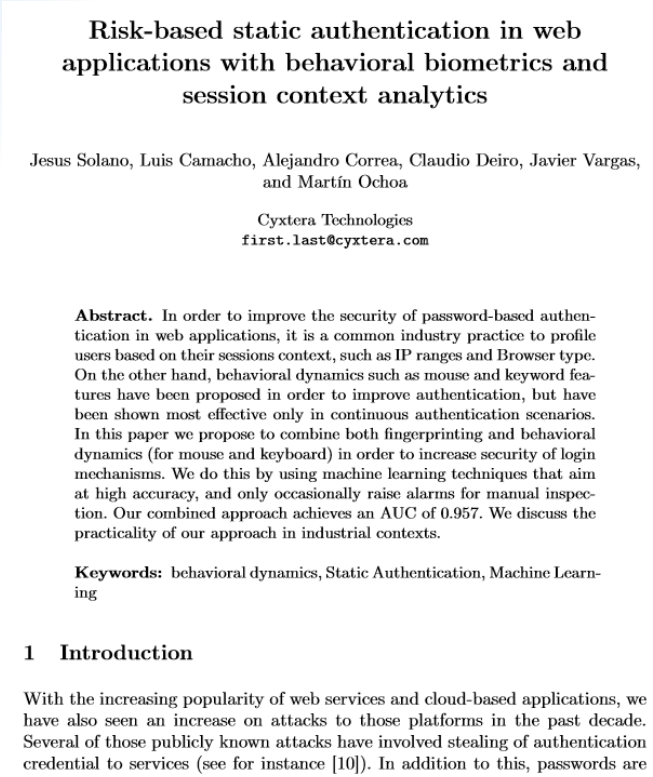
Behavioral Biometrics and Context Analytics

Re-inventing authentication using Python

Jesus Solano

Data Scientist

Cyxtera™



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 Martín Ochoa

Cyxtera Technologies
first.last@cyxtera.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

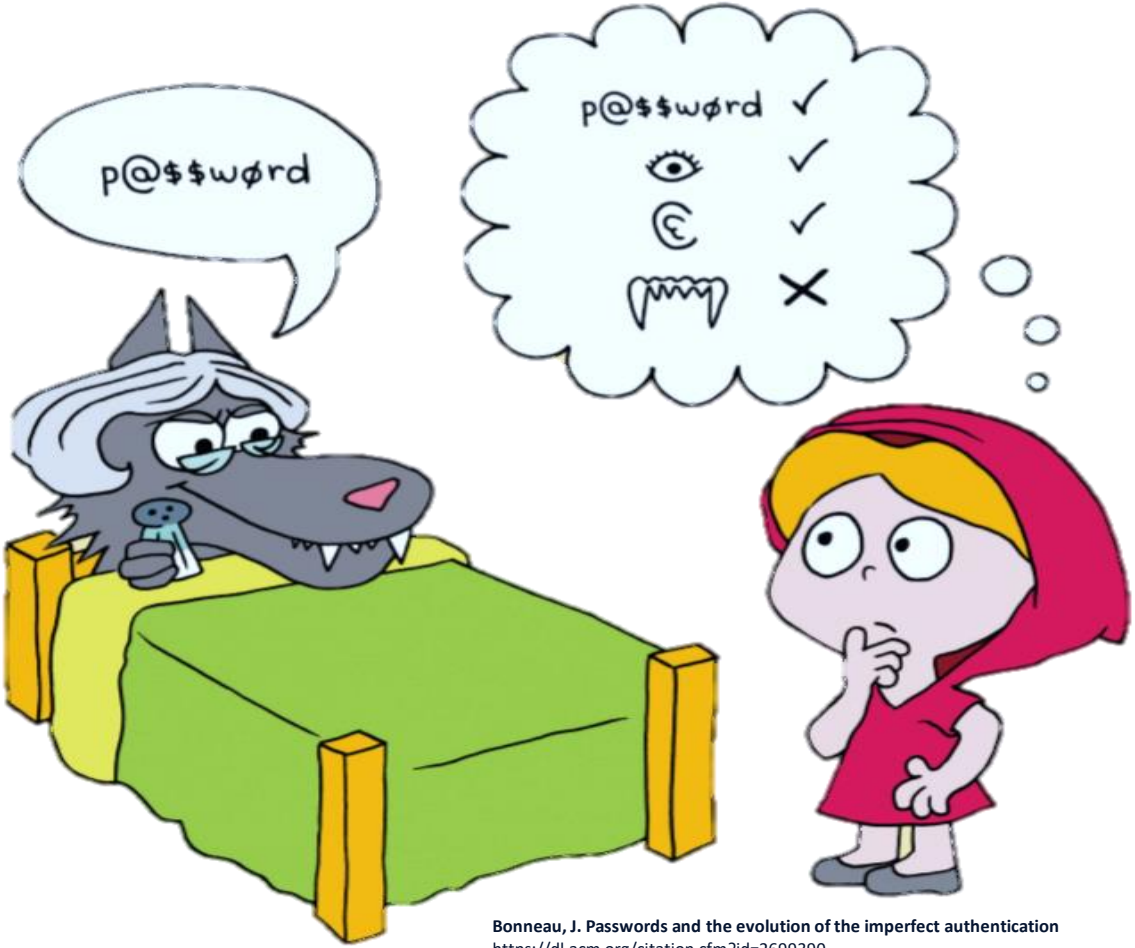
Jesus Solano, Luis Camacho, Claudio Deiro, Javier Vargas, Alejandro Correa, Martin Ochoa

