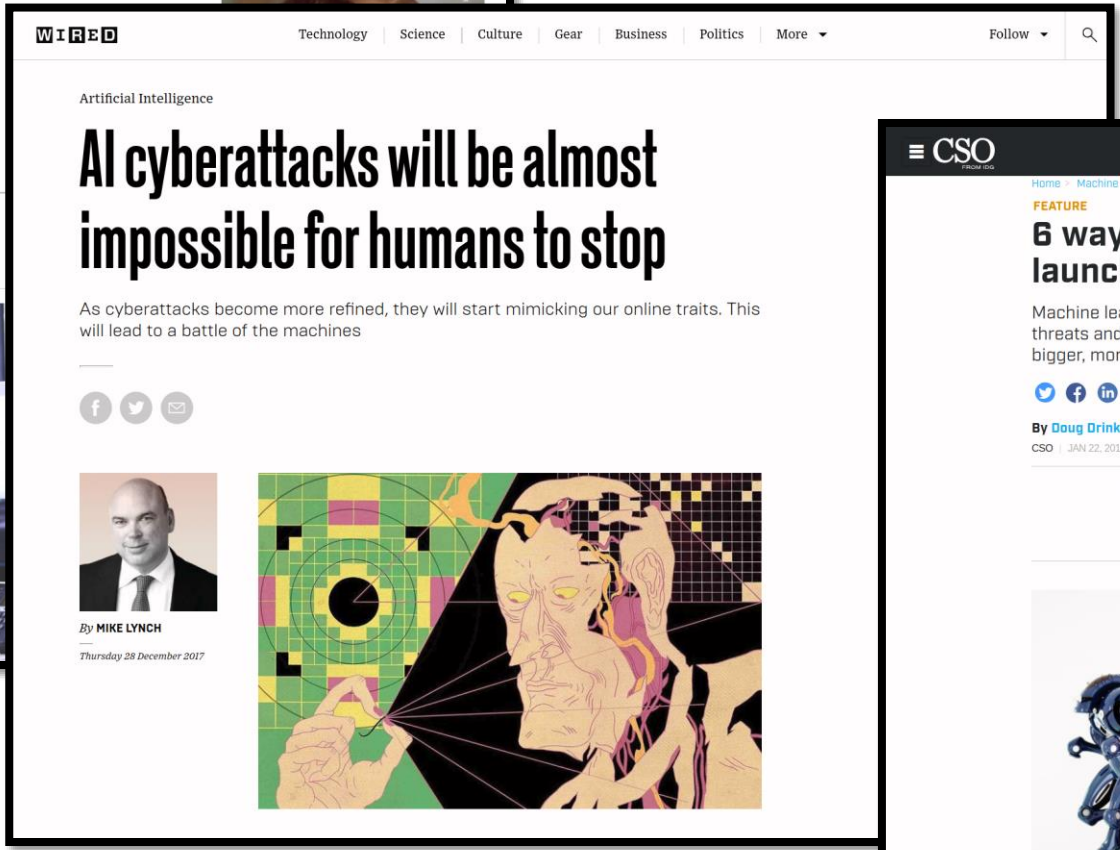
Forbes Billionaires Innovation Leadership Money Consumer Industry Lifestyle Featured BrandVoice

Forbes CommunityVoice Connecting expert communities to the Forbes audience. What is This?

How AI Can Be Applied To Cyberattacks

Ivan Novikov CommunityVoice
Forbes Technology Council
Mar 22, 2018, 08:00am • 1,806 views • #CyberSecurity

POST WRITTEN BY
Ivan Novikov
CEO of Wallarm, a YCombinator-backed AI security startup.



