

# Behavioral Biometrics and Context Analytics

Re-inventing authentication using Python

Jesus Solano

Data Scientist



# Risk-based static authentication in web applications with behavioral biometrics and session context analytics

Jesus Solano, Luis Camacho, Alejandro Correa, Claudio Deiro, Javier Vargas, and Martin Ochoa

Cytxera Technologies  
first.last@cytxera.com

**Abstract.** In order to improve the security of password-based authentication in web applications, it is a common industry practice to profile users based on their sessions context, such as IP ranges and Browser type. On the other hand, behavioral dynamics such as mouse and keyword features have been proposed in order to improve authentication, but have been shown most effective only in continuous authentication scenarios. In this paper we propose to combine both fingerprinting and behavioral dynamics (for mouse and keyboard) in order to increase security of login mechanisms. We do this by using machine learning techniques that aim at high accuracy, and only occasionally raise alarms for manual inspection. Our combined approach achieves an AUC of 0.957. We discuss the practicality of our approach in industrial contexts.

**Keywords:** behavioral dynamics, Static Authentication, Machine Learning

## 1 Introduction

With the increasing popularity of web services and cloud-based applications, we have also seen an increase on attacks to those platforms in the past decade. Several of those publicly known attacks have involved stealing of authentication credential to services (see for instance [10]). In addition to this, passwords are

# Risk-based static authentication in web applications with behavioral biometrics and session context analytics

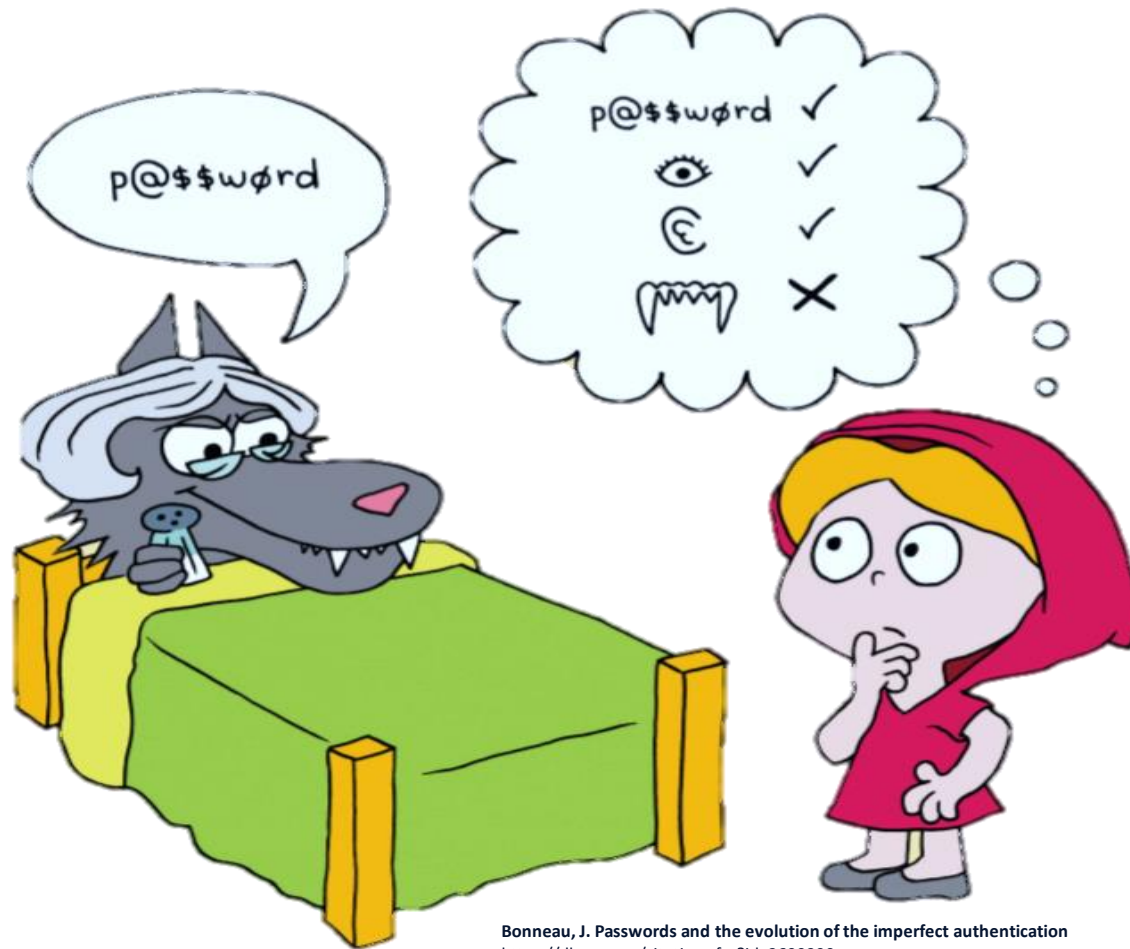
Jesus Solano, Luis Camacho, Claudio Deiro, Javier Vargas, Alejandro Correa, Martin Ochoa