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



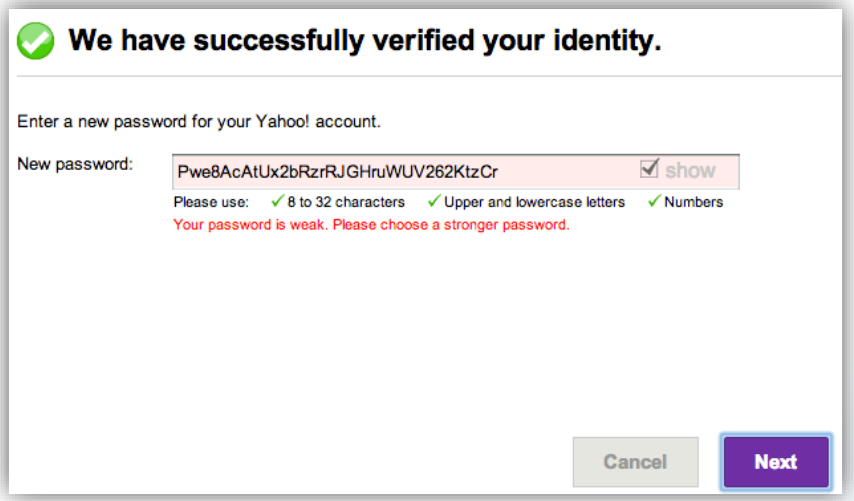
Bonneau, J. Passwords and the evolution of the imperfect authentication
https://dl.acm.org/citation.cfm?id=2699390

PASSWORD AUTHENTICATION IS NOT ENOUGH ANYMORE!

Even the “strongest” password
can be stolen or broken

Todays Password Model

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

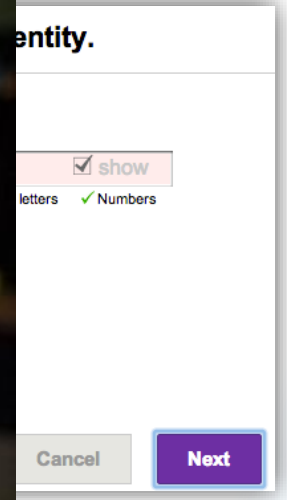


Todays

Passw

huma

easy f



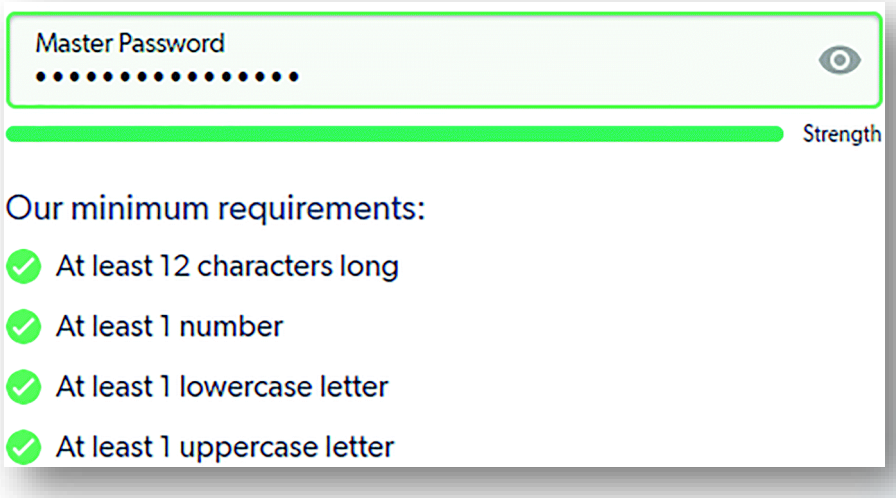
How are online services protecting identities?

