June
2018

HEC PARIS

FinTech Solutions for Banking Operations and Services

MASTER’S THESIS

AUTHORS: LOUIS MARTY, DAMIEN MOSSUZ ET MATTEO SCRENCI
SUPERVISOR: OLIVIER BOSSARD

Behavioral Biometrics: a new era of security for online banking

Master Thesis 2018

Louis Marty, Damien Mossuz, Matteo Screnci

HEC PARIS

Thesis plan

Behavioral Biometrics as a response to increasing online banking fraud

I. **Overview of the secured authentication market**

II. **Behavioral biometrics: a frictionless additional security layer**

III. **One step ahead: evaluating user's reactions with Invisible Challenges**

IV. **Performance evaluation: calculating the ROI of Behavioral Biometrics**

V. **Behavioral Biometrics added value compared to traditional solutions**

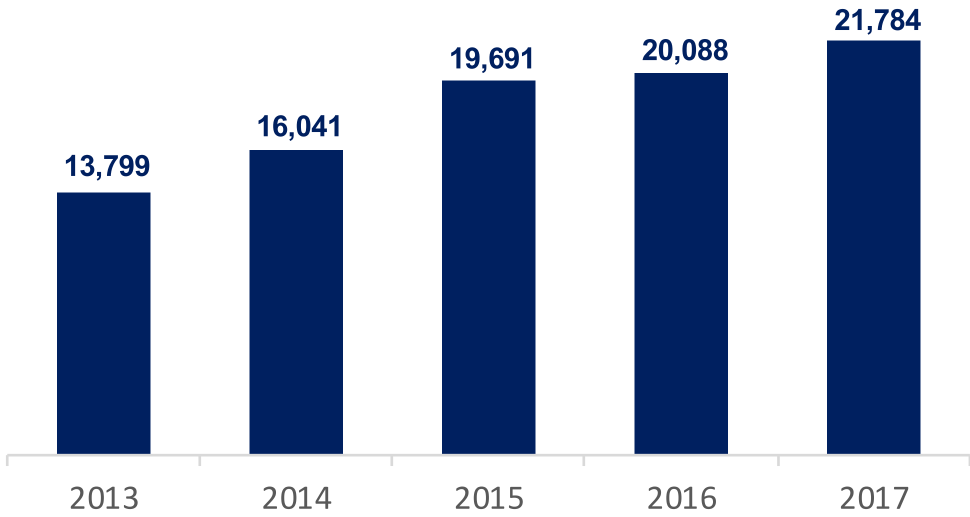
Overview of the secured authentication market

Global trends and forecasts in online banking

Key figures

$ 5.1bn	Losses from account takeover in 2017[19]
73%	73% of people not using mobile banking didn't for security concerns in 2014[14]
$ 50bn	Mobile biometrics market size by 2022[15]
29.3%	Average estimated growth rate of this market from 2016 to 2022[15]

Case volumes for online banking frauds in the UK[16]



Main online banking frauds[2]

Remote Access Trojans
~30% of all frauds

Application Fraud
Most of the increase is in credit card account opening

Social Engineering
Phishing, vishing

Mobile Fraud
Exponential growth with the introduction of full mobile banking functionality

Overview of the secured authentication market

Several reasons to go beyond Multi-Factor Authentication (MFA)

Overview of the key flaws of MFA[11]

Is MFA enough security?

Stolen valid credentials constitute 81% of all breaches

Endpoint security cannot stop attackers with valid credentials

OTP to SMS or Email

Subject to phone fraud – discouraged by NIST

Hard tokens

Compromised in the past, offer a poor experience, and costly to manage

Knowledge-based questions

Easily obtained from social media

Push-to-Accept

Can be manipulated into accepting non-valid requests

Overview of the secured authentication market

Comparison between physical and behavioral biometrics (1/2)

Physical VS. Behavioral authentication[17]

"Behavioral biometrics is the field of study of uniquely identifying and measurable patterns in human activity. The term contrasts with physical biometrics, which involve innate human characteristics such as fingerprints, iris patterns." [3]

Physical Biometrics	**Behavioral Biometrics**

Source: https://nudatasecurity.com/resources/blog/deciding-on-biometrics/

Overview of the secured authentication market

Comparison between physical and behavioral biometrics (2/2)

Physical VS. Behavioral authentication(17)

	Behavioral	Physical
Examples	*Keystrokes, Mouse movements*	*Fingerprint, Iris scan, Face identification*
Process	*Continuous tracking of user's actions: DYNAMIC*	*Physical measurement at an instant: STATIC*
Frictionless	✔	✘
Legal proof of identity	✘	✔
Stored data	*In a form which is not useful for hackers, and no personal identifying information*	*Physical biometrics information*
Exposure to hacking	*Limited*	*High*

Take-away points #1

Overview of the secured authentication market

There is still a lot to be done in security…[24]

- In 2016, in the UK, financial fraud losses totalled **£768.8 million, up 2% from 2015**
- **80%** of these losses came from **remote banking**
- Most part of the growth comes from ***'sophisticated online attacks such as malware and data breaches'***[24]

…and traditional security solutions seem outdated.

Two major limitations for MFA:

- **Stolen credentials** may be used to impersonate the real user
- **After the valid user is properly authenticated,** an unauthorized entity may initiate fraudulent transactions

Could Behavioral Biometrics be the answer?

Behavioral Biometrics are already considered as the **third most popular biometric technology** (after finger and face, tied with iris)[25]

Thesis plan

Behavioral Biometrics as a response to increasing online banking fraud

I. **Overview of the secured authentication market**

II. **Behavioral biometrics: a frictionless additional security layer**

III. **One step ahead: evaluating user's reactions with Invisible Challenges**

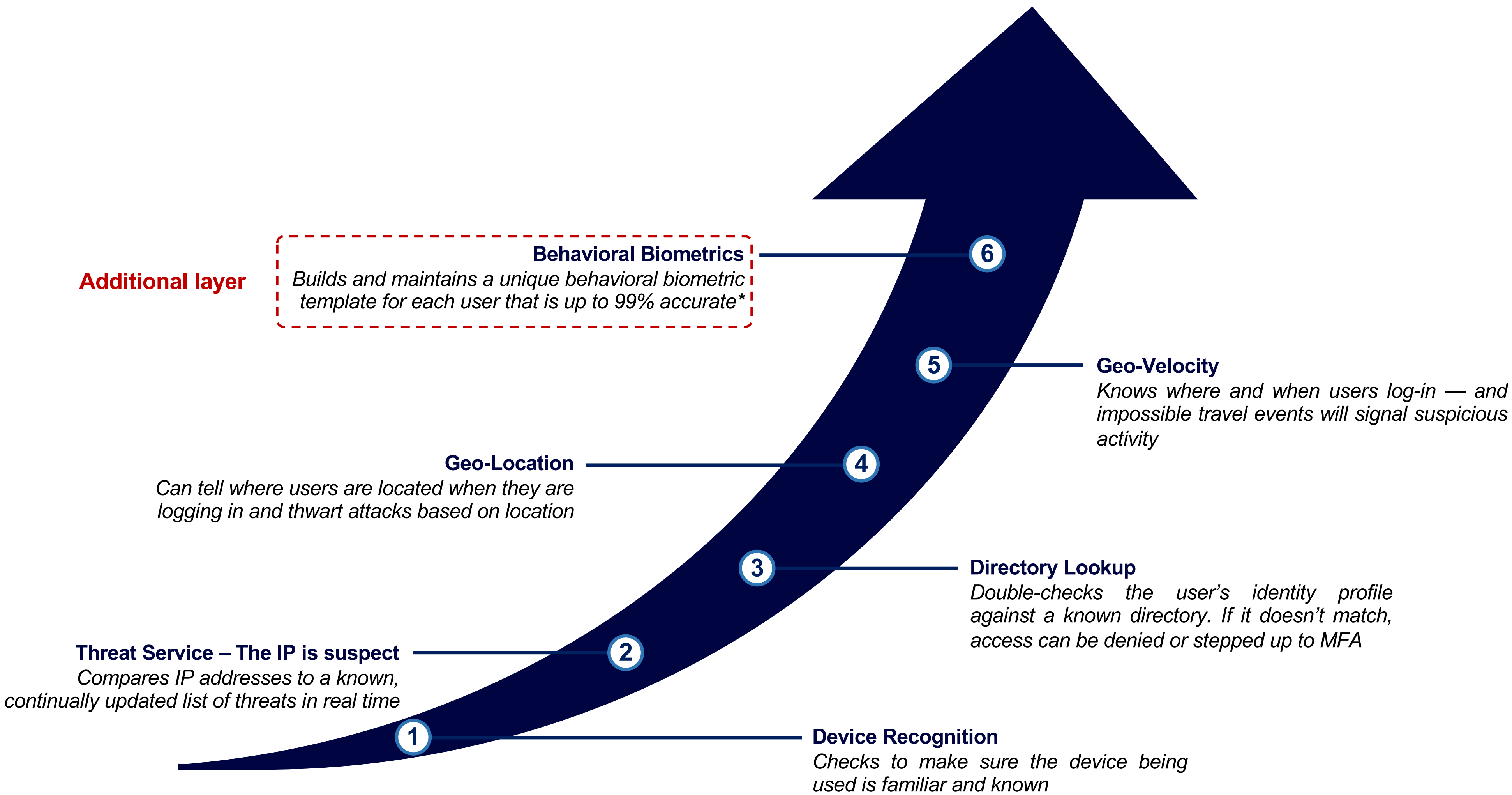
IV. **Performance evaluation: calculating the ROI of Behavioral Biometrics**

V. **Behavioral Biometrics added value compared to traditional solutions**

Behavioral biometrics: a frictionless additional security layer

How it adds up to existing security layers

Behavioral biometrics is a new security layer in the process of secured authentication[13]

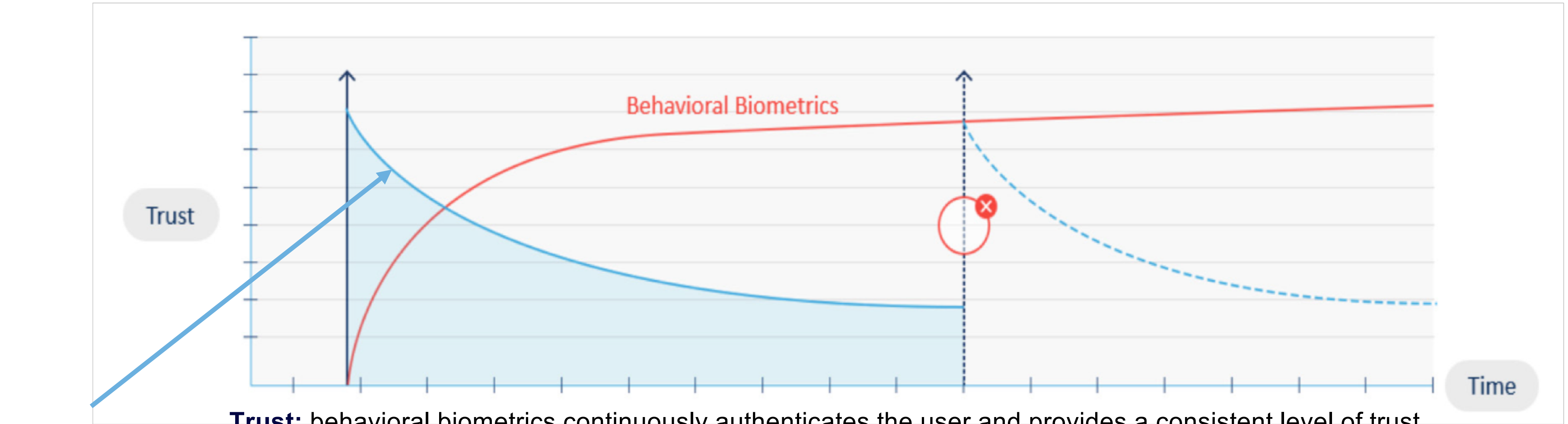


*please see 'Annex: Measuring the performance of a biometric system' to learn more about accuracy.

Behavioral biometrics: a frictionless additional security layer

Enhancing security without adding friction

As a continuous process, behavioral biometrics allow to add security without even noticing the user[(1)]



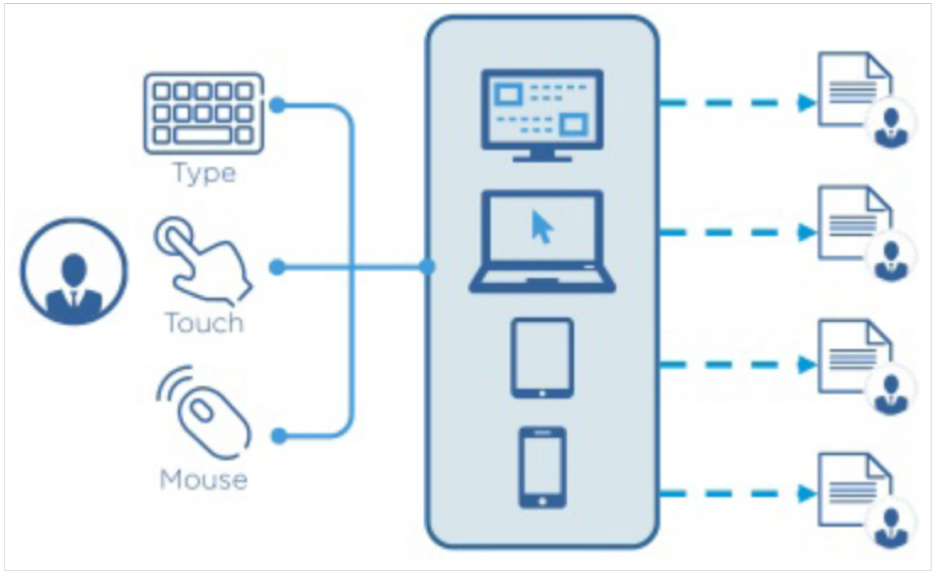
Trust: behavioral biometrics continuously authenticates the user and provides a consistent level of trust.