17th International Conference on  
Applied Cryptography and  
Network Security

Best workshop paper award

1st International Workshop on Security in Machine Learning and its Applications

ACNS 2019



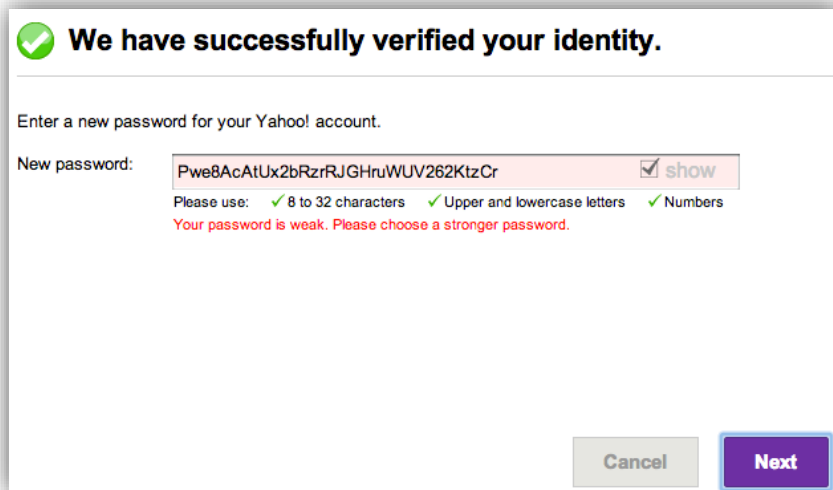
Bonneau, J. Passwords and the evolution of the imperfect authentication  
<https://dl.acm.org/citation.cfm?id=2699390>

# PASSWORD AUTHENTICATION IS NOT ENOUGH ANYMORE!

Even the “strongest” password  
can be stolen or broken

# Today's Password Model

Passwords are **hard for humans** to remember – but easy for computers to guess.



A screenshot of a web interface for creating a new password. At the top, a green checkmark icon is followed by the text "We have successfully verified your identity." Below this, the instruction "Enter a new password for your Yahoo! account." is displayed. A text input field labeled "New password:" contains the password "Pwe8AcAtUx2bRzrRJGHruWUV262KtzCr". To the right of the input field is a "show" button with a checked checkbox. Below the input field, there are three green checkmarks with corresponding text: "8 to 32 characters", "Upper and lowercase letters", and "Numbers". Below these, a red error message states: "Your password is weak. Please choose a stronger password." At the bottom right, there are two buttons: a grey "Cancel" button and a purple "Next" button.

Today's

Passwords

human

easy to



entity.

☒ show

letters ✓ Numbers

Cancel Next

# How are online services protecting identities?



# Current Approaches

Password  
Strength

Secrets Based  
on Challenges

Captcha to Identify  
Human vs Bot

Master Password

●●●●●●●●●●●●●●

Strength

Our minimum requirements:

- ✓ At least 12 characters long
- ✓ At least 1 number
- ✓ At least 1 lowercase letter
- ✓ At least 1 uppercase letter

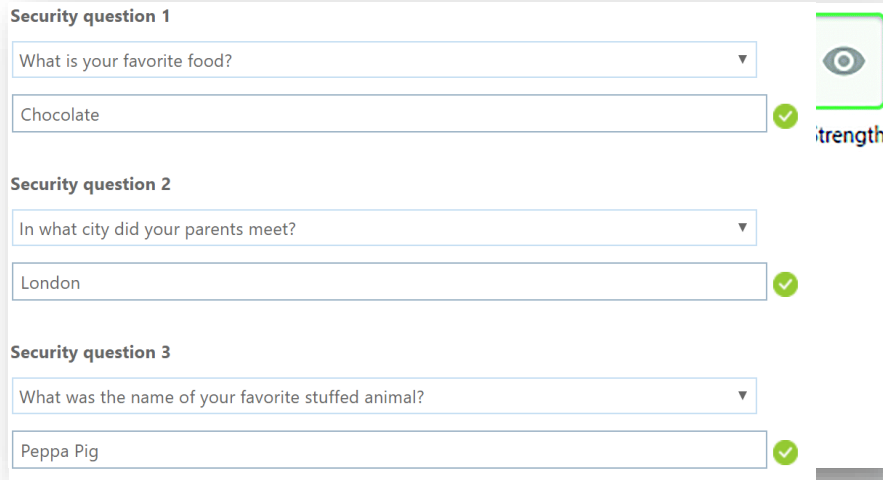


# Current Approaches

Password  
Strength

Secrets Based  
on Challenges

Captcha to Identify  
Human vs Bot



The screenshot shows a security questionnaire interface. It contains three sections, each with a question and an answer field. The first section is titled 'Security question 1' and asks 'What is your favorite food?'. The answer field contains 'Chocolate' and has a green checkmark. The second section is titled 'Security question 2' and asks 'In what city did your parents meet?'. The answer field contains 'London' and has a green checkmark. The third section is titled 'Security question 3' and asks 'What was the name of your favorite stuffed animal?'. The answer field contains 'Peppa Pig' and has a green checkmark. To the right of the questions is a green eye icon and the text 'Strength'.