WIRED Technology Science Culture Gear Business Politics More Follow

Artificial Intelligence

AI cyberattacks will be almost impossible for humans to stop

As cyberattacks become more refined, they will start mimicking our online traits. This will lead to a battle of the machines

By MIKE LYNCH
Thursday 28 December 2017

CSO INSIDER Sign In | Register

Home > Machine Learning

FEATURE

6 ways hackers will use machine learning to launch attacks

Machine learning algorithms will improve security solutions, helping human analysts triage threats and close vulnerabilities quicker. But they are also going to help threat actors launch bigger, more complex attacks.

By Doug Drinkwater
CSO | JAN 22, 2018 3:00 AM PT

Si le parece difícil establecer una red, imagine hacerlo en Marte. Comience el juego › CISCO

MORE LIKE THIS
5 top machine learning use cases for security
How AI can help you stay ahead of cybersecurity threats
The future of AI and endpoint security
VIDEO Who can benefit from Zero Trust Security?

DeepPhish Simulating Malicious AI

The Experiment: Simulating Malicious AI

1. Identify individual threat actors
2. Run them through our own AI detection system
3. Improve their attacks using AI

Uncovering Threat Actors

- Objective: We want to understand effective patterns of each attacker to improve them through a AI model

- As we can not know attackers directly, we must learn from them through their attacks

- Database with 1.1M confirm phishing URLs collected from Phishtank

Threat Actor 1

naylorantiques.com

406 URLs

http://naylorantiques.com/components/com_**contact**/views/**contact**/tmpl/62

http://naylorantiques.com/docs/Auto/**Atendimento**/5BBROPI6S3

http://naylorantiques.com/**Atualizacao Segura**/pictures/XG61YYMT_FXW0PWR8_5P2O7T2U_P9HNDPQR/

http://naylorantiques.com/zifn3p72bsifn9hx9ldecd8jzl2f0xlwf8f

http://www.naylorantiques.com/JavaScript/charset=iso-8859-1/http-equiv/margin-bottom

Visual Check

Keywords

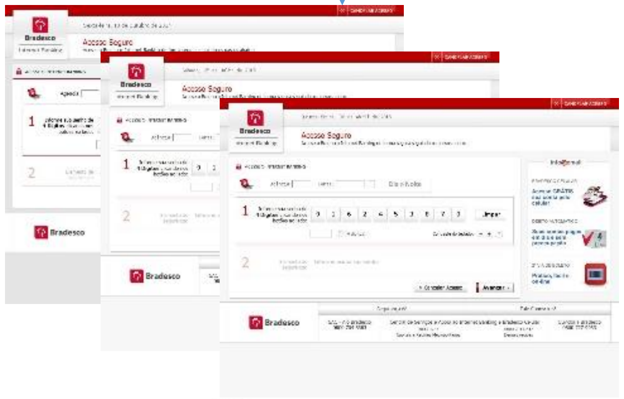
atendimento, jsf, identificacao, ponents, views, TV, mail, SHOW, COMPLETO, VILLA, MIX, ufi, pnref, story, tryy2ilr, Autentico

Check in database

106 domains

naylorantiques.com, netshelldemos.com, debbiebright.co.z, waldronfamilygppractice.co.uk , avea-vacances.com , psncodes2013.com uni5.net , 67.228.96.204, classificadosmaster.com.br, ibjjf.org