Friction: behavioral biometrics continuously monitors the session without disrupting user experience.

Behavioral biometrics: a frictionless additional security layer

How it works

The algorithm listens for events and builds a score[12]



- **Unique to each individual on each device**
- **Accuracy improves over time***
- **Only acts when threat detected – MFA to proceed**
- **Risk assessment for access control**

1. **Monitoring of user's behavior**
   Keystroke speed, mouse movements
2. **Analysis to deduce a score**
   If the score reveals a threat, credentials are asked
3. **Authentication**

please see 'Annex: Measuring the performance of a biometric system' to learn more about accuracy.

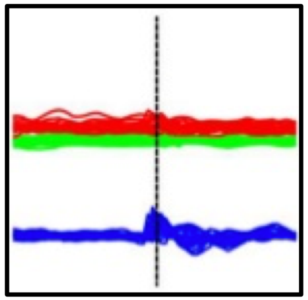
Behavioral biometrics: a frictionless additional security layer

Cognitive Biometric Measurements

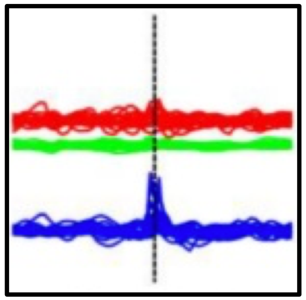
Three types of factors are used to detect a human or non-human imposter[3]

Cognitive factors	Physiological factors	Contextual factors
Eye-hand coordination	Left/right handedness	Transaction
Applicative behavior patterns	Press-size	Navigation
Usage preferences	Hand tremors	Network patterns
Device interaction patterns	Arm size	Geo-Location

How does the user hold the device? What happens when they tap it?[4]



Touch down
User 1



Touch down
User 2

Red/Green: x-y movement of device
Blue: vertical movement (up/down)

User 1: no up/down movement

≠

User 2: visible up/down movement (blue spike)

Source of images[4]

Thesis plan

Behavioral Biometrics as a response to increasing online banking fraud

I. **Overview of the secured authentication market**

II. **Behavioral biometrics: a frictionless additional security layer**

III. **Beyond traditional Behavioral Biometrics: Invisible Challenges[3]**

IV. **Performance evaluation: calculating the ROI of Behavioral Biometrics**

V. **Behavioral Biometrics added value compared to traditional solutions**

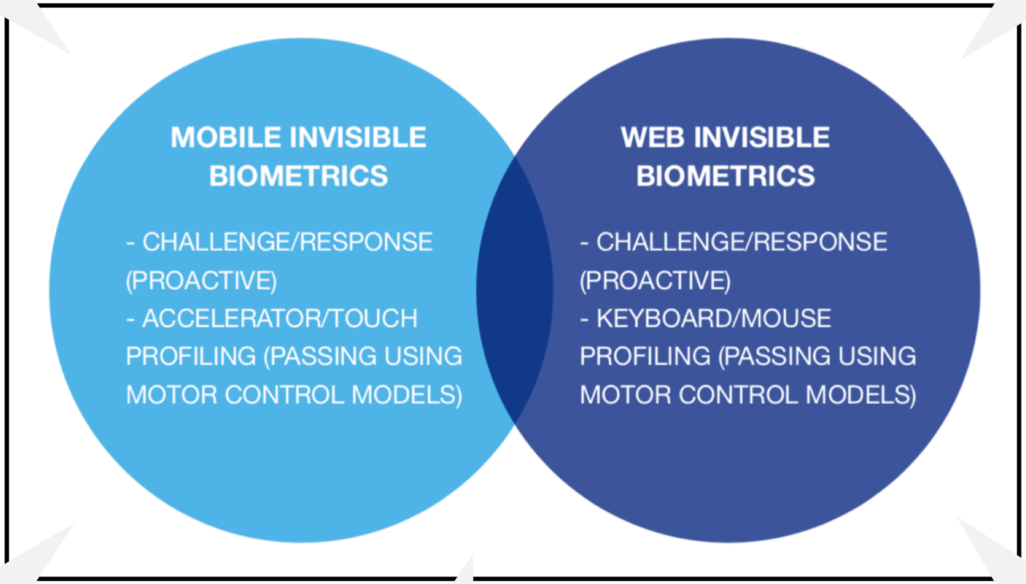
Invisible Challenges

Introduction

Invisible Challenges are at the heart of what makes possible to establish a very accurate profile for each user

Injected at **specific points** within a session

Do not alter user experience



Proactive and **Passive**

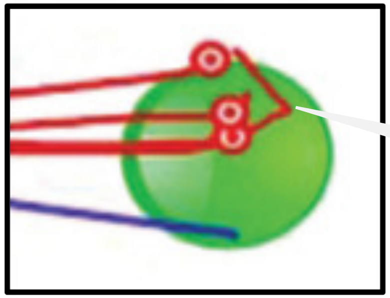
Elicit **unique behavioral and cognitive parameters**

Change each time in a **randomized** way

Invisible Challenges

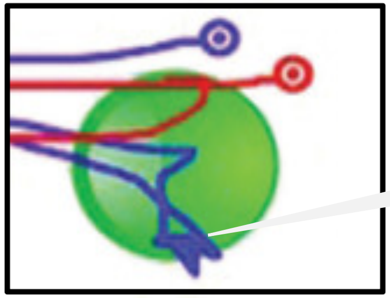
Mobile Behavioral Biometrics: Rotation of Movement

Deviation introduced during a *drag-and-drop*



User 1

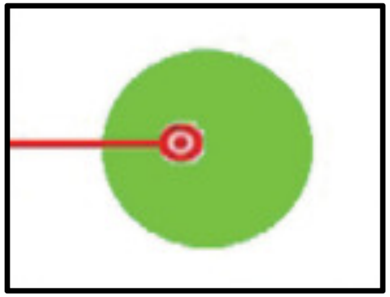
User 1: one small correction during the last 10% of the movement (**red hook**)



User 2

User 2: multiple corrections during the last 20% of the movement (**blue lines**)

Both users reported they did not sense any change in their experience



Robot

Robot: no compensation because no hand-eye coordination

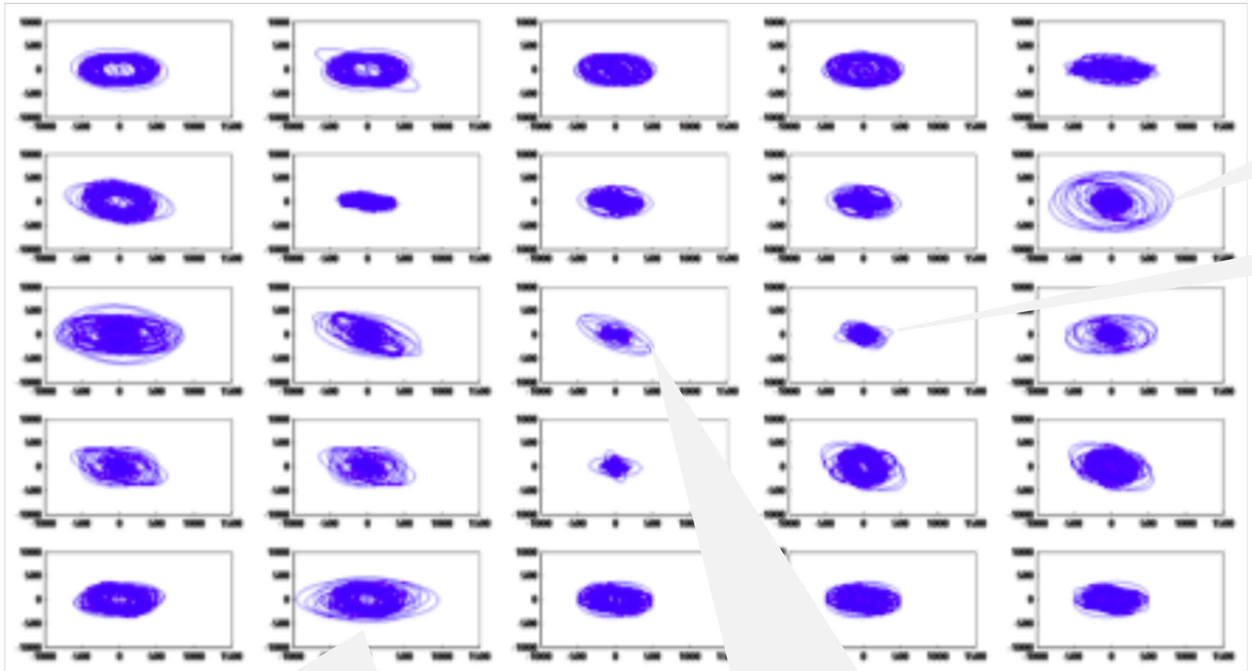
Source of images[(3)]

Invisible Challenges

Computer Behavioral Biometrics: Disappearing Mouse

The cursor is hidden after the user completes a task until they start searching for the mouse

25 users*, each with a slightly different search pattern for a missing cursor.*



Some users have **wide search patterns**…

…others use **small ones**.

Certain users always search **counter-clockwise**.

Some are **horizontal**…

…while others are **diagonal**.

Usually not practical because it takes too much time to capture enough data. This **Invisible Challenge heavily shortens this time** by "forcing" the user to make various mouse movements.

Source of image(3)

Invisible Challenges

Conclusion: advantages of Invisible Challenges

Comparison between Invisible Challenges and Passive Behavioral Biometrics[4]

Identifying real fraud while maintaining low false alarm rates and low user friction: catch-22 for behavioral Biometrics.

Invisible Challenges optimize this balance: a single challenge can lower the EER of any by 3%, adding more challenges drive performance exponentially.*

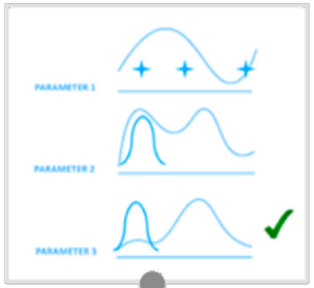
	Behavioral Biometrics with Invisible Challenges	Passive Behavioral Biometrics
Behavioral Parameters	✓	✓
Cognitive Parameters	✓	✗
Device Dependency	*Low*	*High*
Time for building profile	*Short*	*Long*
Excel at	*Free Form Usage*	*Repeat Tasks (e.g. password & PIN typing)*

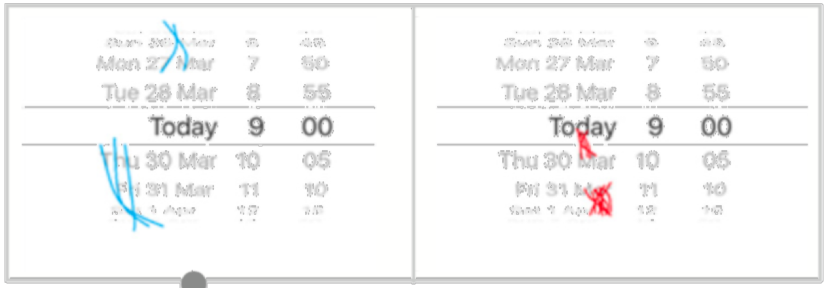
Equal Error Rate: please see *'Annex: Measuring the performance of a biometric system'* to learn more.