Security question 1

What is your favorite food? ▼

Chocolate ✓

Security question 2

In what city did your parents meet? ▼

London ✓

Security question 3

What was the name of your favorite stuffed animal? ▼

Peppa Pig ✓

Strength

# Current Approaches

Password  
Strength

Secrets Based  
on Challenges

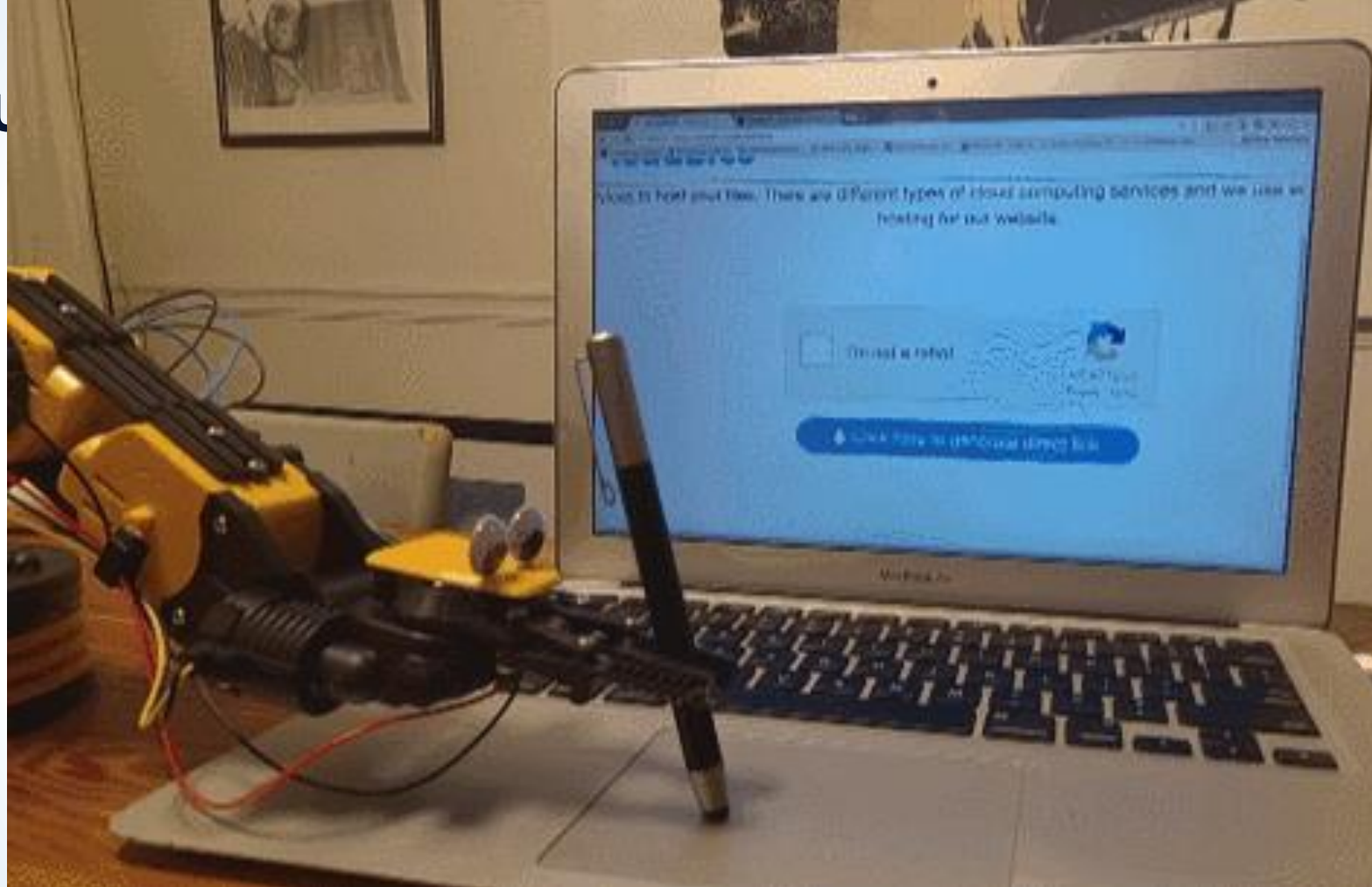
Captcha to Identify  
Human vs Bot

The image is a composite of three overlapping user interface elements representing different security approaches:

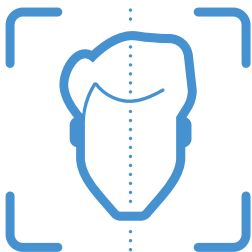
- Security Questions:** A form titled "Security question 1" with the prompt "What is your favorite food?" and the answer "Chocolate". Below it is "Security question 2" with the prompt "In what city did you grow up?" and the answer "London". At the bottom is "Security question 3" with the prompt "What was the name of your first pet?" and the answer "Peppa Pig".
- CAPTCHA:** A grid of nine images with the instruction "Select all images with trees." The images include: a bowl of fruit, a medicine bottle, a beach scene, a forest landscape, a pond with willows, a lake with a bridge, a town square, a mountain, and a basket of chestnuts.
- Password Strength:** A vertical bar on the right showing a password strength indicator with a green eye icon and the word "strength" partially visible. It includes three input fields, each followed by a green checkmark, indicating a strong password.