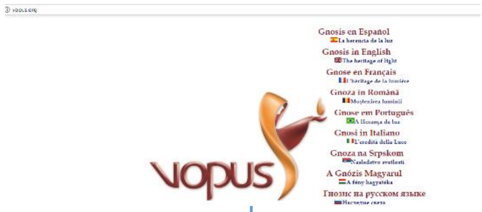
Visual Check



black hat EUROPE 2018

Threat Actor 2

Vopus.org



13 URLs

http://www.vopus.org/es/images/cursos/thumbs/tdcanadatrust

http://www.vopus.org/ru/media/**tdcanadatrust/index.html**

http://vopus.org/common/index.htm

http://www.vopus.org/es/images/cursos/thumbs/**tdcanadatrust/index.html**

http://vopus.org/descargas/otros/**tdcanadatrust/index.html**

Visual Check

Keywords

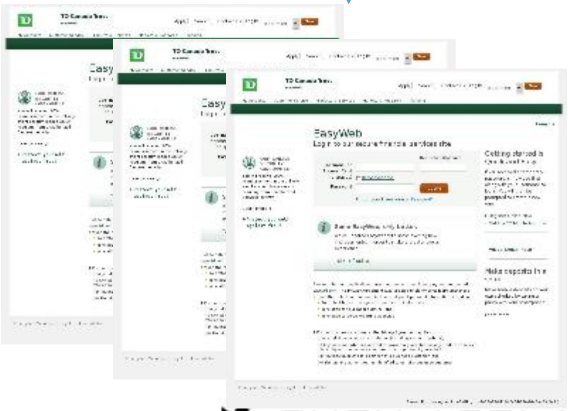
tdcanadatrust/index.html

Check in database

19 domains

friooptimo.com, kramerelementary.org, kalblue.com, vopus.org, artwood.co.kr, stephenpizzuti.com, heatherthinks.com, corvusseo.com, natikor.by, optioglobal.com, backfire.se, fncl.ma, greenant.de, mersintenisakademisi.com, cavtel.net

Visual Check



The Experiment: Simulating Malicious AI

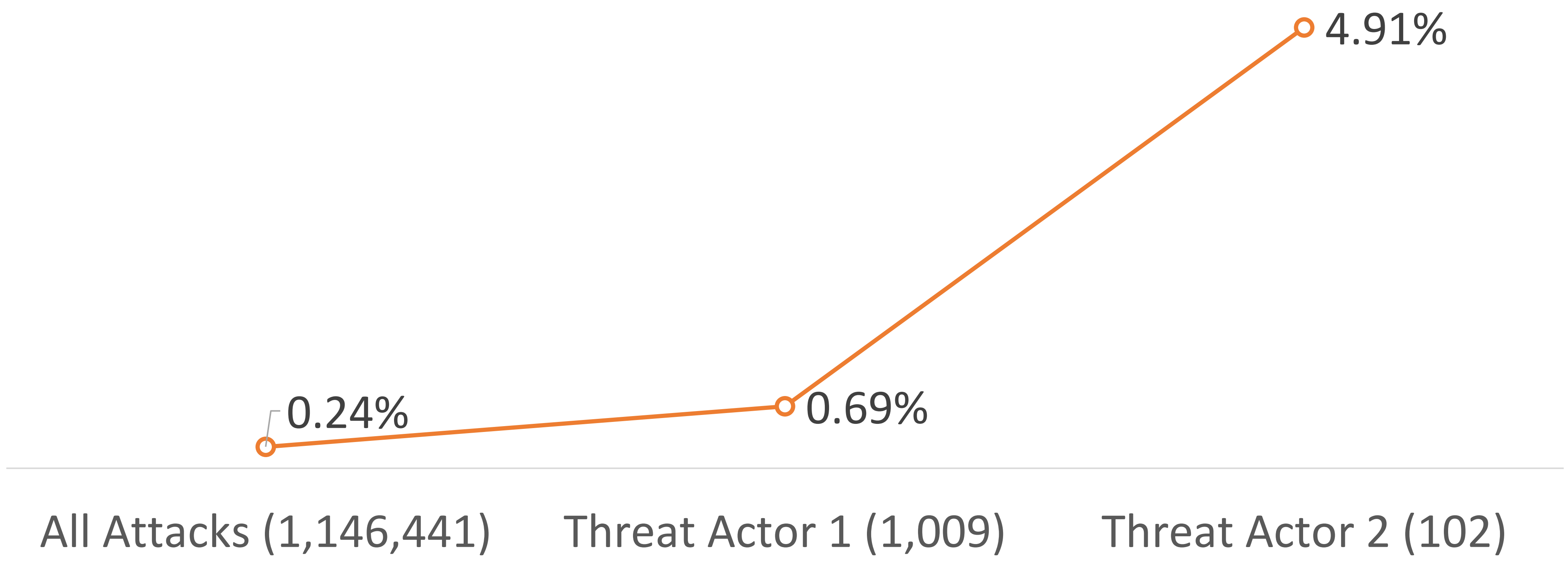
1 Identify individual threat actors

2 Run them through our own AI detection system

3 Improve their attacks using AI

DEMO!!!