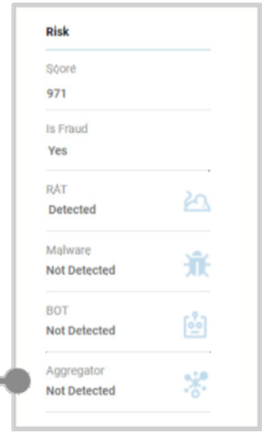
**figures are based on real data coming from the 2 millions transactions BioCatch monitor a month.*

Take-away points #2

How Behavioral Biometrics manage to authenticate users with accuracy



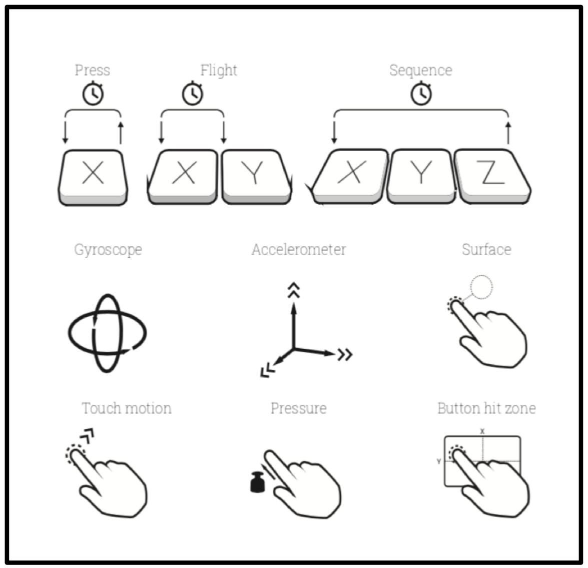




USER PROFILING

INVISIBLE CHALLENGES

RISK SCORE



Source[(10)]

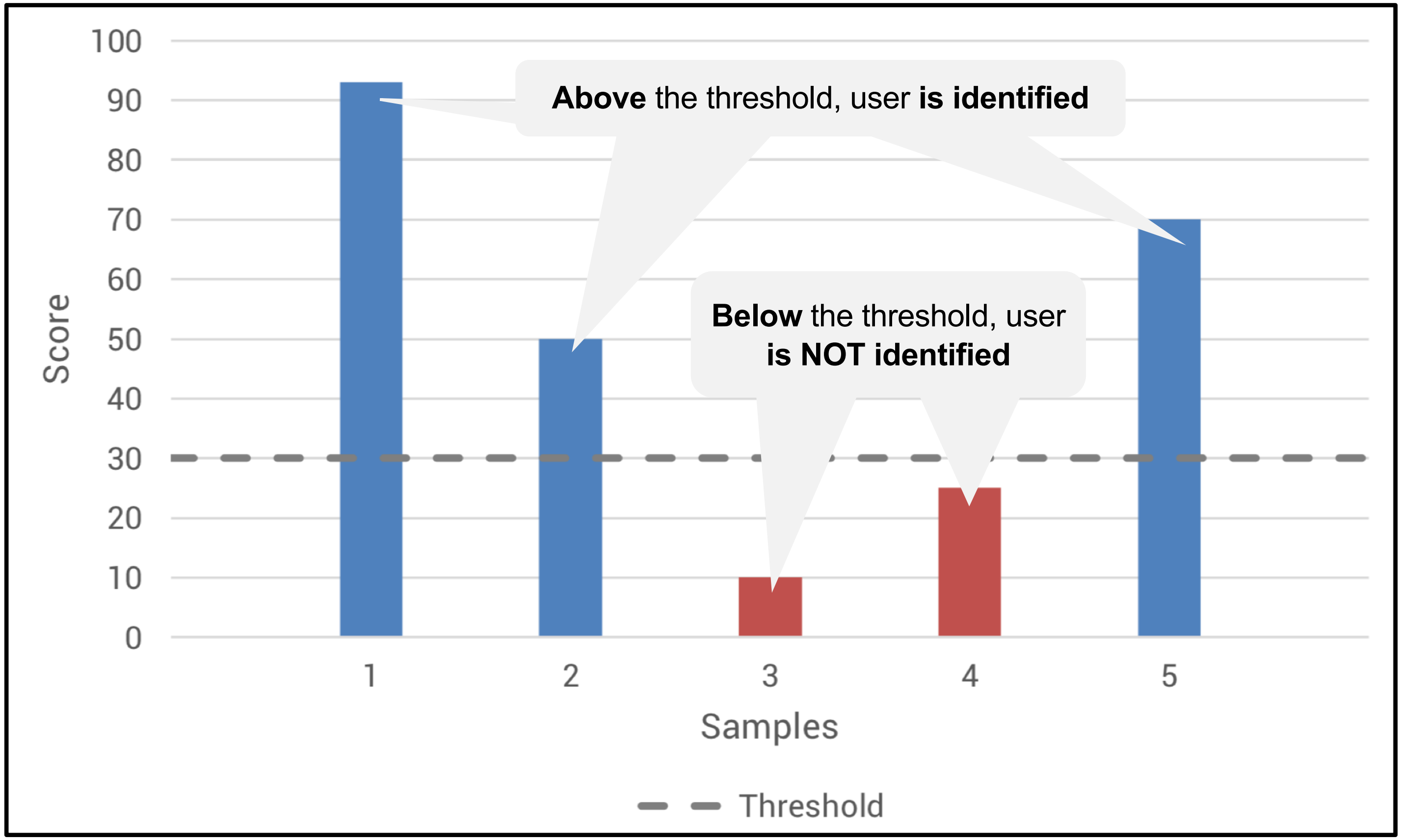
- **Patented technology**
- **Random times**
- **Without user's knowledge**
- **Impossible to mimic**

- **Search for fraudulent activity**
- **Score compared to pre-determined threshold**
- **Real-time alerts**

Annex: Measuring the accuracy of a biometric system[8]

Using the EER to measure the accuracy of a biometric system (1/4)

Some definitions: Threshold, FAR, FRR



Example of threshold-based identification.

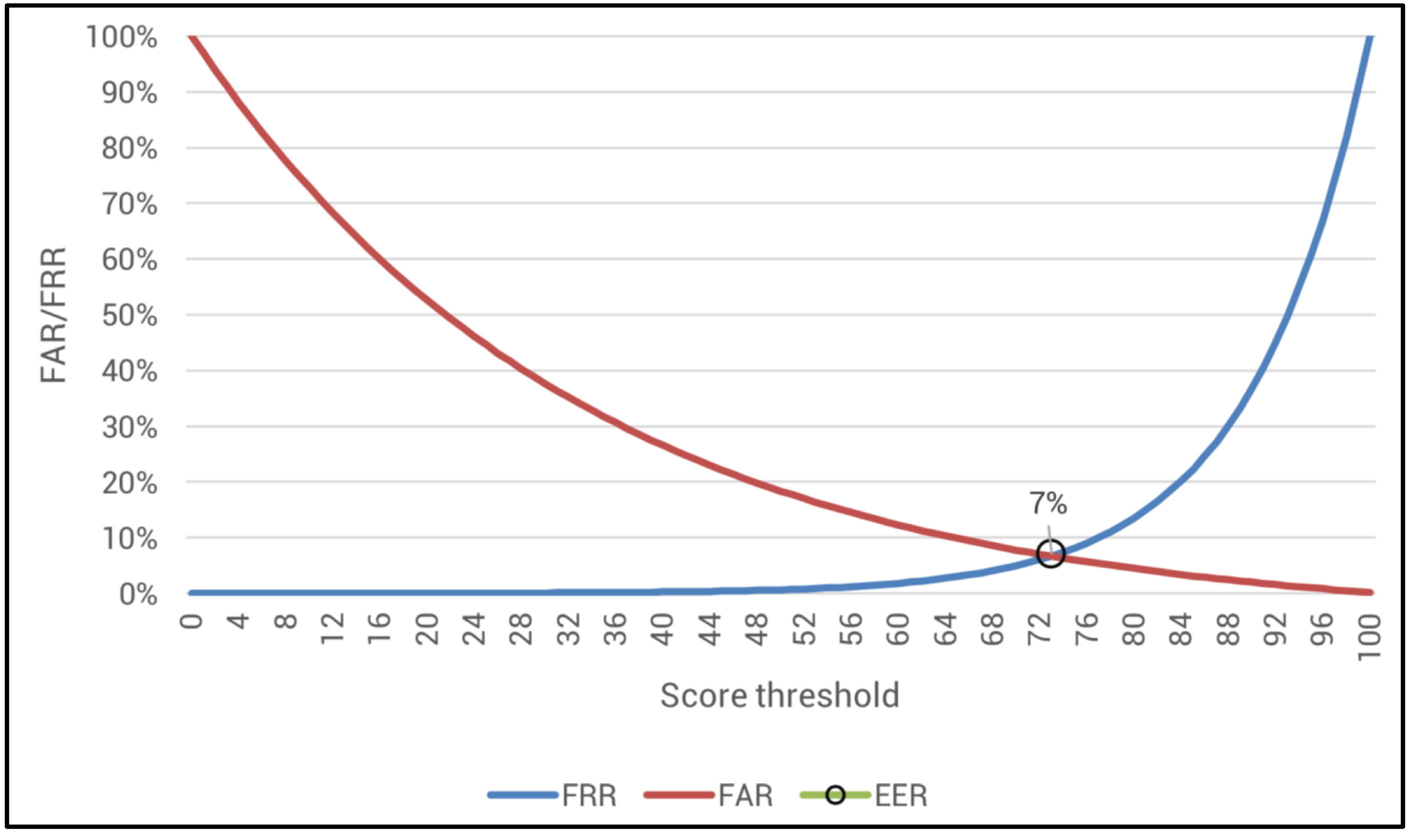
False Acceptance Rate (FAR): percentage of samples that **incorrectly score above the threshold**. A higher threshold reduces the FAR. The FAR is directly connected to the security of a system.
False Rejection Rate (FRR): percentage of samples that **incorrectly score below the threshold**. A lower threshold reduces the FRR. Just a comfort criterion: higher FRR does not make the system less secure.

Source of diagram[8]

Annex: Measuring the accuracy of a biometric system[8]

Using the EER to measure the accuracy of a biometric system (2/4)

Understanding the EER



Example of FAR/FRR curves highlighting the point of the EER.

Equal Error Rate (EER): point at which FAR = FRR. The EER is a quick and well- established threshold-independent method for comparing the accuracy between different systems. The lower the EER, the more accurate the system.

Source of graph[8]

Annex: Measuring the accuracy of a biometric system(8)

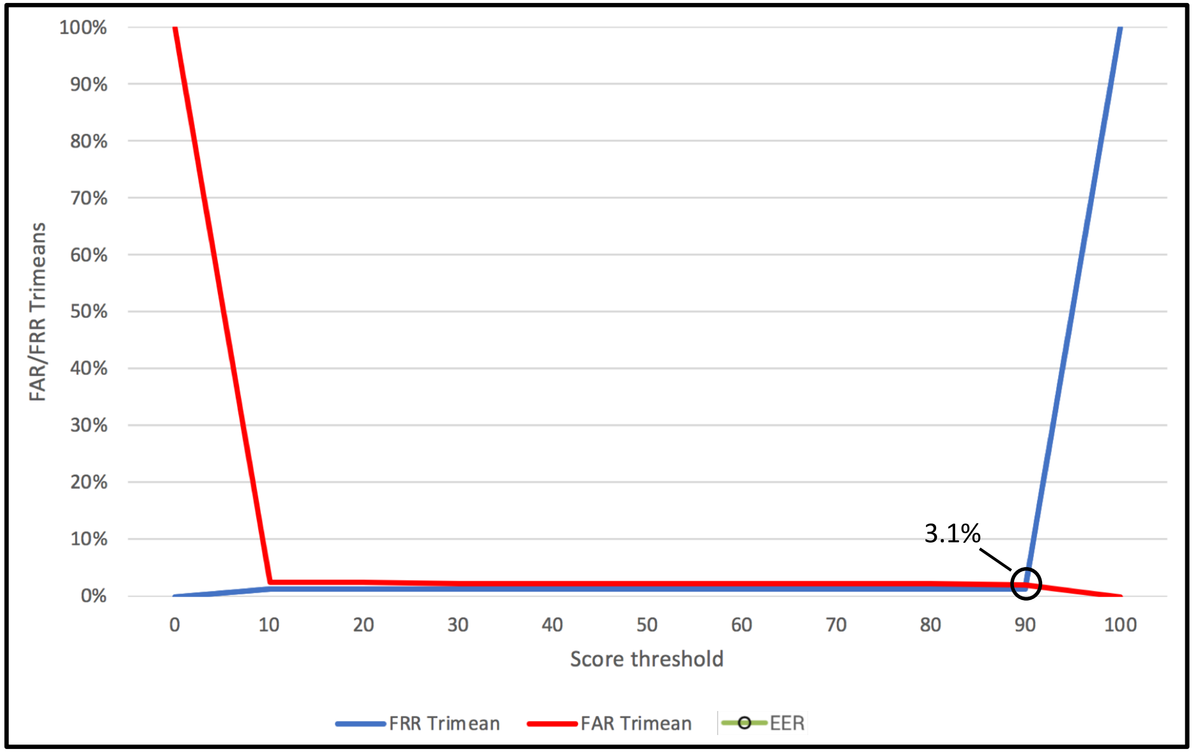
Using the EER to measure the accuracy of a biometric system (3/4)

Results from the analysis of one attempt to login

Threshold	FAR Trimean	FRR Trimean
0	100	0
10	2.50	1.19
20	2.48	1.19
30	2.30	1.19
40	2.30	1.19
50	2.30	1.19
60	2.19	1.19
70	2.19	1.19
80	2.19	1.19
90	1.93	1.19
100	0	100

FAR & FRR trimeans for one attempt to login

Trimean: weighted average of the distributions median and its two quartiles. It is a robust estimator of a population mean.