At the bottom of the CAPTCHA grid, there are icons for a refresh button, a help button, and a "Report a problem" link, along with a blue "Verify" button.

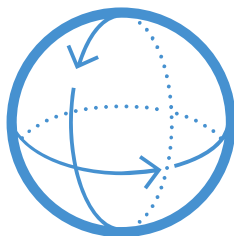
Cu



# Up and Coming Approaches



Passwordless  
Login



Continuous  
Authentication



Connection  
Behavior

# Validating Users' Identity

# Context-Based Authentication



Web-based  
fingerprinting



Connection  
features at each  
login request



Storage of  
users' pattern  
history

# Context-Based Authentication

## Device Fingerprinting for Augmenting Web Authentication: Classification and Analysis of Methods

Furkan Alaca

P.C. van Oorschot

School of Computer Science  
Carleton University, Ottawa, Canada

17' Alaca et al.

- Explores, summarizes and classifies **29 device fingerprinting mechanism** for authentication.
- Combining more vectors tends to improve spoofing resistance.
- Trade-off features vs intrusiveness

## Design of a Risk Based Authentication System using Machine Learning Techniques

Mohammed Misbahuddin<sup>1</sup>, B S Bindhumadhava<sup>2</sup>, B. Dheeptha<sup>3</sup>

<sup>1,2</sup>Computer Networks and Internet Engineering (CNIE) Division,

Centre for Development of Advanced Computing, Electronics City, Bangalore, India - 560100

<sup>3</sup>Dept. of Computer Science & Engineering, Sastra University, Thanjavur, Tamil Nadu - India - 613402

[misbah@cdac.in](mailto:misbah@cdac.in), [bindhu@cdac.in](mailto:bindhu@cdac.in), [dheepthab1210@gmail.com](mailto:dheepthab1210@gmail.com)

17' Misbahuddin et al.

- **Adaptative authentication** model with a user profile analyzer.
- One-class **SVM** to learn patterns from legitimate sessions.



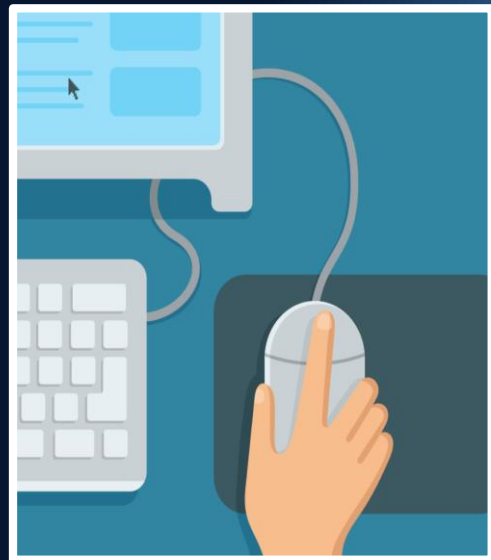
# Behavioral Biometrics

Long time frames required  
to achieve high accuracy

Used for continuous  
authentication, **not**  
**logins**

Privacy issues

Increasing privacy  
reduces accuracy



# Behavioral Biometrics

## Combining Keystroke and Mouse Dynamics for Continuous User Authentication and Identification

Soumik Mondal and Patrick Bours  
Norwegian University of Science and Technology (NTNU)  
Teknologivegen 22, 2815 Gjøvik, Norway  
{soumik.mondal, patrick.bours}@ntnu.no

16' Mondal et al.

- Keystrokes and mouse dynamics for continuous authentication.
- Identification accuracy of 62.2 %
- Focus on continuous identification.

## Combining Mouse and Keystroke Dynamics Biometrics for Risk-Based Authentication in Web Environments

Issa Traore, *Member IEEE*, Isaac Woungang, *Member IEEE*, \*Mohammad S. Obaidat, *Fellow of IEEE*,  
<sup>1</sup>Youssef Nakkabi, and <sup>2</sup>Iris Lai  
Ryerson University, Canada, Monmouth University, USA and Khalifa University, UAE  
<sup>\*</sup>Monmouth University, NJ, USA and Khalifa University, UAE

12' Traore et al.

- Web environments characterized by the limited amount of keystrokes and mouse
- Mouse dynamics and keystroke dynamics biometrics in a multimodal framework.

# MEANWILE IN REAL LIFE ...

# Data From The Wild

## Behavioral Data

---

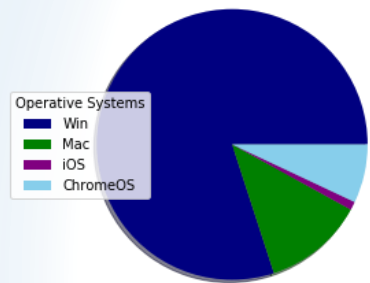
**TWOS**  
320 hours of  
human-computer  
interaction in a  
gamified  
environment