Threat Actors Effectiveness



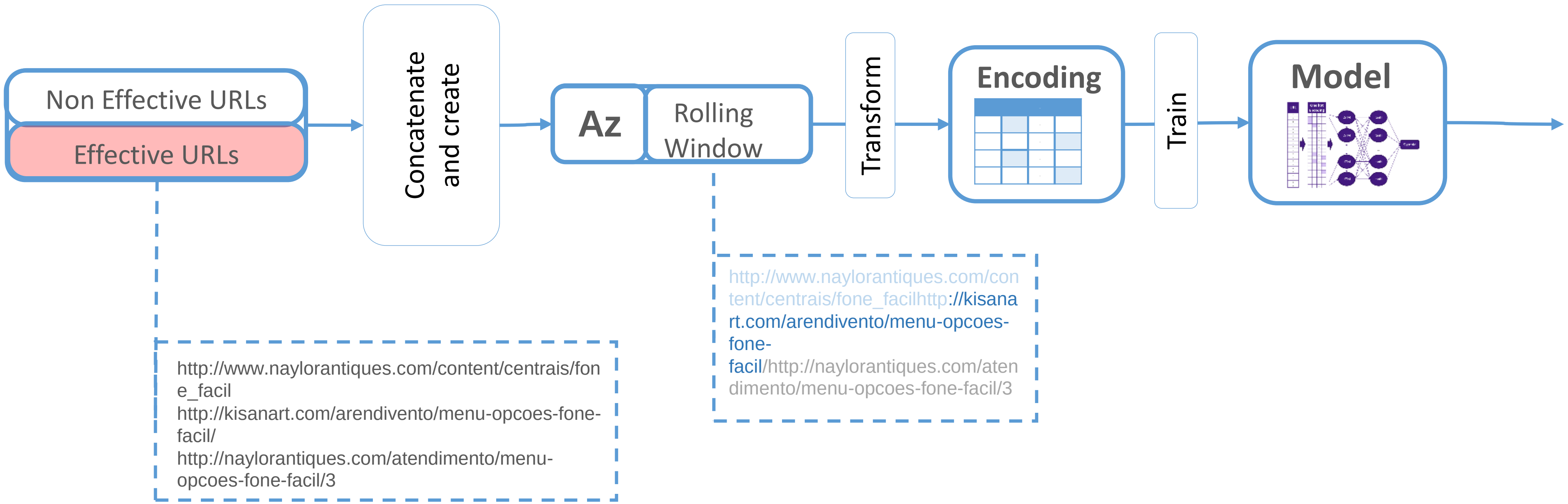
The Experiment: Simulating Malicious AI