Current Approaches

Password Strength

Secrets Based on Challenges

Captcha to Identify Human vs Bot

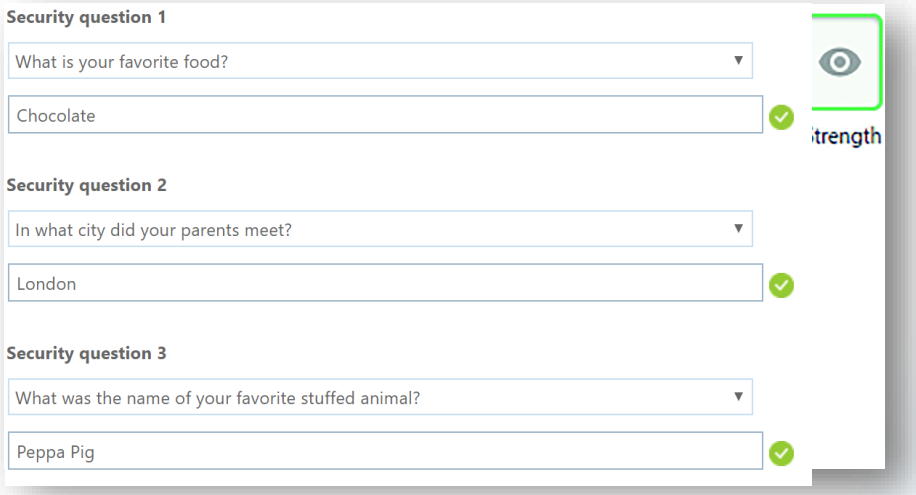


Current Approaches

Password Strength

Secrets Based on Challenges

Captcha to Identify Human vs Bot

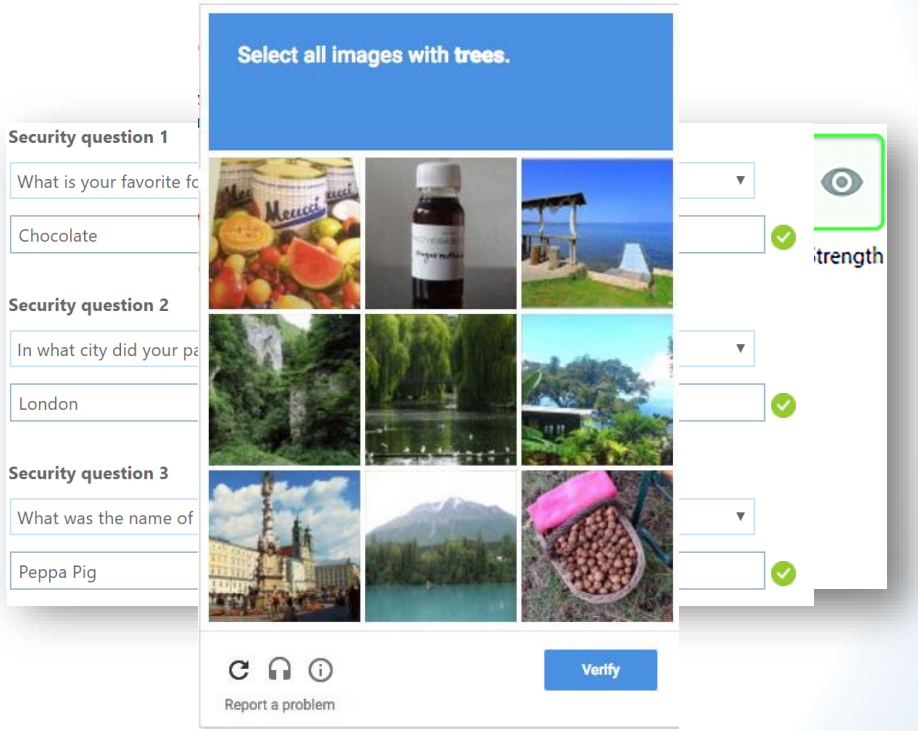


Current Approaches

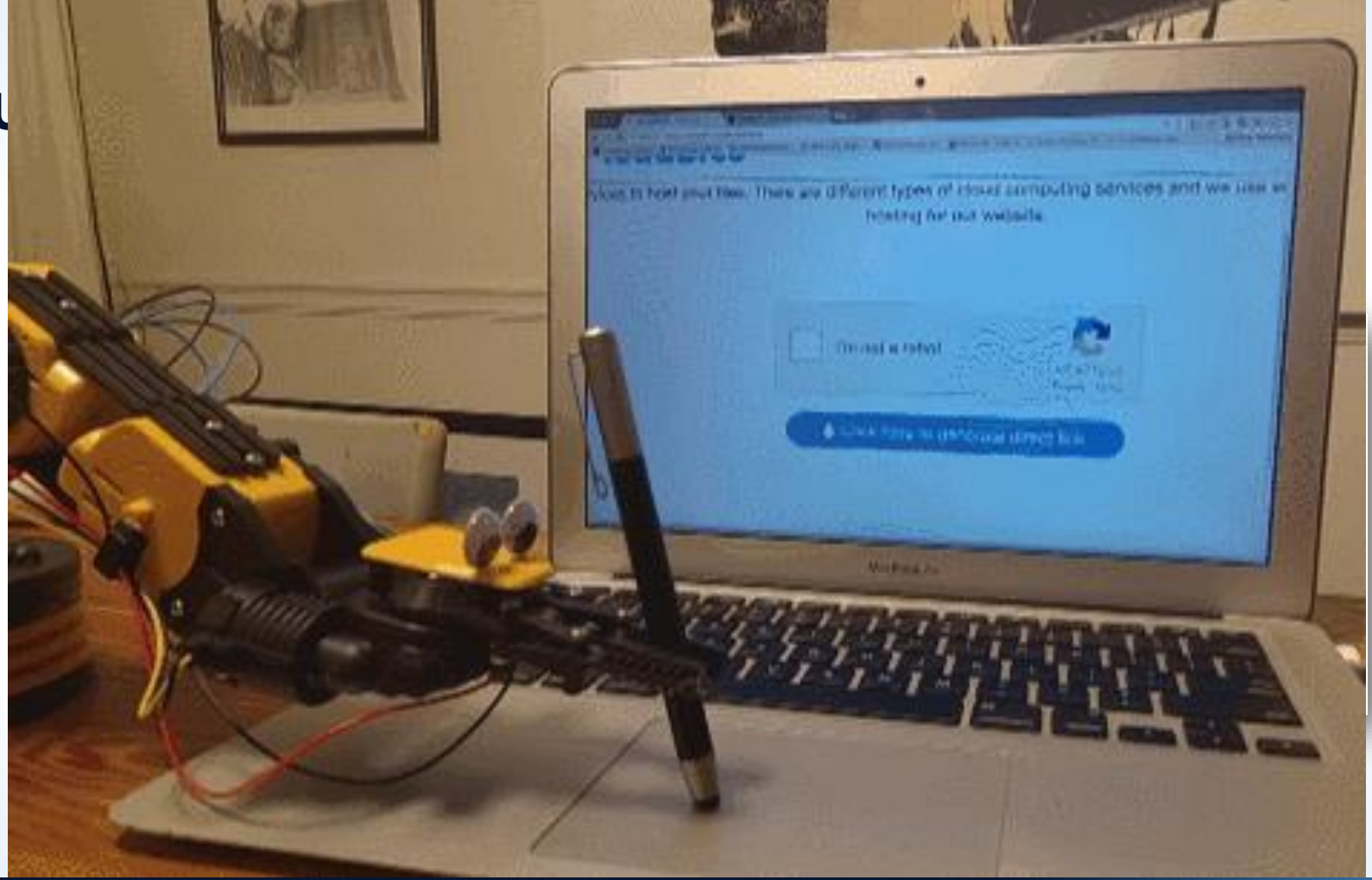
Password Strength

Secrets Based on Challenges

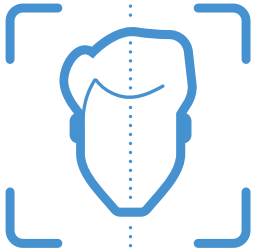
Captcha to Identify Human vs Bot



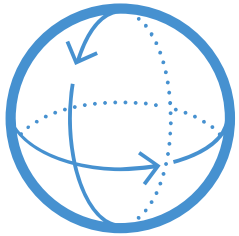
Cu



Up and Coming Approaches



Passwordless Login



Continuous Authentication

Connection Behavior

Validating Users' Identity

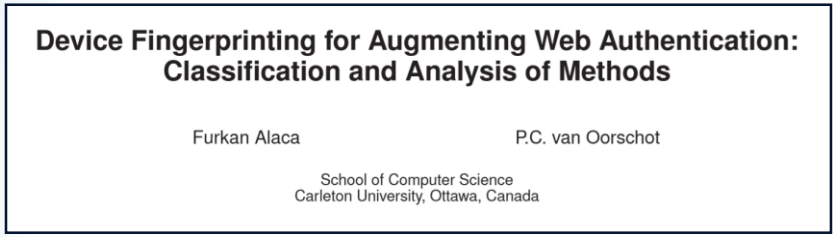
Context-Based Authentication

Web-based fingerprinting