## Context Data

---

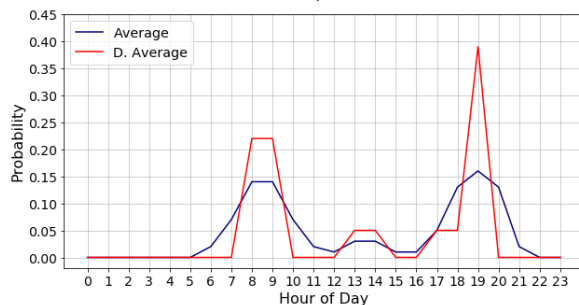
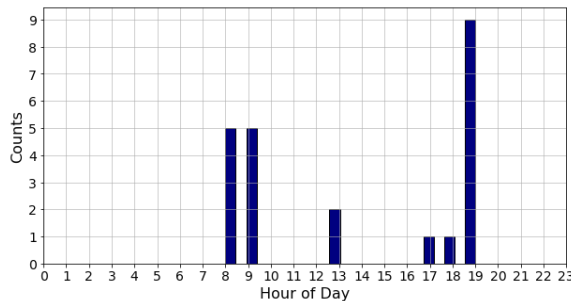
**Financial Clients  
Logins**  
Data from more  
than 2 million real  
users, collected by  
our company's  
products

# Feature Extraction



$$Win = \frac{75}{100} \quad Mac = \frac{13}{100} \quad ChromeOS = \frac{10}{100} \quad iOS = \frac{2}{100}$$

$$Win = \frac{100}{100} \Rightarrow Mac = \frac{25}{100} \Rightarrow ChromeOS = \frac{12}{100} \Rightarrow iOS = \frac{2}{100}$$

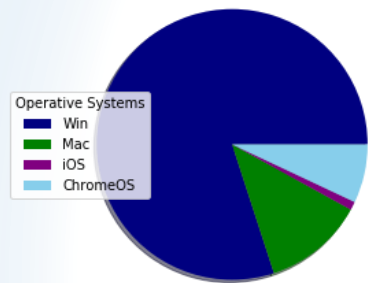


# Session Context

## Financial Clients Logins

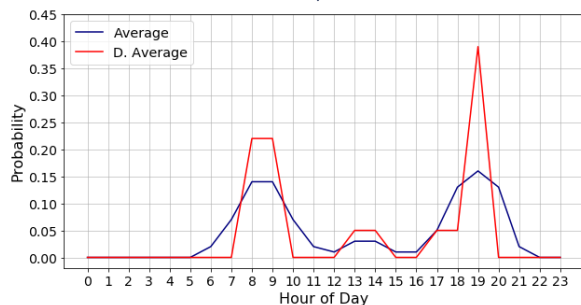
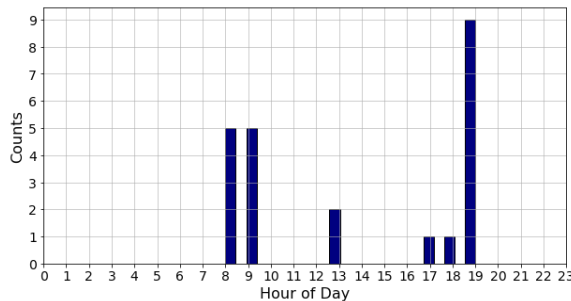
Data from more than 2 million real users, collected by our company's products

# Feature Extraction



$$Win = \frac{75}{100} \quad Mac = \frac{13}{100} \quad ChromeOS = \frac{10}{100} \quad iOS = \frac{2}{100}$$

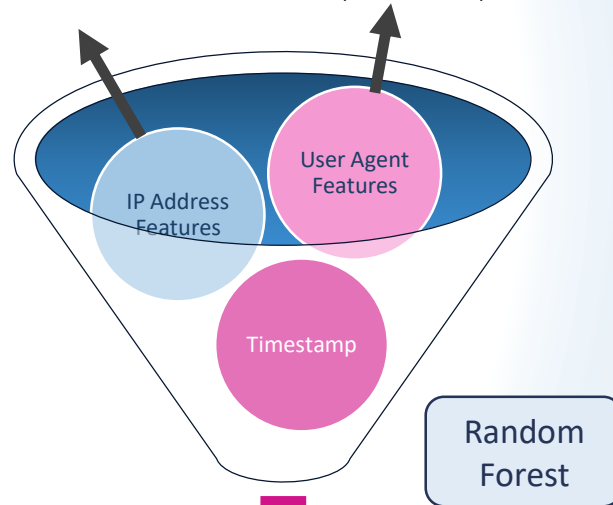
$$Win = \frac{100}{100} \Rightarrow Mac = \frac{25}{100} \Rightarrow ChromeOS = \frac{12}{100} \Rightarrow iOS = \frac{2}{100}$$