1 Identify individual threat actors

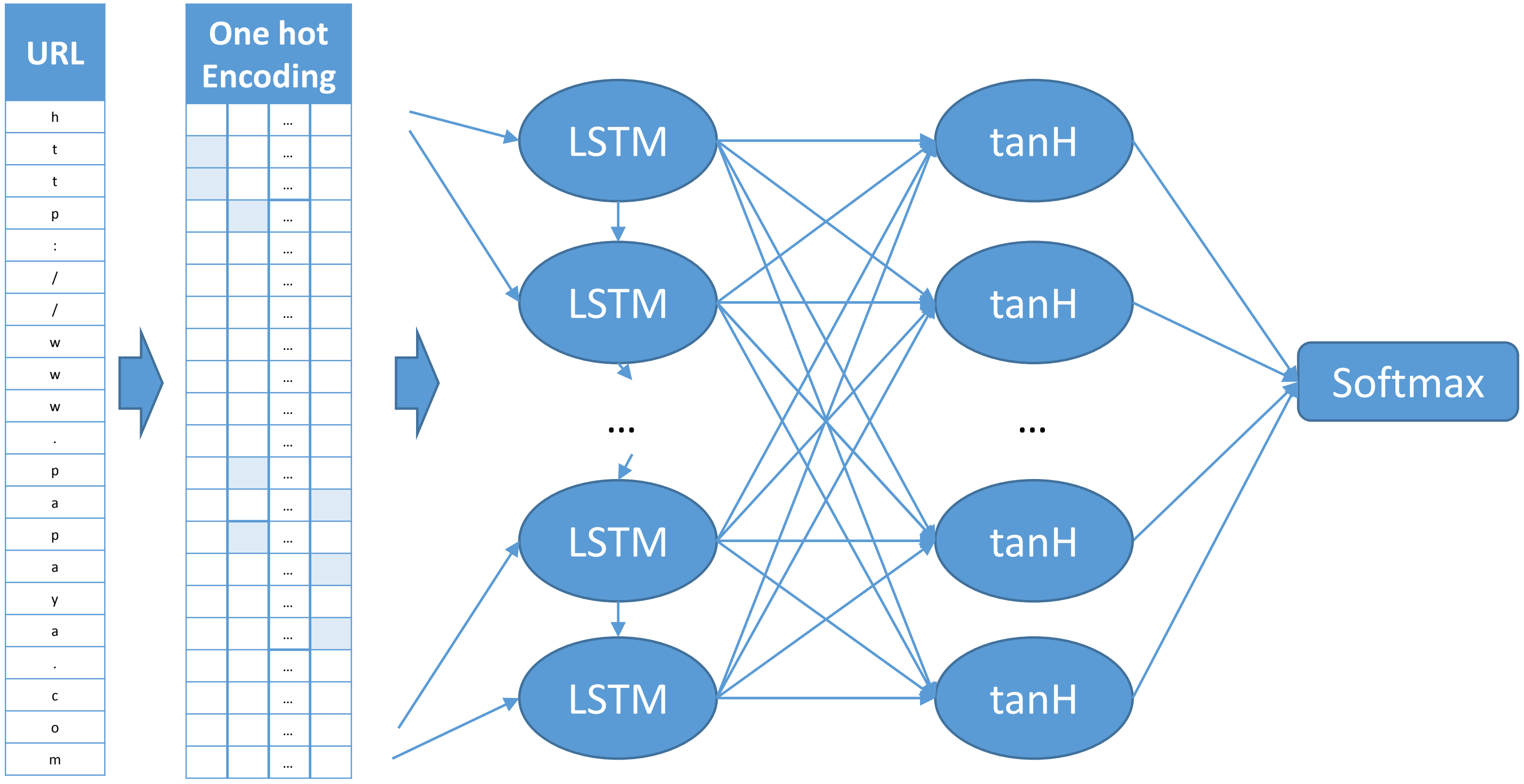
2 Run them through our own AI detection system