Connection features at each login request

Storage of users' pattern history

Context-Based Authentication



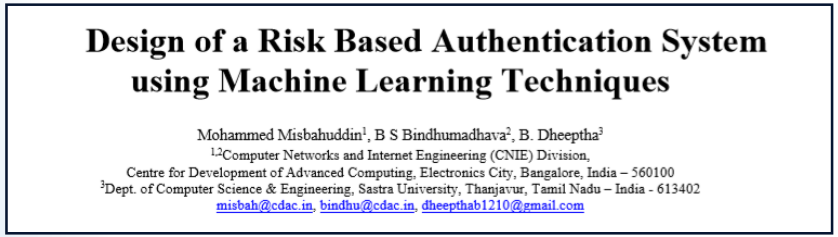
Device Fingerprinting for Augmenting Web Authentication: Classification and Analysis of Methods

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[1], B S Bindhumadhava[2], B. Dheeptha[3]
[1,2]Computer Networks and Internet Engineering (CNIE) Division,
Centre for Development of Advanced Computing, Electronics City, Bangalore, India – 560100
[3]Dept. of Computer Science & Engineering, Sastra University, Thanjavur, Tamil Nadu – India - 613402
misbah@cdac.in, bindhu@cdac.in, 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

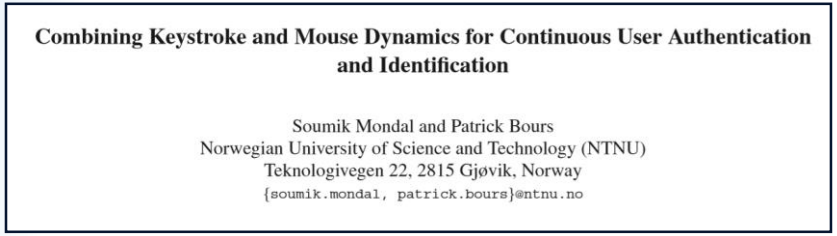
Long time frames required to achieve high accuracy