# Session Context

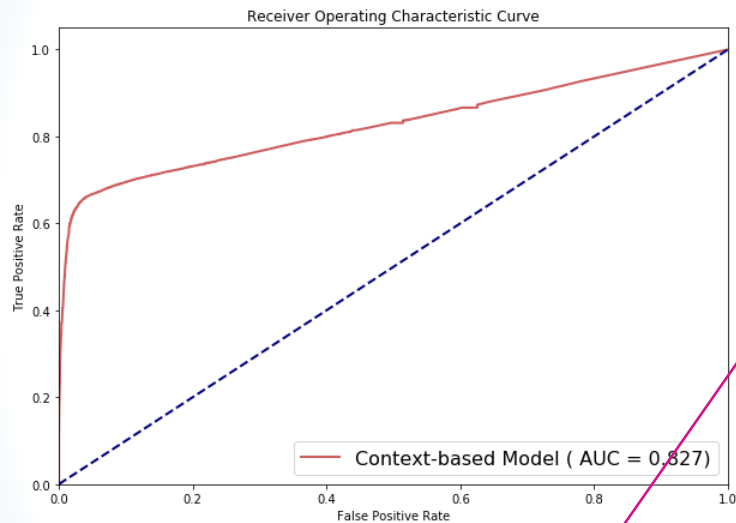
Rank of:  
Country, Region

Rank of:  
Browser, browser-version,  
OS, OS-version, device

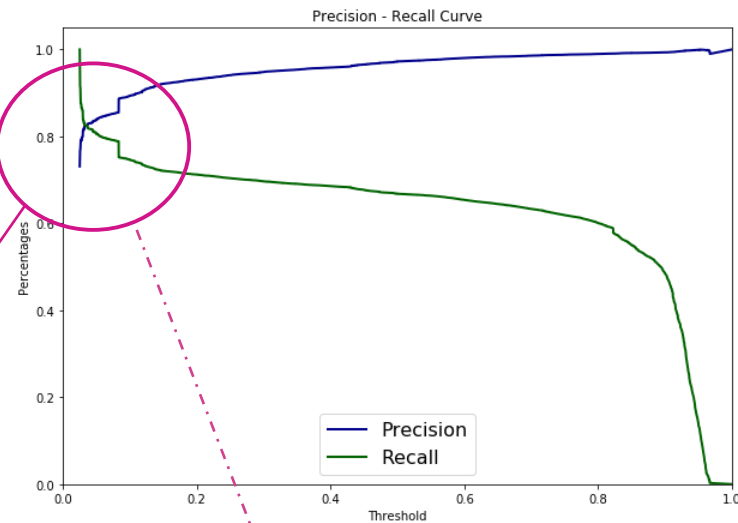


Context Risk Output  
[0,1]

# Testing Context Based Authentication - *On Its Own*



Good performance only  
at lower thresholds



High sensitivity leads  
high misclassification



# Testing Context Based Authentication - *On Its Own*

Decision Threshold	Precision	Accuracy	Recall
0.3	0.948	0.750	0.697
0.5	0.972	0.743	0.668
0.7	0.986	0.725	0.633

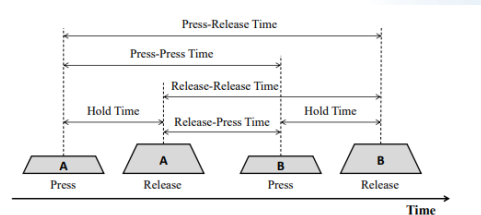
- 
- ➡ Accurate but has a high number of false positives
  - ➡ Sensitive to device impersonation

# Behavioral Data

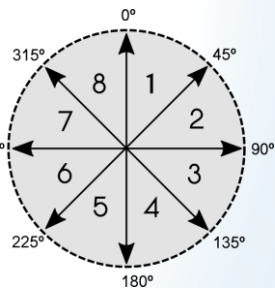
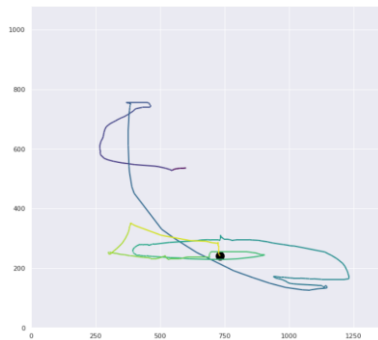
**TWOS**  
320 hours of  
human-computer  
interaction in a  
gamified  
environment

## Feature Extraction

### Keyboard Feature Engineering



### Mouse Feature Engineering

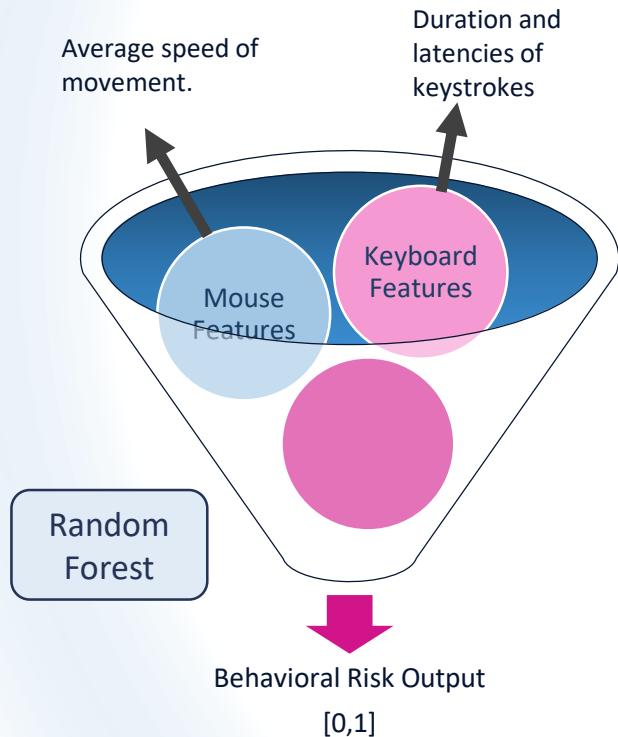


...But this data does not come from logins

# Behavioral Data

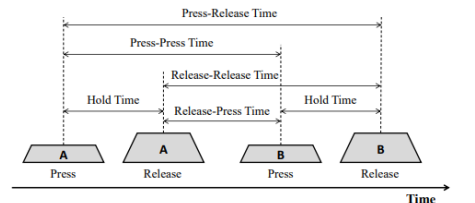
Average speed of movement.