3 Improve their attacks using AI

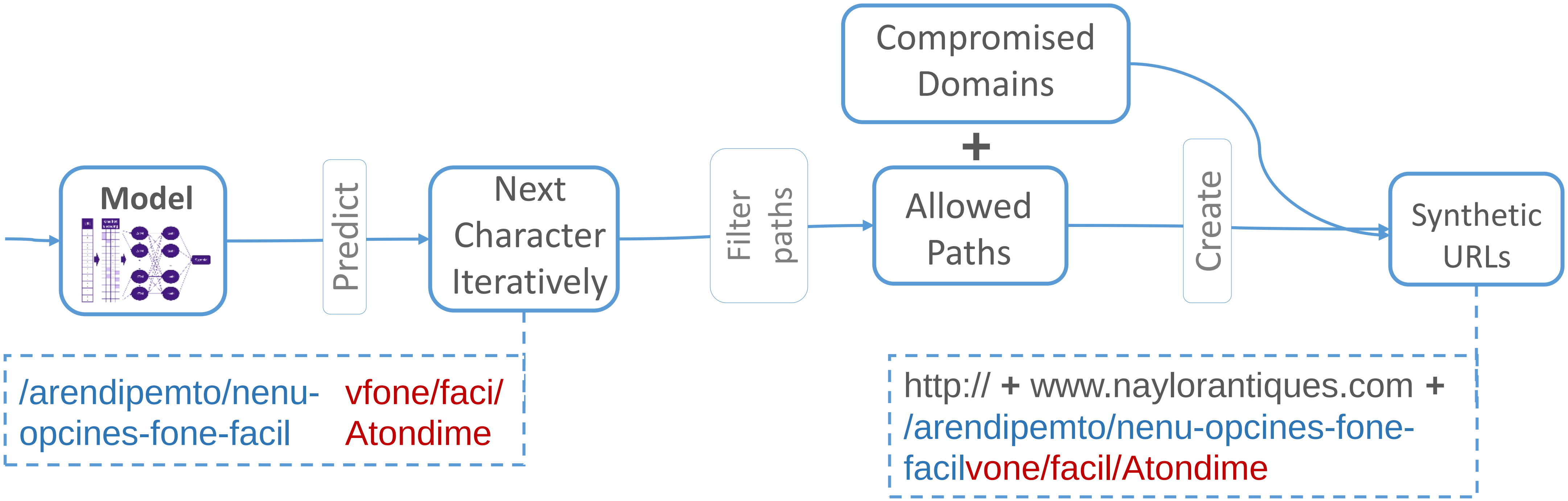
DeepPhish Algorithm - Training



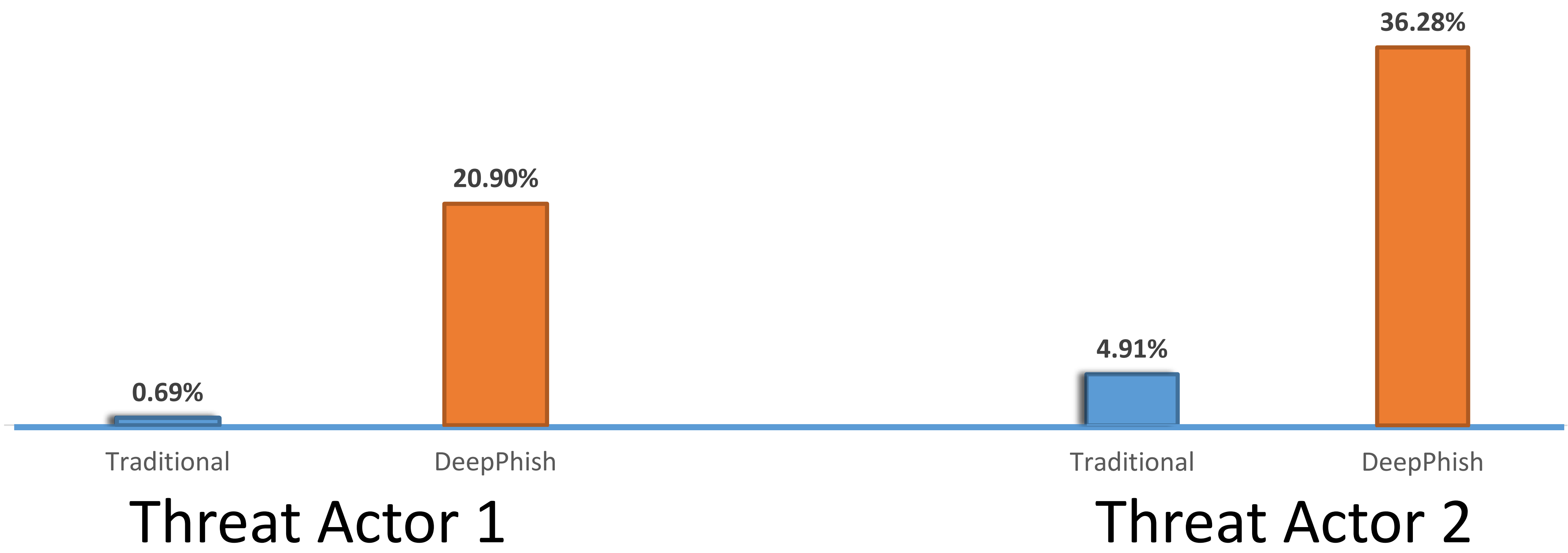
DeepPhish LSTM Network



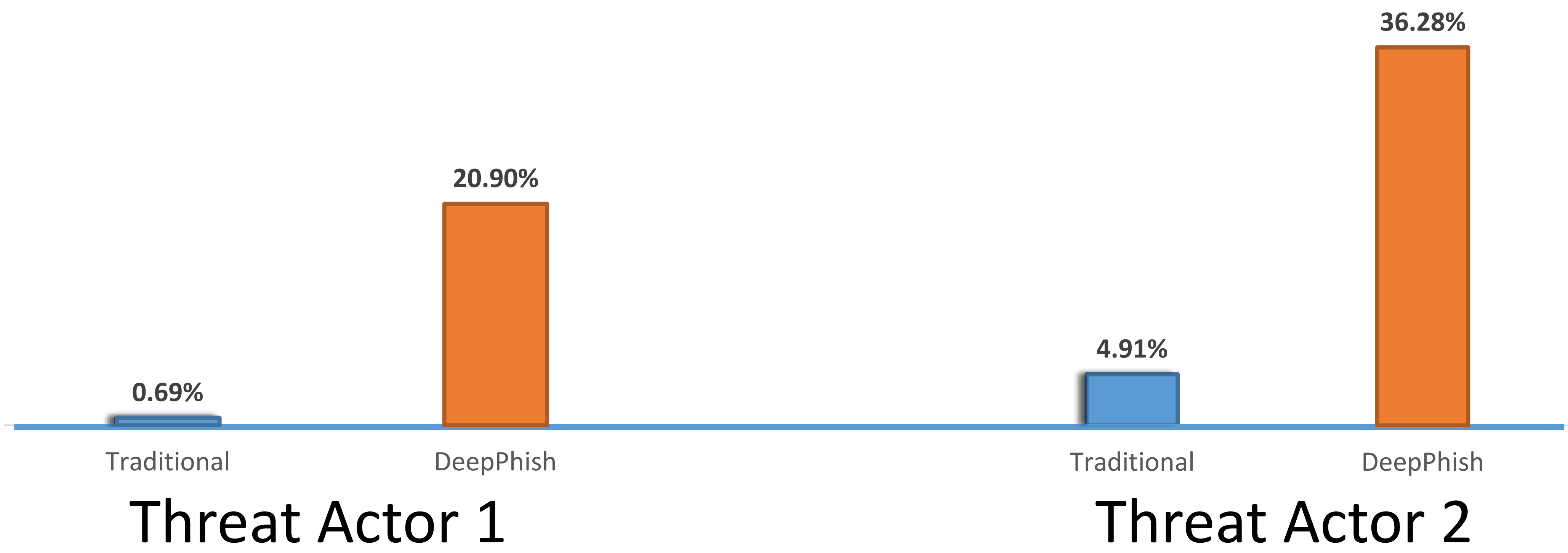
DeepPhish Algorithm - Prediction



Traditional Attacks vs. AI-Driven Attacks



Traditional Attacks vs. AI-Driven Attacks



What's Next??

What's Next??

AI powered Attacks are real, as we probed with Deep Phish experiment.
We need to enhance our own AI detection systems to account for the possibility of attackers using AI.