EER for a full session is 3.1% at a threshold of 91

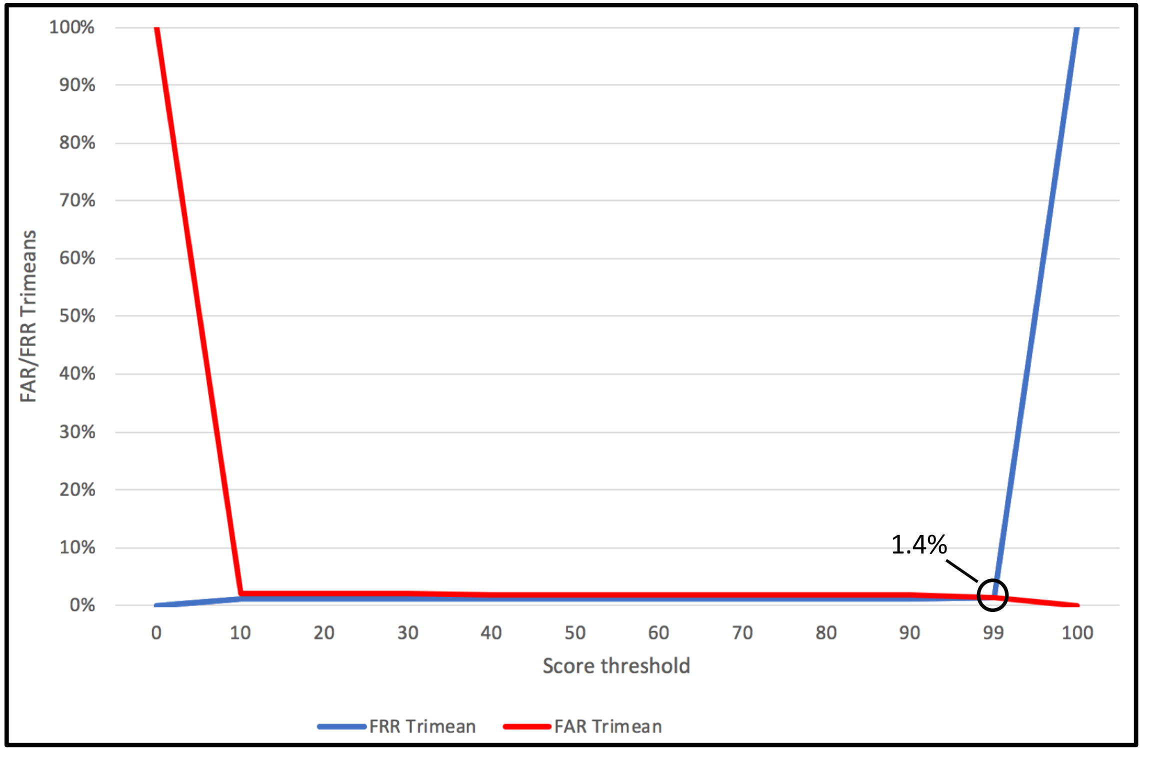
Annex: Measuring the accuracy of a biometric system(8)

Using the EER to measure the accuracy of a biometric system (4/4)

Results from the analysis of a full payment transaction

Threshold	FAR Trimean	FRR Trimean
0	100	0
10	2.10	1.19
20	2.04	1.19
30	1.97	1.19
40	1.85	1.19
50	1.85	1.19
60	1.85	1.19
70	1.79	1.19
80	1.79	1.19
90	1.79	1.19
100	0	100

FAR & FRR trimeans for a full payment transaction



EER for a full session is 1.4% at a threshold of 99

Thesis plan

Behavioral Biometrics as a response to increasing online banking fraud

I. **Overview of the secured authentication market**

II. **Behavioral biometrics: a frictionless additional security layer**

III. **Beyond traditional Behavioral Biometrics: Invisible Challenges**

IV. **Performance evaluation: calculating the ROI of Behavioral Biometrics[5]**

V. **Behavioral Biometrics added value compared to traditional solutions**

Calculating the Return On Investment

Part 1 – Comparing Periodic to Continuous Authentication

The quality of each fraud detection consists of three dimensions

- **How effective is the solution in detecting as many actual fraud cases as possible**
- **How effective is the solution in avoiding false alarms**
- **How operationally-efficient is the solution (no additional intrusive operational steps, no barriers to deployment, etc.)**

Comparing Periodic to Continuous Authentication

Quantification of the three dimensions in dollar terms for both:

Periodic User-Authentication

PINs, passwords or MFA

Continuous User-Authentication

Behavioral Biometrics

Double positive impact**: detects a higher number of fraudulent attempts** and **rejects or escalates a smaller number of valid transactions.**

Data collection is performed **transparently and passively**.

Can provide complementary and incremental value when **used in conjunction with other solutions.**

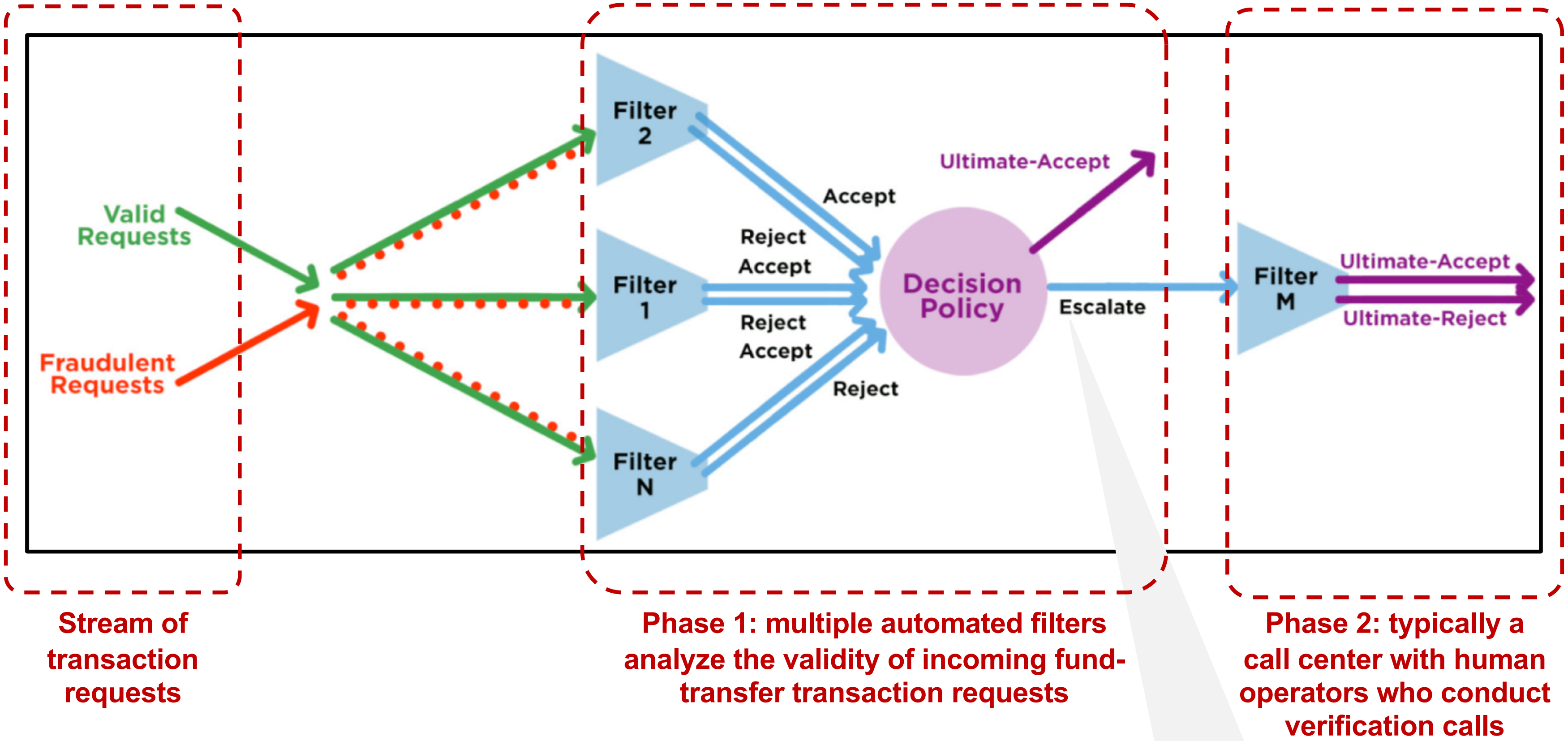
Comparison of the dollar-term quantifications

Calculation of the ROI

Calculating the Return On Investment

Part 2 – Defining a Filter Model for Fraud Detection Solutions

Model of a typical hybrid online transactional environment



- Goal of each filter: conjointly **maximize Fraud Rejection** percentage (FR) and **minimize Valid Rejection** percentage (VR)
- Other possible outputs such as a **score indicating a decision confidence interval**
- To calculate **performance of a precise filter in a multi-filter scheme**, the model considers the fraudulent transaction rejected by the filter but not by the system without this filter

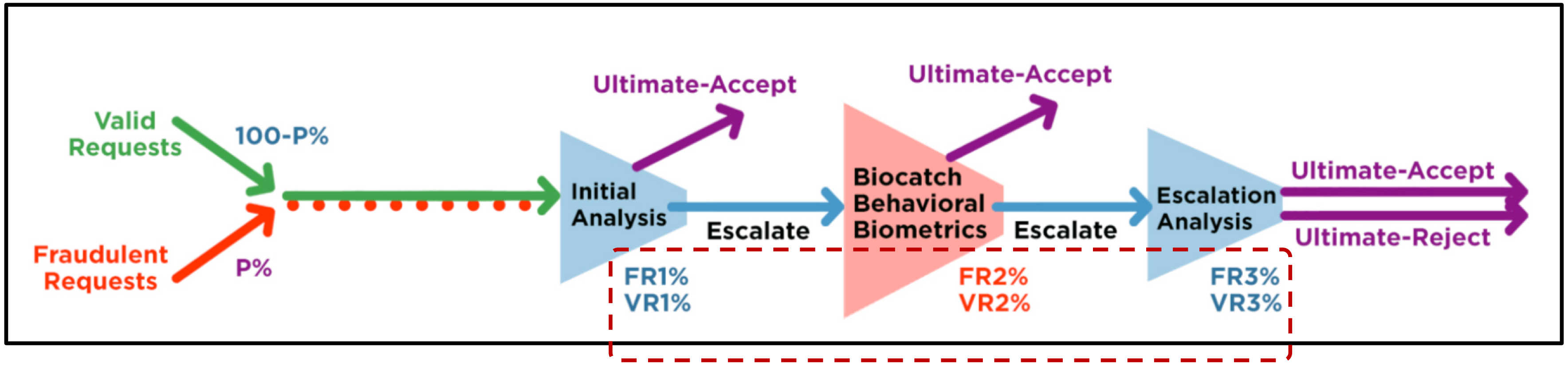
Suspected-fraudulent transactions are escalated serially to a consequent filter

Source of diagram[5]

Calculating the Return On Investment

Part 3 – Business Value Quantification (1/3)

Environment of the study: three serially operating filters



For each of the three filters, the pair (FR; VR) must be assessed

Example of additional parameters that can be incorporated

Transactions	
Number of transactions per month	2,000,000
Average $ value of a transaction	$150
Total lifetime value of an average customer	$100,000
Estimated likelihood of a customer switching provider if rejected	12%

Frauds	
Average % of transactions that are fraudulent	1%
Average recovery costs for a fraudulent transaction	$3,000
Estimated likelihood of a customer switching provider if victim of fraud	8%

Filters	
Initial pre-BehavioralFilter estimated % of escalated fraudulent transactions	55%
Initial pre-BehavioralFilter estimated % of escalated valid transactions	0.7%
Post-BehavioralFilter (call center) - Estimated % of rejected fraudulent transactions	85%
Post-BehavioralFilter (call center) – Average operational costs for a transaction	$200

Source of diagram[(5)]

Calculating the Return On Investment

Part 3 – Business Value Quantification (2/3)

Calculating the monetary value or cost of each ultimate decision

The monetary value or cost of each decision is calculated according the following formulas:

- Accepted **Valid Transaction**: *Average revenue of a transaction*
- Accepted **Fraudulent Transaction:** *Average chargeback of a transaction* **+** *Avg additional fraud recovery costs* **+** *Avg lifetime value of a customer * likelihood to switch provider if victim of fraud*
- Rejected **Valid Transaction:** *Average lost revenue of a transaction* **+** *Avg lifetime value of a customer * likelihood to switch provider if rejected*
- Escalated **Transaction:** *Average operational cot of performiong the escalated decision process* **+** *Avg lifetime value of a customer * likelihood to switch provider if escalated*

	Intermediate monetary values
Average revenue from each Accepted Valid Transaction	$200
Average cost of each Rejected Valid Transaction	$680
Average cost of each Accepted Fraudulent Transaction	$4,400
Average escalation cost for each Escalated Transaction	$105

Example of the derived intermediary monetary values

The model then calculates the total number of transactions of each of the following types

- ***Valid*** *Transactions ultimately* ***Accepted***
- ***Valid*** *Transactions ultimately* ***Rejected***
- ***Escalated*** *Transactions*
- ***Fraudulent*** *Transactions ultimately* ***Accepted***
- ***Fraudulent*** *Transactions ultimately* ***Rejected***

Calculating the Return On Investment

Part 3 – Business Value Quantification (3/3)

Calculating the Value of the Behavioral Biometrics filter

- Total number of **Valid** Transactions ultimately **Accepted** * avg value of each
- Total number of **Valid** Transactions ultimately **Rejected** * avg cost of each
- Total number of **Fraudulent** Transactions ultimately **Accepted** * avg cost of each
- Total number of **Escalated** Transactions * avg cost of each

	Scenario 1 Escalate transactions based on current method (pre-BehavioralFilter)	Scenario 2 Escalate transactions based on BehavioralFilter *Threshold option 1*	Scenario 3 Escalate transactions based on BehavioralFilter *Threshold option 2*
Avg monthly total revenue from transactions	**$118,775,052**	**$118,771,488**	**$118,760,796**
Avg monthly total cost from transactions frauds, rejections and escalations	**$18,504,433**	**$9,141,701**	**$5,539,264**
Avg monthly total net profits from transactions	**$100,270,619**	**$109,629,787**	**$113,221,532**
Monthly BehavioralFilter Value		**$9,359,168**	**$12,950,914**