Duration and latencies of keystrokes

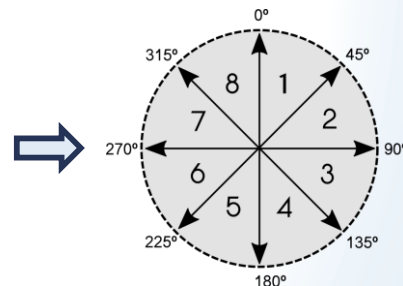
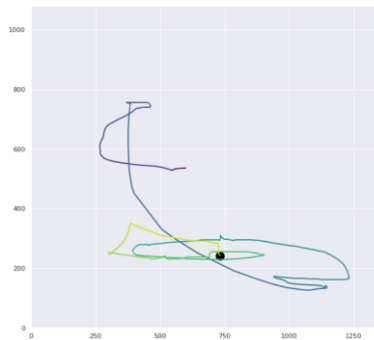


# Feature Extraction

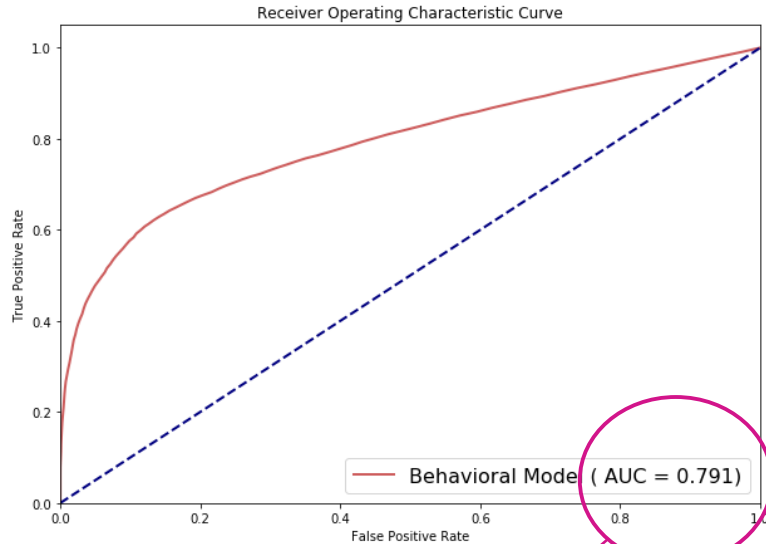
## Keyboard Feature Engineering



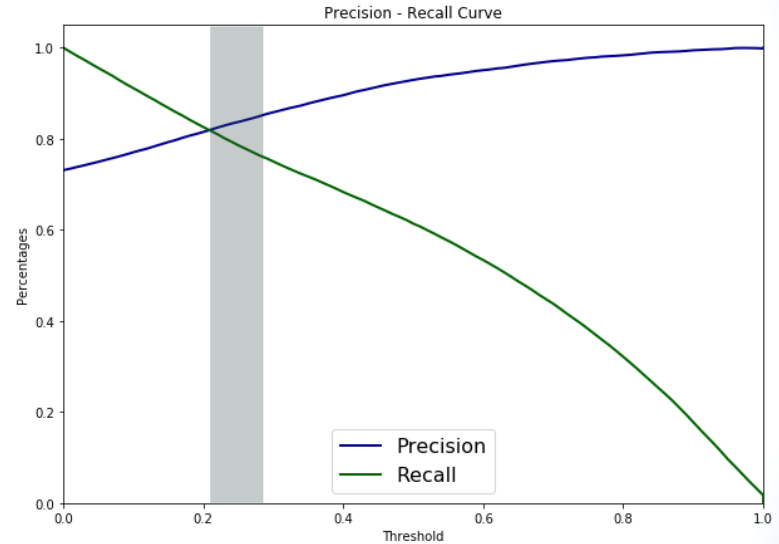
## Mouse Feature Engineering



# Testing Behavioral Biometrics – *On Its Own*




**Better than state-of-art  
separability performance**



**Best classification  
threshold**










# Testing Behavioral Biometrics - *On Its Own*

Decision Threshold	Precision	Accuracy	Recall
0.3	0.862	0.725	0.743
0.5	0.932	0.680	0.607
0.7	0.972	0.572	0.427

- 
- ☛ Accurate but sensitive on training data
  - ☛ May produce a high number of false positives

# Who are we defending users from?

# Attack Types

SIMPLE	CONTEXT	PHYSICAL
		
		
		