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

Privacy issues

- Increasing privacy reduces accuracy

Behavioral Biometrics

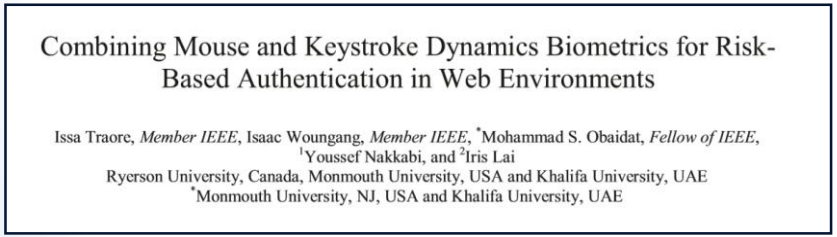


Combining Keystroke and Mouse Dynamics for Continuous User Authentication and Identification

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*,
[1]Youssef Nakkabi, and [2]Iris Lai
Ryerson University, Canada, Monmouth University, USA and Khalifa University, UAE
*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

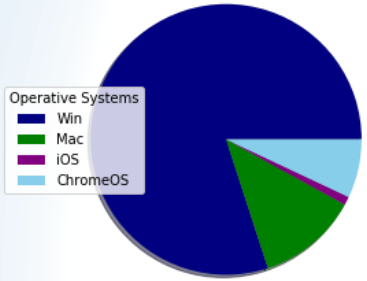
320 hours of human-computer interaction in a gamified environment

Context Data

Financial Clients Logins

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

Feature Extraction



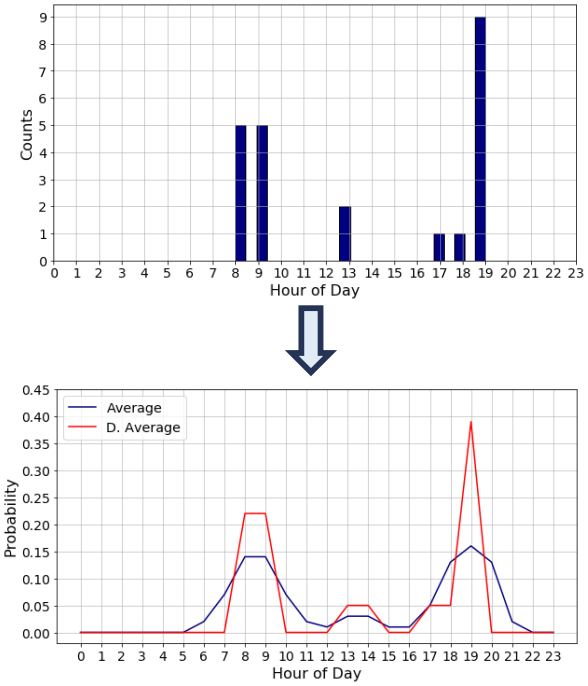
$Win = \frac{75}{100}$ $Mac = \frac{13}{100}$ ChromeOS= $\frac{10}{100}$ iOS= $\frac{2}{100}$

Win= $\frac{100}{100}$

$Mac = \frac{25}{100}$

ChromeOS= $\frac{12}{100}$

$iOS = \frac{2}{100}$

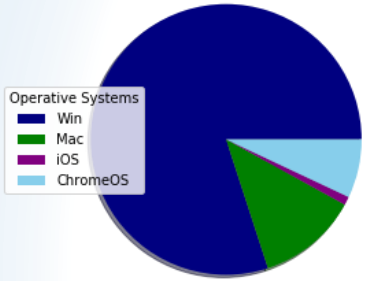


Session Context

Financial Clients Logins

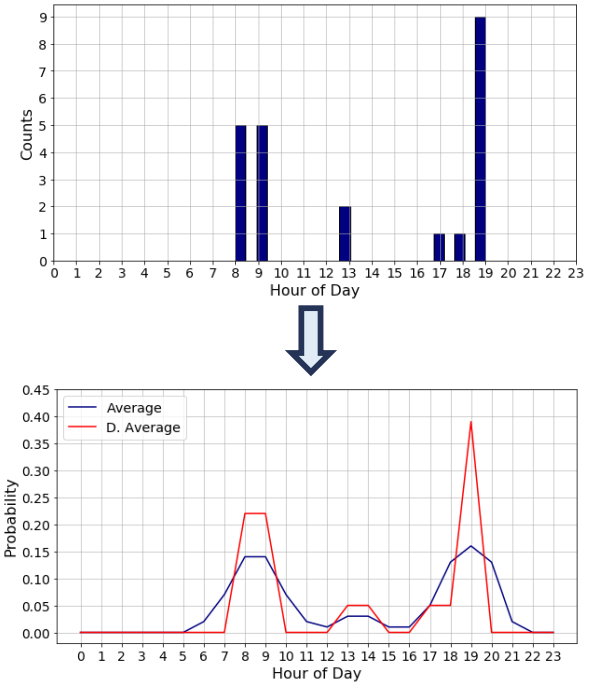
Data from more than 2 million real users, collected by our company's products