DeepPhish: Simulating Malicious AI

Alejandro Correa Bahnsen, Iv
Cyber Thr
Email: {alejandro.correa, ivan.t

Abstract—In this work we describe how threat
use AI algorithms to bypass AI phishing detection s
analyzed more than a million phishing URLs to und
different strategies that threat actors use to create phis
Assuming the role of an attacker, we simulate how diff
actors may leverage Deep Neural Networks to en
effectiveness rate. Using Long Short-Term Memory
we created DeepPhish, an algorithm that learns to c
phishing attacks. By training the DeepPhish algorit
different threat actors, they were able to increase the
ness from 0.69% to 20.9%, and 4.91% to 36.28%,

Keywords—*Malicious AI; phishing detection; cyb
current neural networks; long-short term memory net
adversarial learning.*

I. INTRODUCTION

Machine Learning (ML) and Artificial Intelli

Classifying Phishing URLs Using Recurrent

Alejandro Correa Bahnsen[†], Edu
Javier Vargas
[†]Easy
*MindLab Research Group,
Email: acorrea@easysol.net, econtrerasb@unal.edu.co,

Abstract—As the technical skills and costs associated
the deployment of phishing attacks decrease, we are witn
an unprecedented level of scams that push the need for
methods to proactively detect phishing threats. In this
we explored the use of URLs as input for machine lea
models applied for phishing site prediction. In this wa
compared a feature-engineering approach followed by a ra
forest classifier against a novel method based on recurrent n
networks. We determined that the recurrent neural net
approach provides an accuracy rate of 98.7% even withou
need of manual feature creation, beating by 5% the random
method. This means it is a scalable and fast-acting pro
detection system that does not require full content analysis

Keywords—*Phishing detection; Cybercrime; Feature engi
ing; Recurrent neural networks; Long short term memory netw*

I. INTRODUCTION

Hunting Malicious TLS Certificates with Deep Neural Networks

Ivan Torroledo, Cyxtera Technologies, ivan.torroledo@cyxtera.com
Luis David Camacho, Cyxtera Technologies, luis.camacho@cyxtera.com
Alejandro Correa Bahnsen, Cyxtera Technologies, alejandro.correa@cyxtera.com

ABSTRACT

Encryption is widely used across the internet to secure communications and ensure that information cannot be intercepted and read by a third party. However, encryption also allows cybercriminals to hide their messages and carry out successful malware attacks while avoiding detection. Further aiding criminals is the fact that web browsers display a green lock symbol in the URL bar when a connection to a website is encrypted. This symbol gives a false sense of security to users, who are in turn more likely to fall victim to phishing attacks. The risk of encrypted traffic means that information security researchers must explore new techniques to detect, classify, and take countermeasures against malicious traffic. So far there exists no approach for TLS detection in the wild. In this paper, we propose a method for identifying malicious use of web certificates using deep neural networks. Our system uses the content of TLS certificates to successfully identify legitimate certificates as well as malicious patterns used by attackers. The results show that our system is capable of identifying malware certificates with an accuracy of 94.87% and phishing certificates with an accuracy of 88.64%.

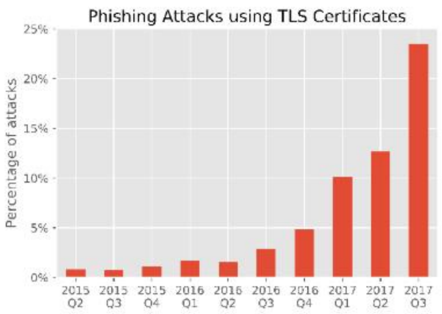


Figure 1: Evolution of phishing attacks using TLS [16].

black hat®
EUROPE 2018
DECEMBER 3-6, 2018
EXCEL LONDON / UNITED KINGDOM

Thanks!!

Alejandro Correa Bahnsen, PhD
VP, AI & Research
alejandro.correa@cyxtera.com
@albahnsen