CONTEXT BASED



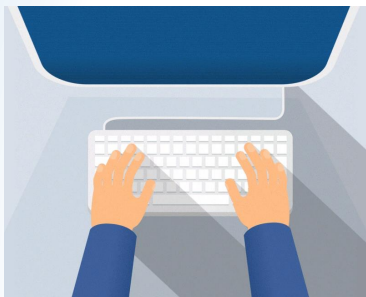
BEHAVIORAL BIOMETRICS



**COMBINED APPROACH**



# Proposed Model



Behavioral  
Biometrics



Web-Based  
Fingerprinting



Enhanced  
Risk-Based  
Authentication

## How?

Parametric linear  
combination of both  
machine learning  
algorithm outputs

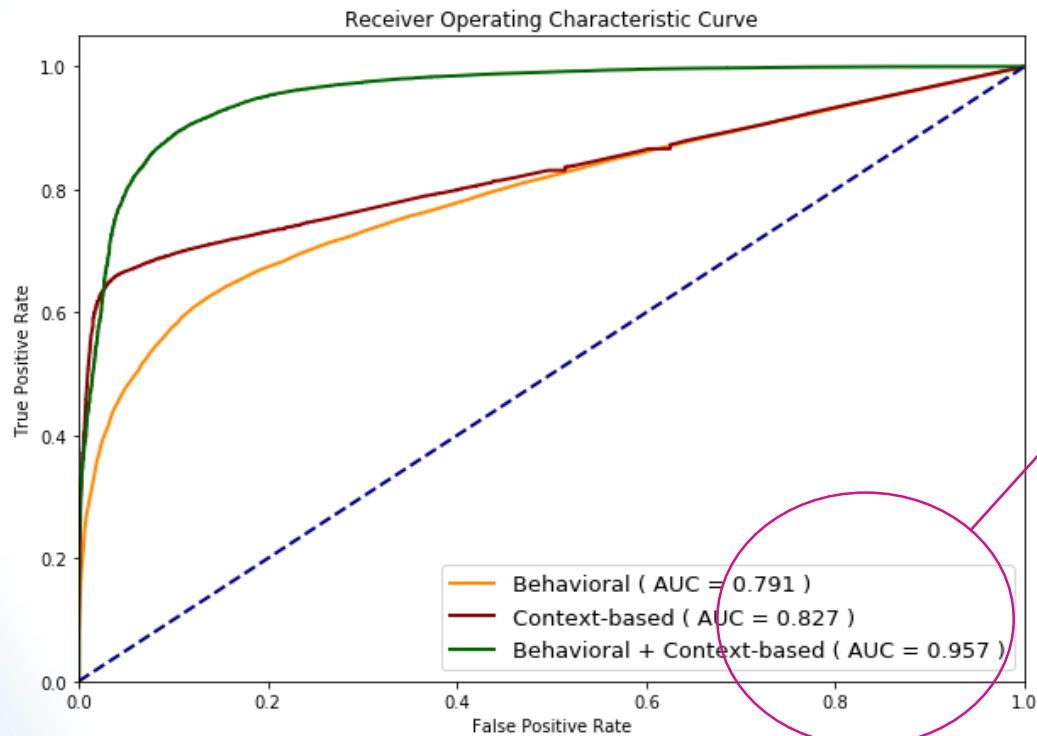


$$\hat{y}_t = \alpha_c \cdot \hat{y}_c + \alpha_b \cdot \hat{y}_b$$

Context Risk      Behavioral Risk

$\alpha_c + \alpha_b = 1$  ← Coefficient Parameters

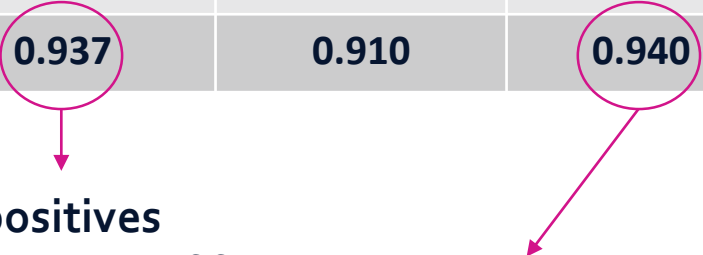
# Proposed Model Results



20.98% better AUC than the worst single model

# Proposed Model Results

Decision Threshold	F1-Score	Precision	Accuracy	Recall
Behavioral	0.798	0.862	0.725	0.743
Context-Based	0.803	0.948	0.750	0.697
<b>Behavioral + Context-Based</b>	<b>0.939</b>	<b>0.937</b>	<b>0.910</b>	<b>0.940</b>

- 
- ✓ Reduce friction by reducing false positives
  - ✓ Increase security by reducing the number of false negatives
  - ✓ Increase robustness of identity verification focused on the user

# Conclusions

- ✓ The proposed model outperforms both individual models.
- ✓ Our proposed model reduces friction and increases security in static authentication.
- ✓ Our proposed model is easily extensible to continuous authentication.

# Takeaways

Everyone's **behavior is** a reflection of their **identity** – learn from it!

Reimagining security is about **more certainty** and **less friction**.

**Python** is a tool to **make** the limitless of science a **reality** – play with it!

# Question & Answers



**Jesus Solano**  
Data Scientist

[linkedin.com/in/jesus-solano-go](https://www.linkedin.com/in/jesus-solano-go)  
[Jesus.solano@Cyxtera.com](mailto:Jesus.solano@Cyxtera.com)