Feature Extraction



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

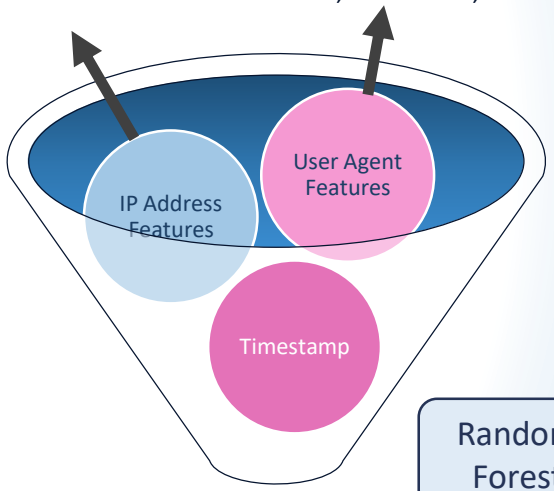
Win$= \frac{100}{100}$ → $Mac = \frac{25}{100}$ → ChromeOS$= \frac{12}{100}$ → $iOS = \frac{2}{100}$



Session Context

Rank of:
Country, Region

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

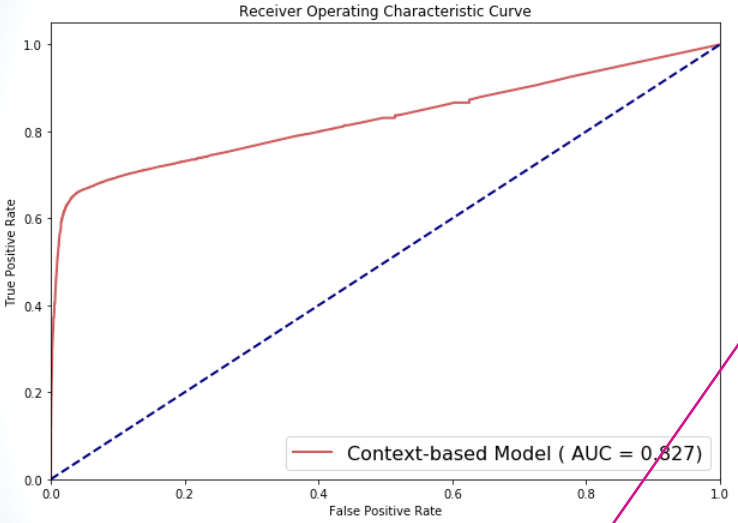


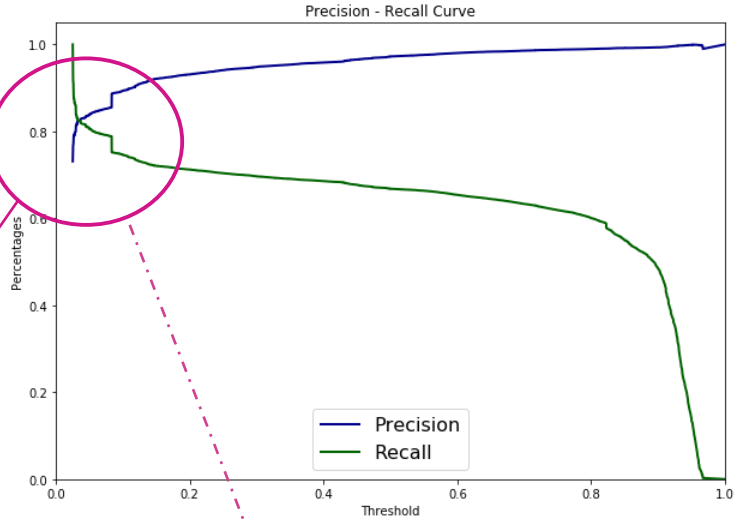
Random Forest

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

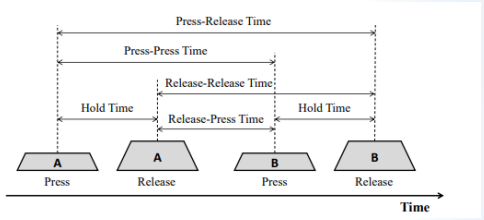
Behavioral Data

TWOS

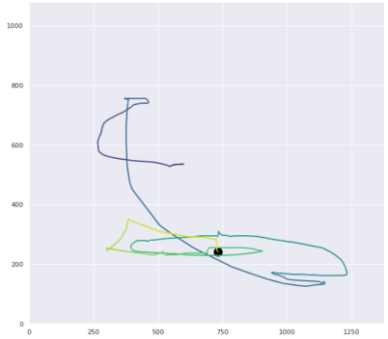
320 hours of human-computer interaction in a gamified environment

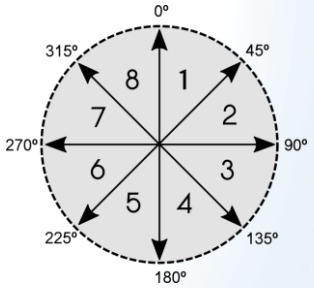
Feature Extraction

Keyboard Feature Engineering



Mouse Feature Engineering



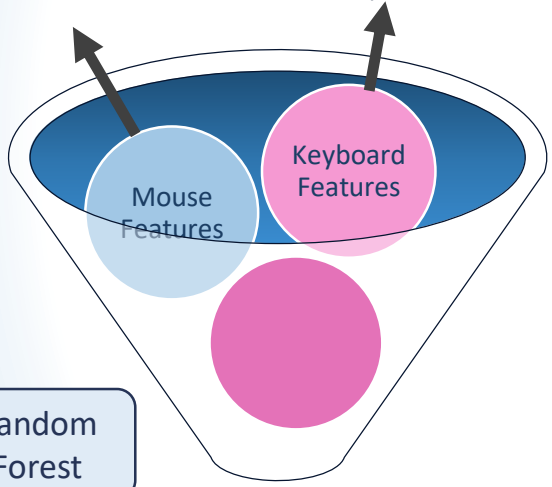


...But this data does not come from logins

Behavioral Data

Average speed of movement.