Example of ROI calculation for a Behavioral Biometrics filter

Thesis plan

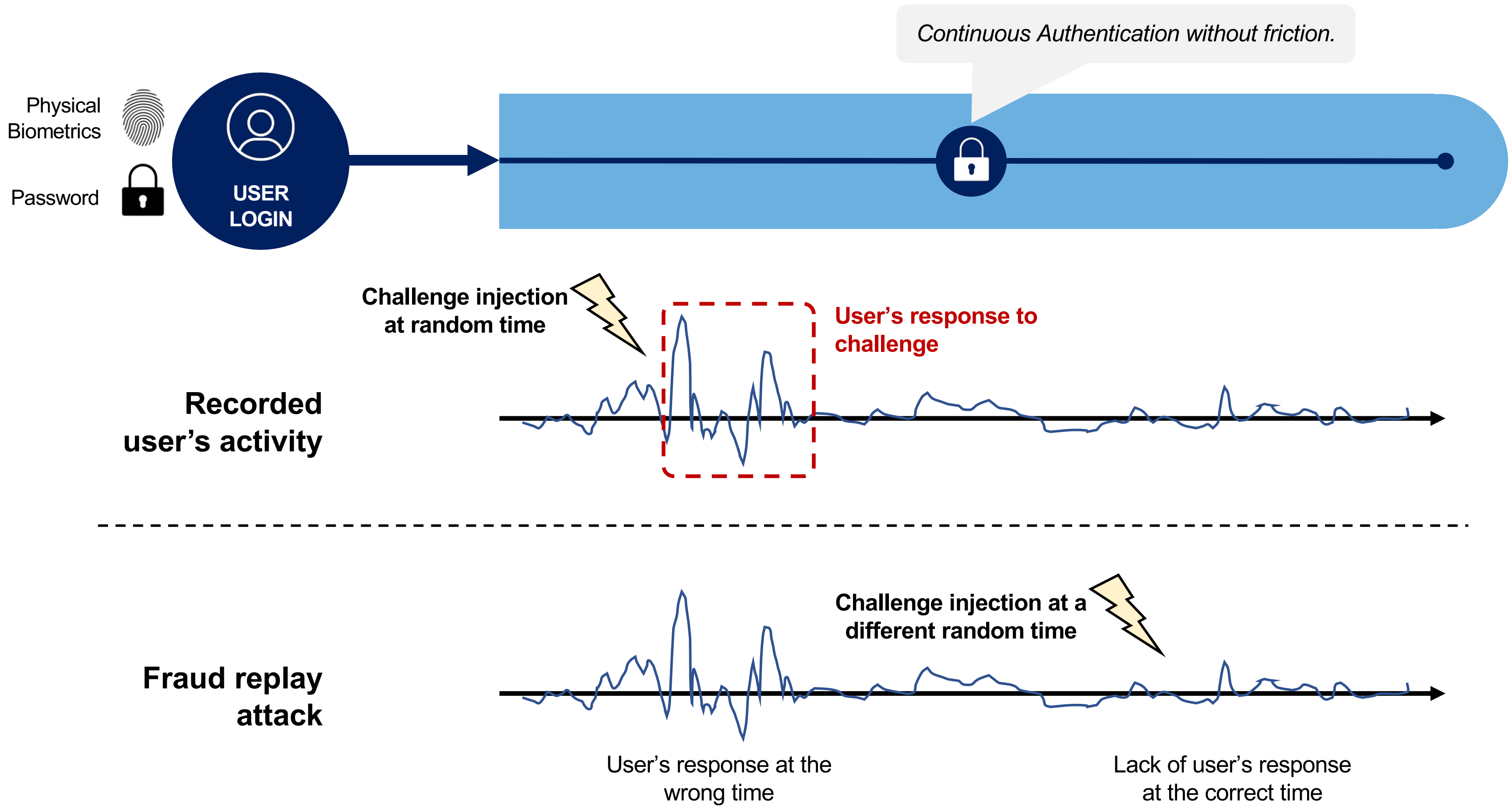
Behavioral Biometrics as a response to increasing online banking fraud

- **I.** **Overview of the secured authentication market**
- **II.** **Behavioral biometrics: a frictionless additional security layer**
- **III.** **Beyond traditional Behavioral Biometrics: Invisible Challenges**
- **IV.** **Performance evaluation: calculating the ROI of Behavioral Biometrics**
- **V.** **Behavioral Biometrics added value compared to traditional solutions**

Behavioral Biometrics added value

1 – Preventing Replay Attacks[(1)(4)]

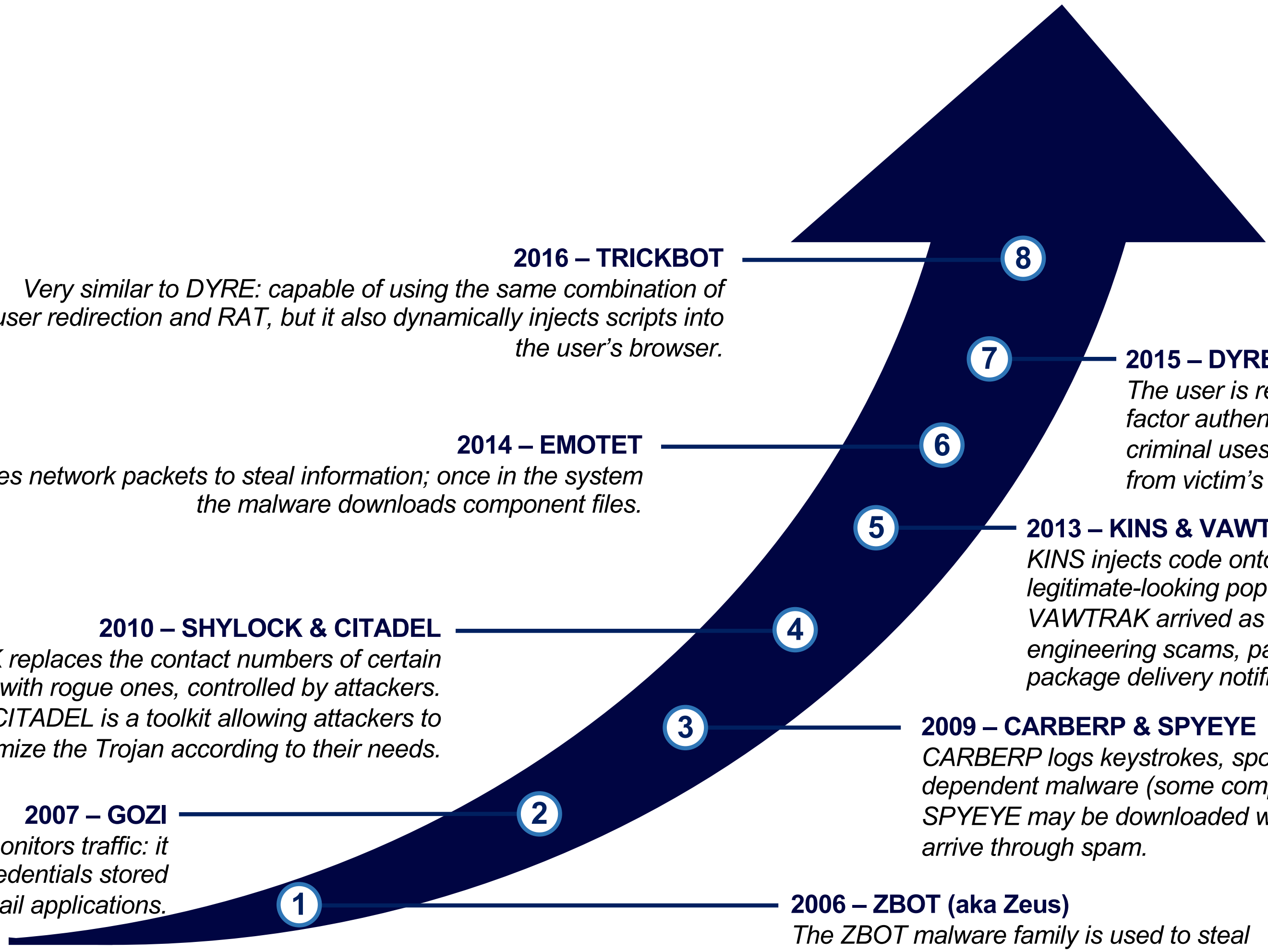
Temporal randomness allows detecting replay attacks



Behavioral Biometrics added value

2 – Detecting Remote Access Trojans (RATs) (1/2)

An history of RATs[(6)(18)]



1. **2006 – ZBOT (aka Zeus)**
The ZBOT malware family is used to steal data or account details.

2. **2007 – GOZI**
Spyware that monitors traffic: it can obtain login credentials stored in browsers and mail applications.

3. **2009 – CARBERP & SPYEYE**
CARBERP logs keystrokes, spoofs websites and is a plugin-dependent malware (some components need to be downloaded). SPYEYE may be downloaded when visiting malicious sites, or arrive through spam.

4. **2010 – SHYLOCK & CITADEL**
SHYLOCK replaces the contact numbers of certain banks with rogue ones, controlled by attackers. CITADEL is a toolkit allowing attackers to customize the Trojan according to their needs.

5. **2013 – KINS & VAWTRAK**
KINS injects code onto the user's browser, then shows legitimate-looking popups that ask for banking credentials. VAWTRAK arrived as a ZIP file attachment in social engineering scams, particularly spam emails disguised as package delivery notifications.

6. **2014 – EMOTET**
Analyses network packets to steal information; once in the system the malware downloads component files.

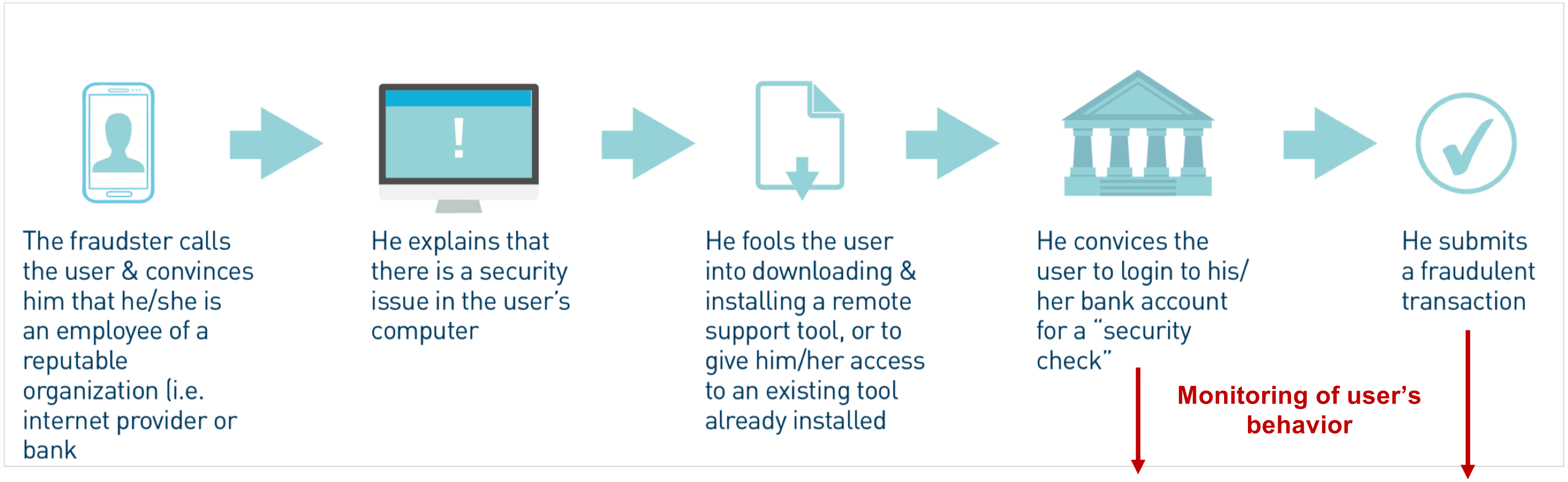
7. **2015 – DYRE**
The user is redirected to a fake site to steal their 2-factor authentication codes, and in parallel the criminal uses RAT to conduct fraudulent activities from victim's device.

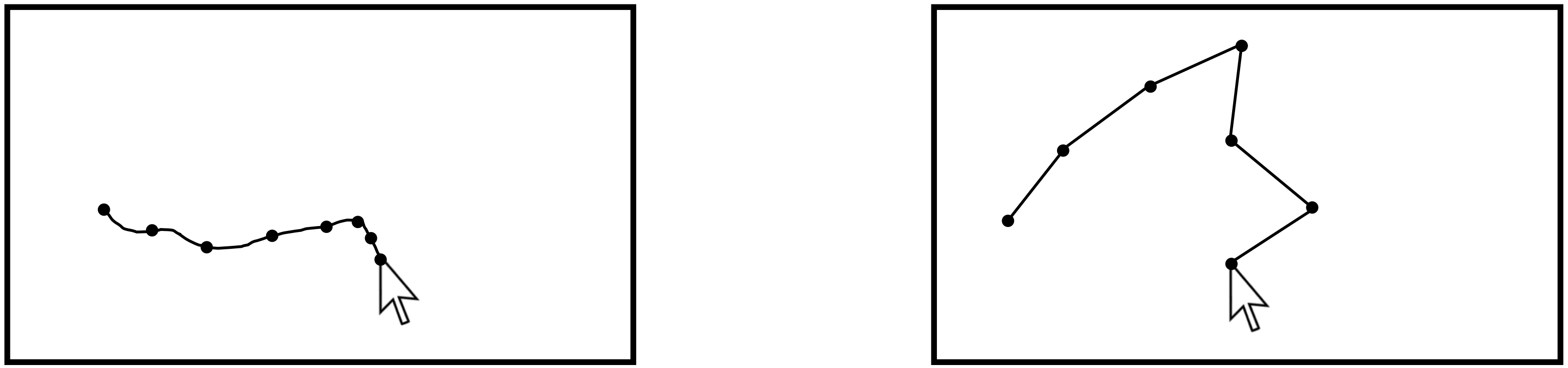
8. **2016 – TRICKBOT**
Very similar to DYRE: capable of using the same combination of user redirection and RAT, but it also dynamically injects scripts into the user's browser.