Duration and latencies of keystrokes



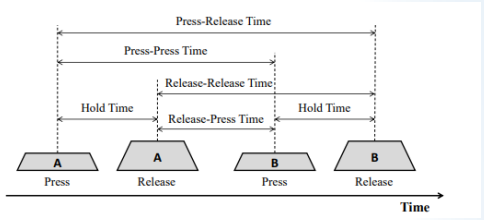
Random Forest

Behavioral Risk Output

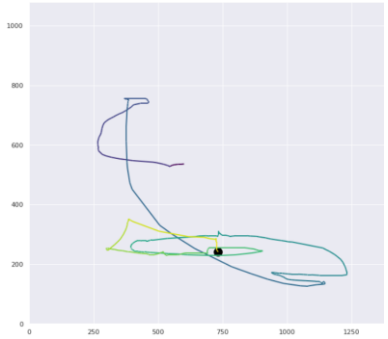
[0,1]

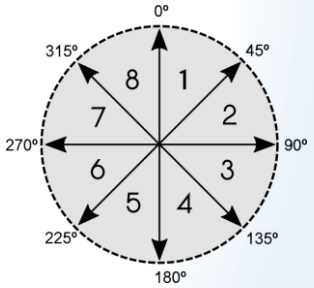
Feature Extraction

Keyboard Feature Engineering

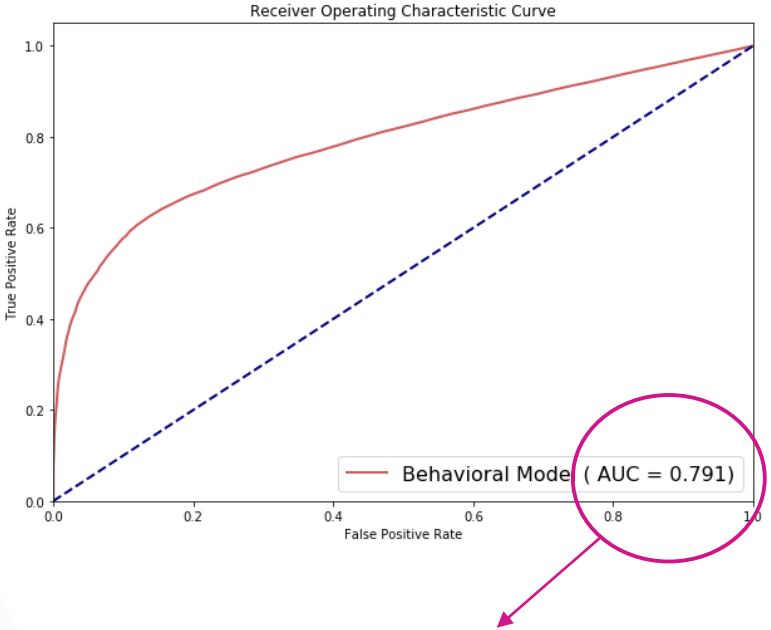


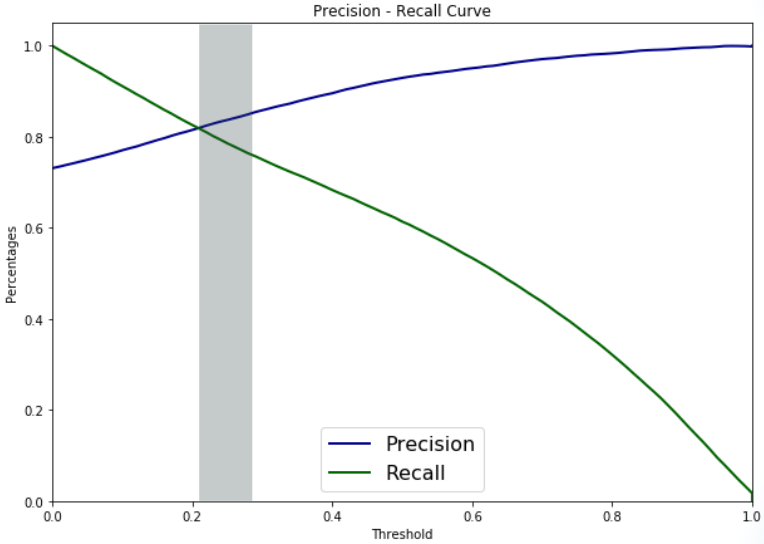
Mouse Feature Engineering





Testing Behavioral Biometrics – *On Its Own*





Better than state-of-art separability performance

Best classification threshold

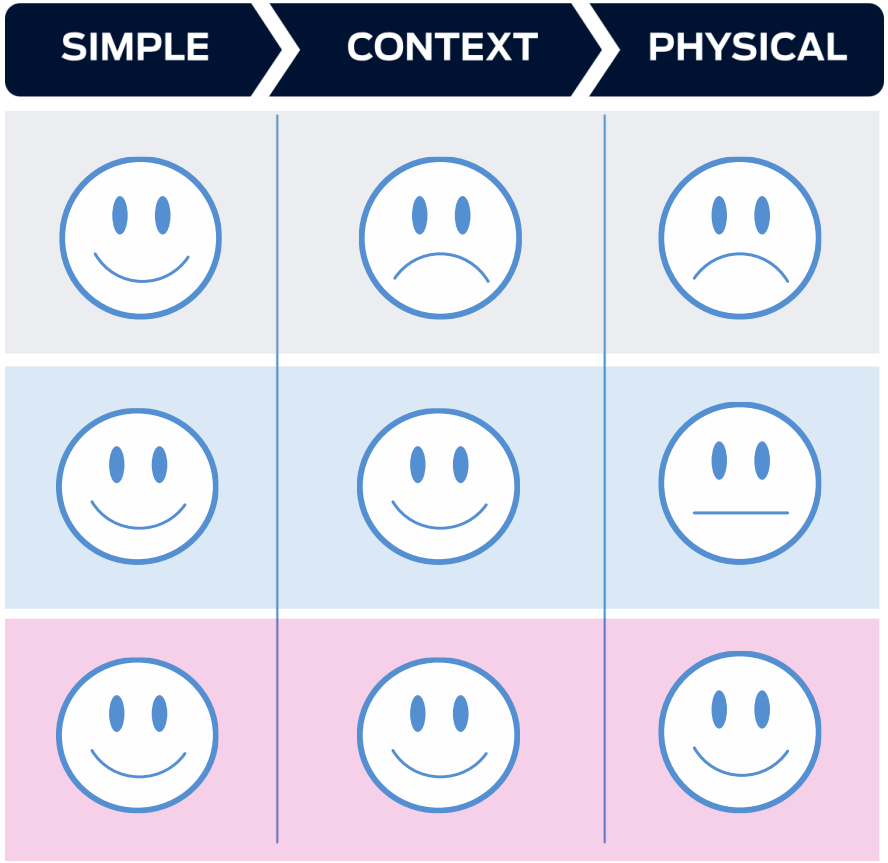