Behavioral Biometrics added value

2 – Detecting Remote Access Trojans (RATs) (2/2)

How RAT in the Browser works(6)





Regular user: no interruption of hand-eye coordination

RAT session: jumpy inconsistent motion of the mouse

Source of diagram(6)

Behavioral Biometrics added value

3 – Preventing New Account Fraud(7) (1/2)

Introduction – What is New Account Fraud?

Key figures(19)
- 1.5+ million victims
- $2.8 billion losses
- 40% increased in 2016
- 20% of all fraud losses

What data is needed to open a new account?
- Security social number
- Emails, Addresses, Phone number
- Device / network attributes

What's next?
- Phishing, Social Engineering
- Fake account is created with stolen data

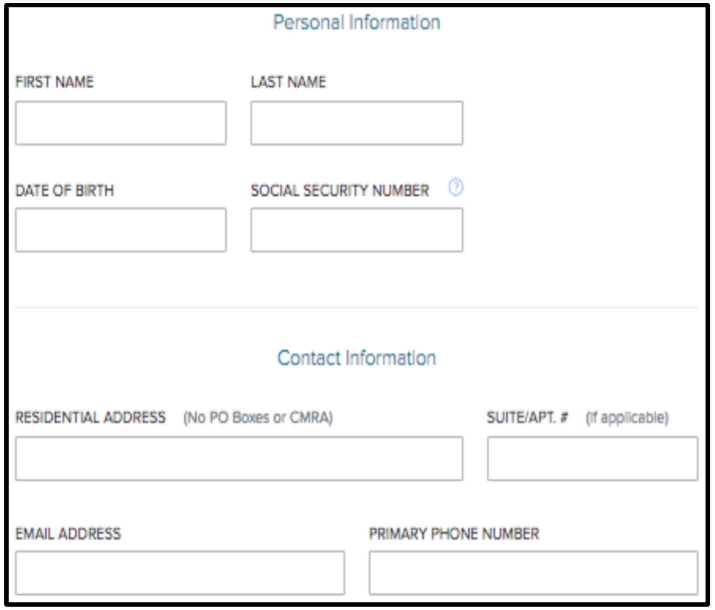
New Account Fraud
- When the fraud occurs within 90 days

Behavioral Biometrics added value

3 – Preventing New Account Fraud[7] (2/2)

3 main areas where behavioral analysis can be used to detect new account fraud

Application Fluency	Expert Users	Low Data Familiarity
Fraud can be detected in real-time by examining the **time it takes to fill out an online application**: **a new user will require more time** whereas a fraudster who repeatedly attacks a site is much faster.	**Fraudsters often use advanced computer skills** that are rarely seen among real users: **keyboard shortcuts** and function keys ('Shift+Tab' shortcut is used by only 13% of the general population).	Fraud can also be detected in real-time by examining **data familiarity: genuine users are usually very quick with fields that include personal data** (e.g. address), and fraudsters can make mistakes on details that should be intuitive.


Personal Information

FIRST NAME

LAST NAME

DATE OF BIRTH

SOCIAL SECURITY NUMBER

Contact Information

RESIDENTIAL ADDRESS (No PO Boxes or CMRA)

SUITE/APT. # (if applicable)

EMAIL ADDRESS

PRIMARY PHONE NUMBER

Example of fields for online application

### ##	[Long Pause]	####
Type First 5 SSN digits		**Type** Last 4 SSN Digits

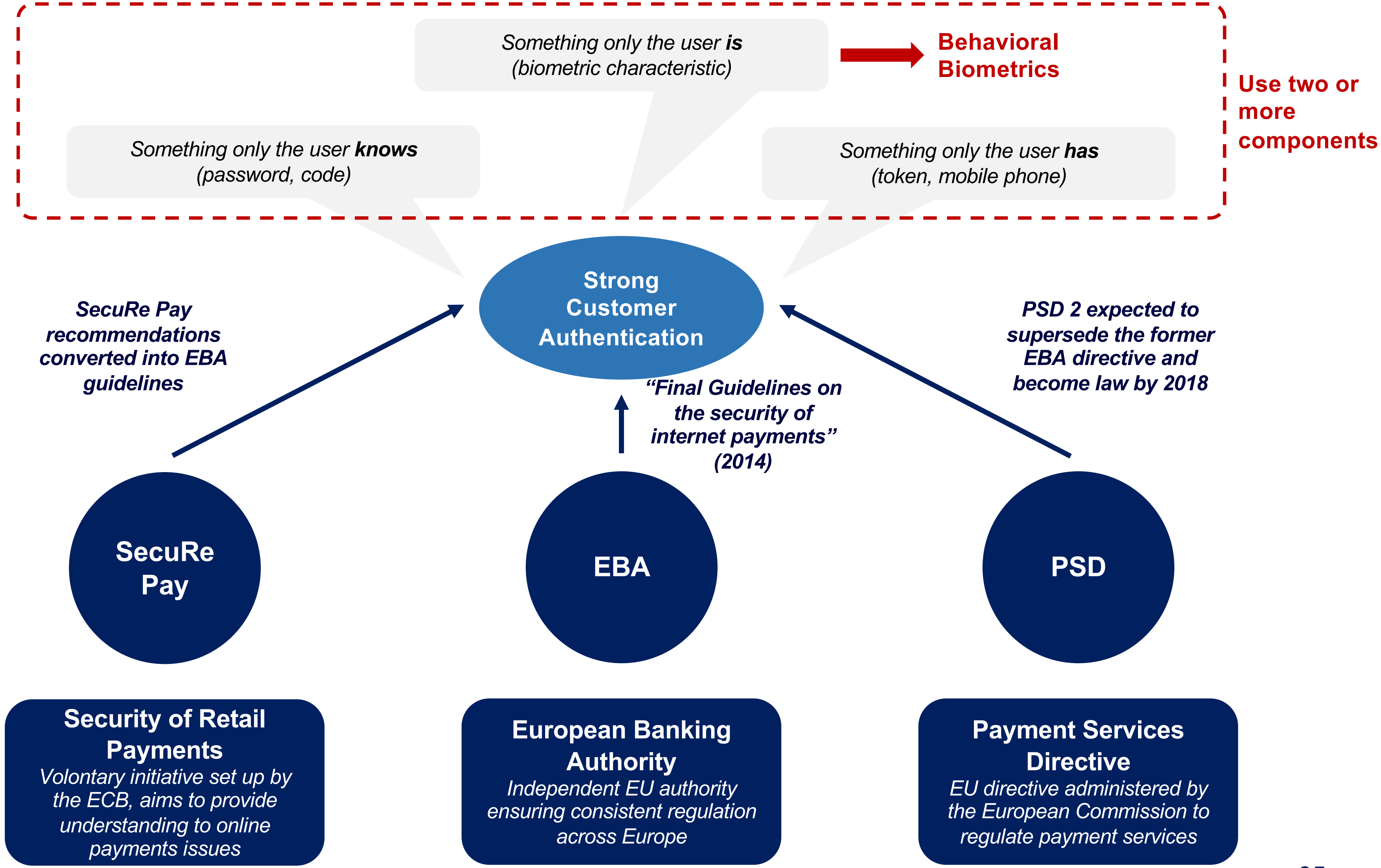
Uncommon behavior when entering personal data

Source of images[7]

Behavioral Biometrics added value

4 – Behavioral Biometrics as a tool to comply with regulations[9]

The regulatory landscape in the EU

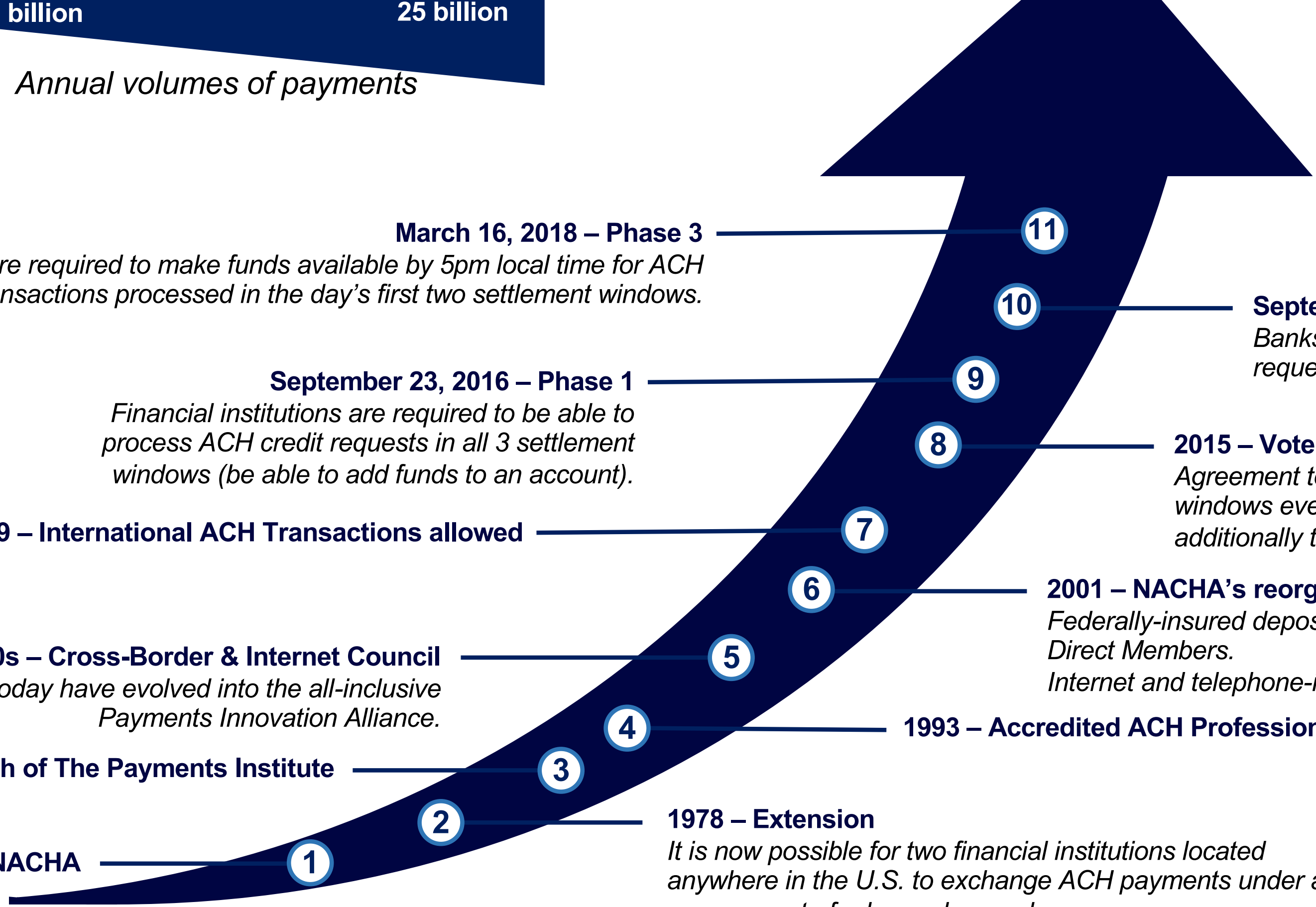


Behavioral Biometrics added value

5 – Validating Same-Day ACH Payments

The Need for Speed – A history of NACHA & ACH[8][22][23]

Annual volumes of payments



1. **1974 – Creation of NACHA**
   ACH Associations from California, Georgia, New England and the Upper Midwest form NACHA.
2. **1978 – Extension**
   It is now possible for two financial institutions located anywhere in the U.S. to exchange ACH payments under a common set of rules and procedures.
3. **1987 – Launch of The Payments Institute**
4. **1993 – Accredited ACH Professional program**
5. **1990s – Cross-Border & Internet Council**
   Those two groups today have evolved into the all-inclusive Payments Innovation Alliance.
6. **2001 – NACHA's reorganization**
   Federally-insured depository financial institutions become Direct Members.
   Internet and telephone-initiated payments become effective.
7. **2009 – International ACH Transactions allowed**
8. **2015 – Vote of *Same-Day ACH* Plan**
   Agreement to implement two new settlement windows every business day: 1pm & 5pm additionally to the existing one at 9:30am.
9. **September 23, 2016 – Phase 1**
   Financial institutions are required to be able to process ACH credit requests in all 3 settlement windows (be able to add funds to an account).
10. **September 15, 2017 – Phase 2**
    Banks additionally need to accept debit requests in the 3 settlement windows.
11. **March 16, 2018 – Phase 3**
    Banks are required to make funds available by 5pm local time for ACH credit transactions processed in the day's first two settlement windows.

Behavioral Biometrics added value

Appendix: How ACH works[21]

OFDI: Originating Depository Financial Institution

ACH File

Federal Reserve

ACH Return

RFDI: Receiving Depository Financial Institution

Bob's bank
$ +100
$ -100
$ 0

Common reasons for an ACH file to be returned:
- **Non-sufficient funds**
- **Invalid account number**
- **Payment stopped** by the owner of the bank account

Alice's bank
~~**$ -100**~~

Day 1	Day 2	Day 3	Day 4
OFDI sends ACH request to Fed, who then sends to RFDI	RFDI processes ACH request	RFDI sends ACH return to Fed, who then sends to OFDI	OFDI processes ACH return

Source of image: *https://bipartisanpolicy.org/person/jerome-powell-0/*

Take-away points #3

The ROI of Behavioral Biometrics and their added value

Behavioral Biometrics help reduce costs…

- They are a profitable investment mainly because they reduce costs from transaction frauds, rejections and escalations
- On the example we analysed, they provided around 10% value of the total net profits from transactions

…and have advantages going beyond a secure authentication.

Prevent **Replay Fraud Attacks**

Prevent **New Account Fraud**

Detect **Remote Acces Trojans**

Same-Day ACH Payments

Strong Customer Authentication

Disclaimer

Disclaimer

This presentation has been prepared for informational and educational purposes only. Although the information contained in this presentation has been obtained from sources which the authors believes to be reliable, it has not been independently verified and no representation or warranty, express or implied, is made and no responsibility is or will be accepted by the authors as to or in relation to the accuracy, reliability or completeness of any such information.

Opinions expressed herein reflect the judgement of the authors as of **[June 2018]** and may be subject to change without notice if the authors become aware of any information, whether specific or general, which may have a material impact on any such opinions.

The information of this presentation is not intended as and does not constitute investment advice or legal or tax advice or an offer to sell any securities/tokens to any person or a solicitation of any person of any offer to purchase any securities/tokens.

The authors will not be responsible for any consequences resulting from the use of this presentation as well as the reliance upon any opinion or statement contained herein or for any omission.

This presentation is confidential and may not be reproduced (in whole or in part) nor summarised or distributed without the prior written permission of the authors.

IV. Appendix

Bibliography (1/3)

BioCatch Resources

1. **BioCatch White Paper –** *From Login to Logout: Continuous Authentication with Behavioral Biometrics*
2. **BioCatch White Paper –** *Global Trends in Online Fraud (2016)*
3. **BioCatch White Paper –** *Invisible Challenges™*
4. **BioCatch Webinar –** *Frictionless Authentication and Advanced Threat Detection:*

   https://www.slideshare.net/Yanivt/bio-catch-38634545
5. **BioCatch White Paper –** *The Promise of Behavioral Biometrics: Calculating the Return On Investment*
6. **BioCatch White Paper –** *Protect Online Banking from Remote Access Trojan (RAT) Attacks*
7. **BioCatch White Paper –** *Identity Proofing in the Age of Hacks: Preventing New Account Fraud with Behavioral Biometrics (May 2017)*
8. **BioCatch White Paper –** *Validating Same-Day ACH Payments with Behavioral Biometrics*

BehavioSec Resources

8. **BehavioSec White Paper –** *Accuracy Report for Native Mobile Application*
9. **BehavioSec White Paper –** *The Payment Service Provider Challenge: Meeting EBA and PSD2 Guidelines for Strong Authentication*
10. **BehavioSec Executive Summary –** *Human Behavior as an Extra Layer of Security*

Bibliography (2/3)

SecureAuth Resources

11. SecureAuth Website

https://www.secureauth.com/solutions/two-factor-authentication

12. SecureAuth Webinar

https://www.slideshare.net/SecureAuth2FASSO/whats-new-in-idp-90-behavioral-biometrics-and-more

13. SecureAuth

https://www.secureauth.com/products/secureauth-idp/adaptive-authentication

Other Resources

14. Statista Study – *Mobile Banking*

15. Markets and Markets

https://www.marketsandmarkets.com/Market-Reports/mobile-biometric-market-255843667.html

16. Statista

https://www.statista.com/statistics/466656/telephone-banking-fraud-cases-uk

17. NuData Security

https://nudatasecurity.com/resources/blog/deciding-on-biometrics

Bibliography (3/3)

Other Resources (continued)

18. **Trend Micro –** *A Brief History of Notable Online Banking Trojans*

    https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/online-banking-trojan-brief-history-of-notable-online-banking-trojans

19. **Javelin Strategy & Research –** *2018 Identity Fraud: Fraud Enters a New Era of Complexity*

20. **Javelin Strategy & Research –** *2017 Identity Fraud: Securing the Connected Life*

21. **Gusto Engineering Blog**

    https://engineering.gusto.com/how-ach-works-a-developer-perspective-part-4/

22. **Abacus Blog – What Does Same-Day ACH Really Mean?**

    http://blog.abacus.com/what-does-same-day-ach-really-mean/

23. **National Automated Clearing House Association – Timeline**

    https://www.nacha.org/ach-network/timeline

24. **Financial Fraud Actionn UK –** *Fraud the Facts 2017*

    https://www.financialfraudaction.org.uk/fraudfacts17/assets/fraud_the_facts.pdf

25. **Findbiometrics–** *Year in Review 2016, The Most Exciting Modalities*

    https://findbiometrics.com/year-review-2016-modalities-401120/

Credits

All logos are from The Noun Project

- Trojan Horse by TNS
- OTP by Neha Shinde
- Security by Aneeque Ahmed
- Authenticator by Isman Fromes
- iPhone by Andrey Vasiliev
- Online Application by Robiul Alam
- Password Phishing by Creative Stall
- Skull Phone by Till Teenck
- Keyboard by Chanaky
- Touch by creative outlet
- Mouse by Thengakola
- Profile by Oksana Latysheva
- Fingerprint by Stephen Kelly
- Lock by Aya Sofya
- Geo location by Alex Muravev
- Network by Alexander
- Transaction by Chanut is Industries
- Bump right by Saeful Muslim
- Push button by Guilhem
- Two finger drag left by ivisual
- Measuring tape by Gan Khoon Lay
- Tablet by Lara
- Behavior by Nithinan Tatah
- Search by Luis Prado
- Bank by anbileru adaleru
- Document by arjuazka

Picture of the first page is from PYMTS.com – *BioCatch Boosts Behavioral Biometrics Tech*

Payment solutions: from contactless to Apple Pay

Master Thesis 2018

Matteo Screnci, Louis Marty, Damien Mossuz

HEC PARIS

Introduction (1/3)

The development of new payment solutions from contactless to Apple Pay

A sector impacted by technologies …

Speed — **Speed is crucial in both payment and clearing of the transaction**

Big Data — **New ways of payments allow to have access to more data**

Security — **Safety is of tremendous importance while new threats appeared with contactless and mobile payments**

Crypto — **Can cryptocurrencies and more precisely Bitcoin change the way we pay?**

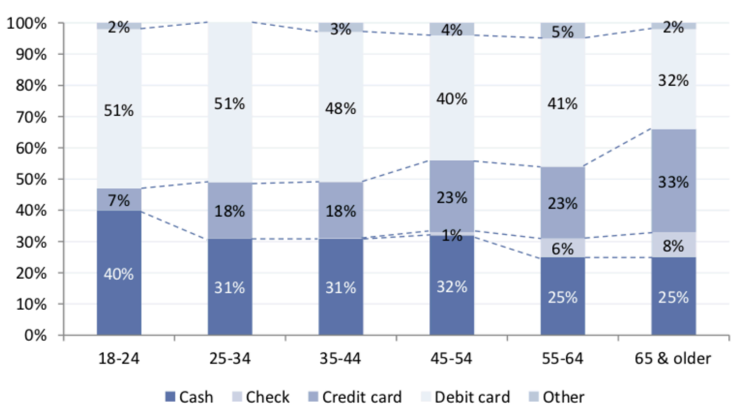
… but facing regulations

Consumer Protection Law: *essential to frame the use of these new ways of payments without threatening the safety of users' data*

Compliance Requirements: *to fight against money laundering and fraud, we will notably see that with cash payments*

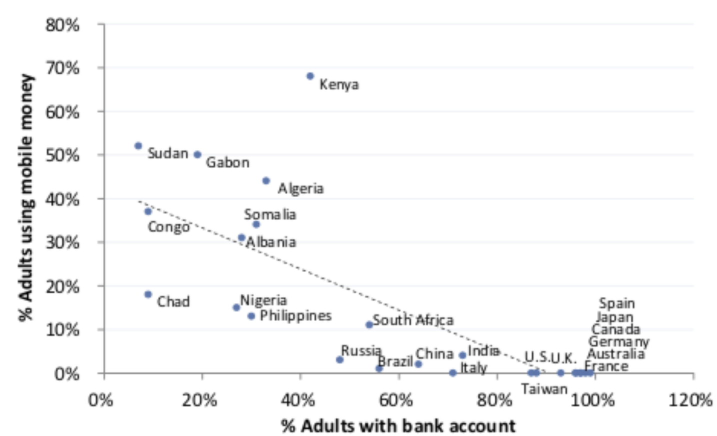
Interchange Rules: *managing the fees paid to issuer, the most important part of the merchant discount rate and that could potentially be disrupted by new ways of payment*

With surprising data on both demography…



Source: US Federal Reserve.

... and countries



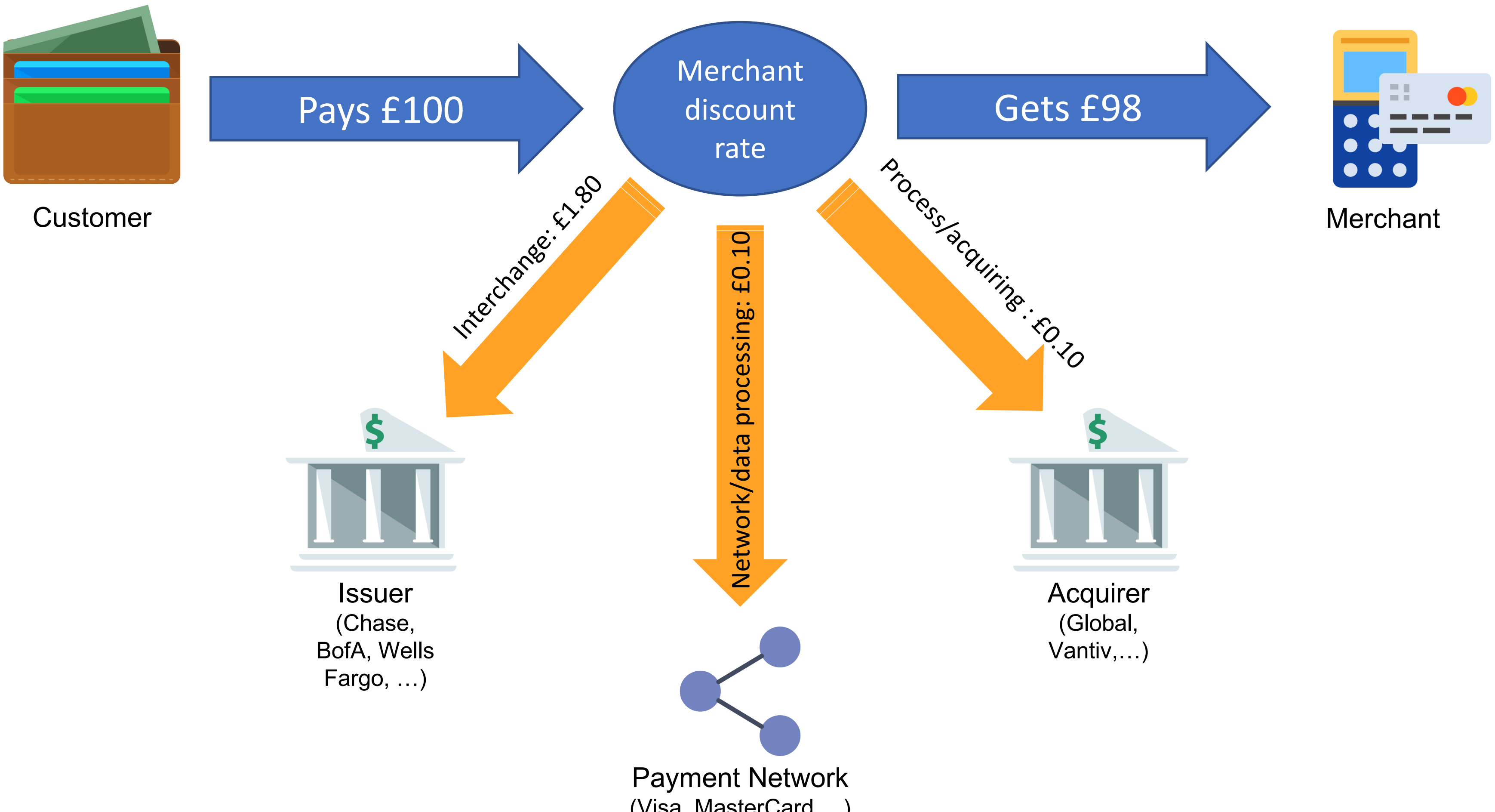
Source: World Bank's Global Findex Database, 2012

Source: *Goldman Sachs Global Investment Research : The Future of Finance Part 2*

Introduction (2/3)

The actual payment ecosystem

What are the fees and the actors when you use a credit card?