Testing Behavioral Biometrics - *On Its Own*

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



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

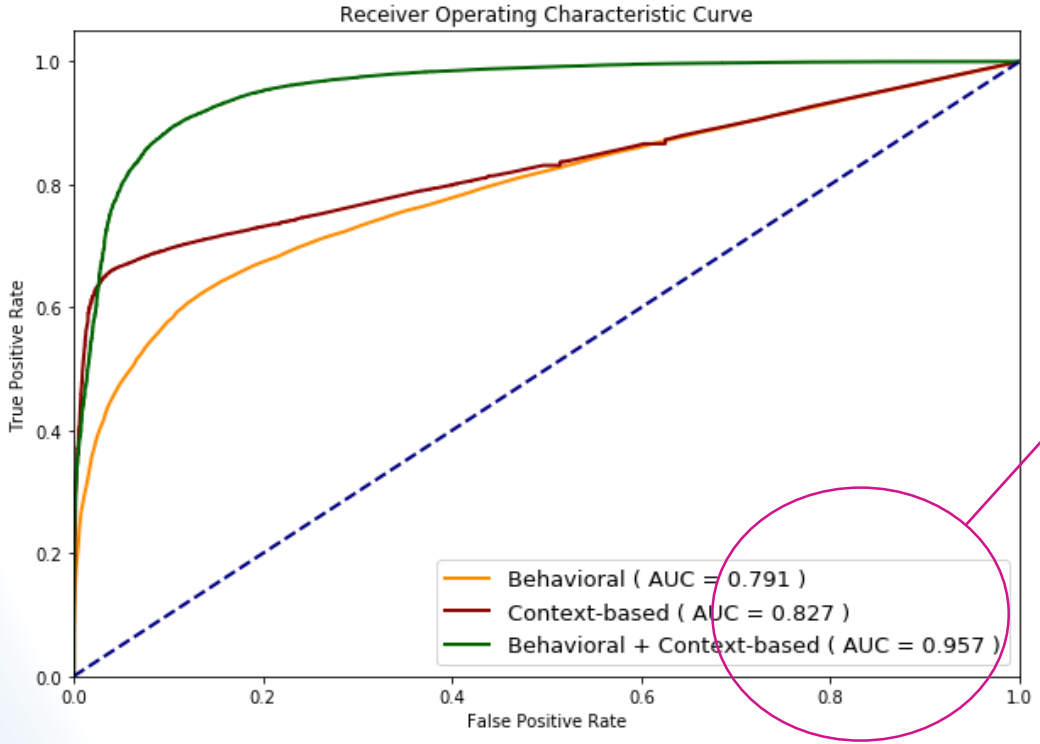
Context Risk

Behavioral Risk

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

$\alpha_c + \alpha_b = 1$ ← 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
Behavioral + Context-Based	**0.939**	**0.937**	**0.910**	**0.940**

- ✓ 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
Jesus.solano@Cyxtera.com

Cyxtera