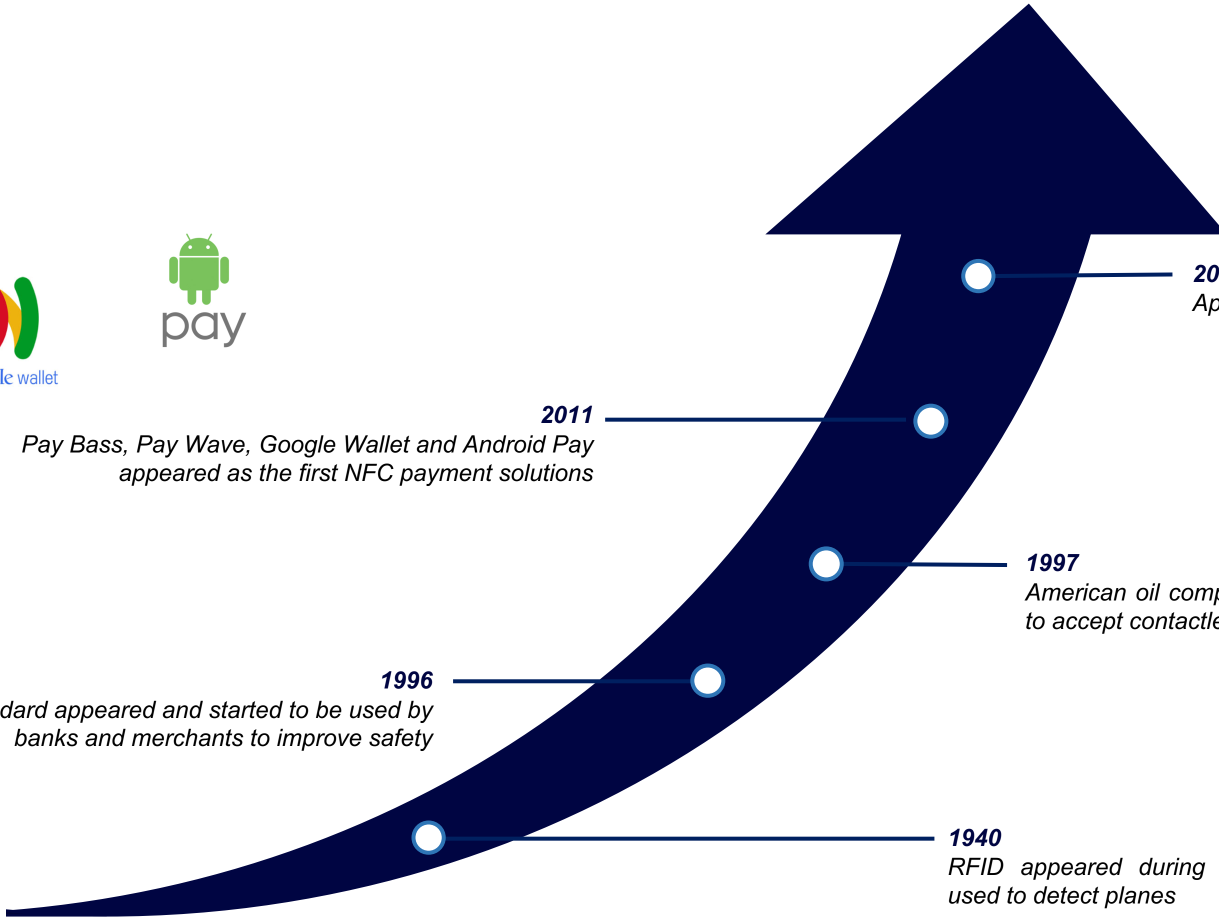
Source: *Goldman Sachs Global Investment Research : The Future of Finance Part 2*

Note: Icon made by Freepik, RoundIcons, Smashicons from www.flaticon.com

Introduction (3/3)

The development of new payment solutions from contactless to Apple Pay

History of contactless payments



2014
Apparition of Apple Pay

2011
Pay Bass, Pay Wave, Google Wallet and Android Pay appeared as the first NFC payment solutions

1997
American oil company Mobil became the first to accept contactless payment

1996
EMV standard appeared and started to be used by banks and merchants to improve safety

1940
RFID appeared during WW2 and was used to detect planes

Source: https://www.payfirma.com/payments-101/contactless-payments-a-history/ and http://www.centrenational-rfid.com/introduction-a-la-rfid-article-15-fr-ruid-17.html

Thesis plan

The development of new payment solutions from contactless to Apple Pay

I.	The RFID Technology
II.	Risks of RFID Technology
III.	The NFC Technology
IV.	Impacts on merchants
V.	Impacts on customers
VI.	Impacts on issuers

HEC PARIS

The RFID Technology

Payment standards before contactless payments: EMV

EMV standards…

- ***Founded*** *in 1993 by Europay, Mastercard an Visa, joined lately by JCB International and American Express.*
- *Standards for payment cards since* ***1995***
- *Imposed in France and cards and payment terminals since* ***2006***

Define standards for ***cards with integrated circuit*** *in the goal to* ***improve safety*** *in all type of transactions (including today contactless and NFC payments).*
Second objective is to allow ***"interoperability and compatibility"*** *of all credit cards and payment terminals around the world .*

100% *of European Cards follow EMV standards since* ***2010****. In* ***2016****,* ***70%*** *of US Cards are EMV and 50% of of merchants are EMV compliant.*
Transitions pushed by a ***"liability shift":*** *all merchants non-EMV compliant will be found liable for any fraud (started in 2005 in EU and 2015 in the US)*

… are described in four « books »

Book 1: Application Independent and ICC Control

Describe the parameters (physics, software and electrics) the card needs to respect and the way it will exchange with the payment terminal

Book 2: Security and Key Management

Describe the way the card will be authenticated by the payment terminal (with Static Data Authentication and Dynamic Data Authentication). It should protect against data modification and cloning

Book 3: Application Specification

Organize the way the payment terminal access the files on the card and verify the identity of the client through a Cardholder Verification Method (most of the time the PIN).

Book 4: Cardholder, Attendant and Acquirer Interface Requirements

Describe the way interface should be designed and should access data, notably for online transactions. Details about the CVV (Card Verification Value).

Source: ***https://en.wikipedia.org/wiki/EMV#North_America*** ***and*** ***https://cedric.cnam.fr/~bouzefra/cours/Cartes_Bouzefrane_EMV_nov2009.pdf***

The RFID Technology

Radio-Frequency Identification Technology is at the origin of contactless payments

Overview of the RFID Technology with electronic chip

What is RFID?

Data

Allows to transfer data using radio frequencies

RFID Tag

Objects wear small chips, called RFID Tag were data are stored

Antenna

RFID Tags exchange information with an interrogator/reader through an antenna

Memory

Memory of the tags is limited, most of them only contain specific information

WORM

Information stored are WORM (Write Once Read Multiple) most of the time, meaning they cannot be changed

Source: *lhttp://www.centrenational-rfid.com/fonctionnement-dun-systeme-rfid-article-17-fr-ruid-17.html and https://cedric.cnam.fr/~bouzefra/cours/CoursNFC_Bouzefrane_Decembre2013.pdf*

Note: Icon made by Smashicons, Freepik from www.flaticon.com

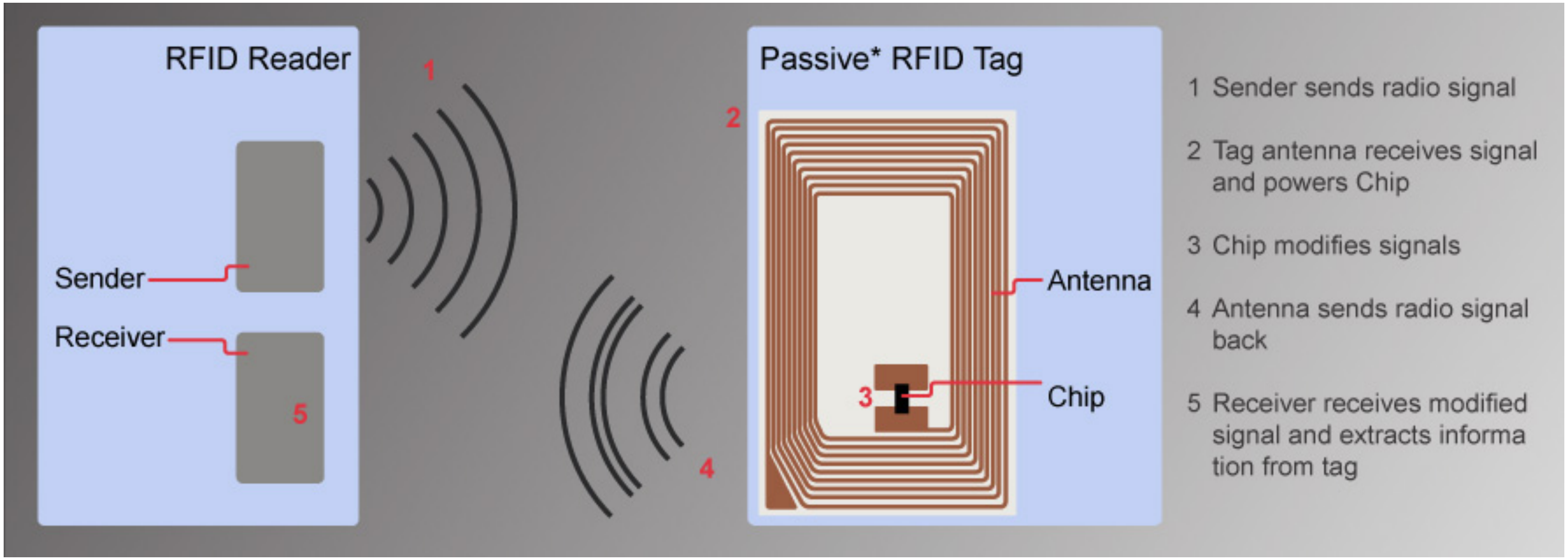
The RFID Technology

Different types of RFID Tags

Most of RFID Tags have integrated circuits, however we can still distinguish three different types of Tags

Active		Semi-Active		Passive	
✔	Transmitter	✘	Transmitter	✘	Transmitter
✔	Power supply	✔	Power supply	✘	Power supply

Passive is the most common type, it is used for contact less credit cards



Source: http://www.centrenational-rfid.com/fonctionnement-dun-systeme-rfid-article-17-fr-ruid-17.html and https://electronics.stackexchange.com/questions/334419/how-does-rfid-reader-reads-a-passive-rfid-tag

The RFID Technology

Different protocols and frequency range

Protocols …	
ITF	**Interrogator Talk First**
TTF	**Tag Talk First**

… and frequency ranges

Low Frequency (LF): *125 kHz-134.2 kHz*
Expensive and bad performances in metal or liquid environment

High Frequency (HF): *13.56 MHz*
Cheap and single frequency in the world

Ultra High Frequency (UHF): *860 MHz-960 MHz*
Good for high distances but less performant than HF in liquid and metal

Super High Frequency (SHF): *2.45 GHz*
Highly sensible to liquid and metal

- **Contact less credit cards use ITF protocol and High Frequency range**
- **High Frequency is also used for NFC systems, that we will discuss later**

Source: http://www.centrenational-rfid.com/fonctionnement-dun-systeme-rfid-article-17-fr-ruid-17.html
Note: *(1) As of January 7th 2018*

Thesis plan

The development of new payment solutions from contactless to Apple Pay

I.	**The RFID Technology**
II.	**Risks of RFID Technology**
III.	**The NFC Technology**
IV.	**Impacts on merchants**
V.	**Impacts on customers**
VI.	**Impacts on issuers**

Risks of RFID Technology

Data stored on the chip of contact less credit card

Data stored on the chip of contact less credit card have evolved

Card	CVV	Name	Card Number	Expiration Date
Old Chip		Name	Card Number	Expiration Date
New Chip (since 2012)			Card Number	Expiration Date

Storing the same information allows to keep using same POS system (no additional cots)

CVV are not stored on the chip but dynamically generated at each transaction for more safety

HEC PARIS

Source: *http://www.nfc.cc/2012/04/02/android-app-reads-paypass-and-paywave-creditcards/* ***and*** *https://www.dlapiper.com/en/singapore/insights/publications/2015/10/contactless-credit-cards/*

Risks of RFID Technology

Even if CVV are not stored there are still some risks

Payments where CVV is not required …

- ***Some websites don't ask for CVV at the payment:*** *(1)*
  - ***Amazon***
  - ***Target***
  - ***Rakuten***
- ***You are also bypassing the value of transaction limit***

… or seizing the dynamic CVV

- ***Faking a POS with a card reader to obtain the one-time CVV***
- ***You cannot bypass the value of transaction limit***

*A **card reader** can be bought for $50 on eBay. You can even find an **app on Android NFC phones** that enables you to read information stored on the chip. Then, using a card magnetizing tool you can encode the data onto a fake blank card that you can use to pay.*

Source: https://www.forbes.com/sites/andygreenberg/2012/01/30/hackers-demo-shows-how-easily-credit-cards-can-be-read-through-clothes-and-wallets/#486084d278a6 and http://www.nfc.cc/2012/04/02/android-app-reads-paypass-and-paywave-creditcards/

The question of whether the name should be stored or not

	Advantages	Disadvantages
Storing the name	Should bring more safety if compared to the identity of the person trying to pay	Can be read by a card reader and used to create a fake card

HEC PARIS

Note: *(1) https://www.quora.com/What-online-shopping-stores-dont-require-CVV-code*

Risks of RFID Technology

Mitigating those risks

What are the ways to reduce the risks of contact less credit cards?

Transaction Value Limit	Encryption	RFID Shield
Transaction Value Limit to limit losses in case your card get stolen: - France: €30 - UK: £30	Cryptography according to EMV standards that can only be decrypted by a genuine POS provided by a genuine bank	Wallets, backpacks and jeans with RFID shield are now sold. They use metal fibers that block RFID communication.

However nothing can protect you against getting your card stolen and used against your will!

Or maybe NFC enabled device can?

Source: https://www.corporatetravelsafety.com/safety-tips/rfid-pickpockets/ and https://www.gemalto.com/brochures-site/download-site/Documents/documentgating/fs-why-contactless-pickpocketing-impossible.pdf?webSyncID=079d946a-00b7-01bb-740d-9c2445a59040&sessionGUID=64b82cf1-9134-4e28-890b-666ea54cd4a9

Thesis plan

The development of new payment solutions from contactless to Apple Pay

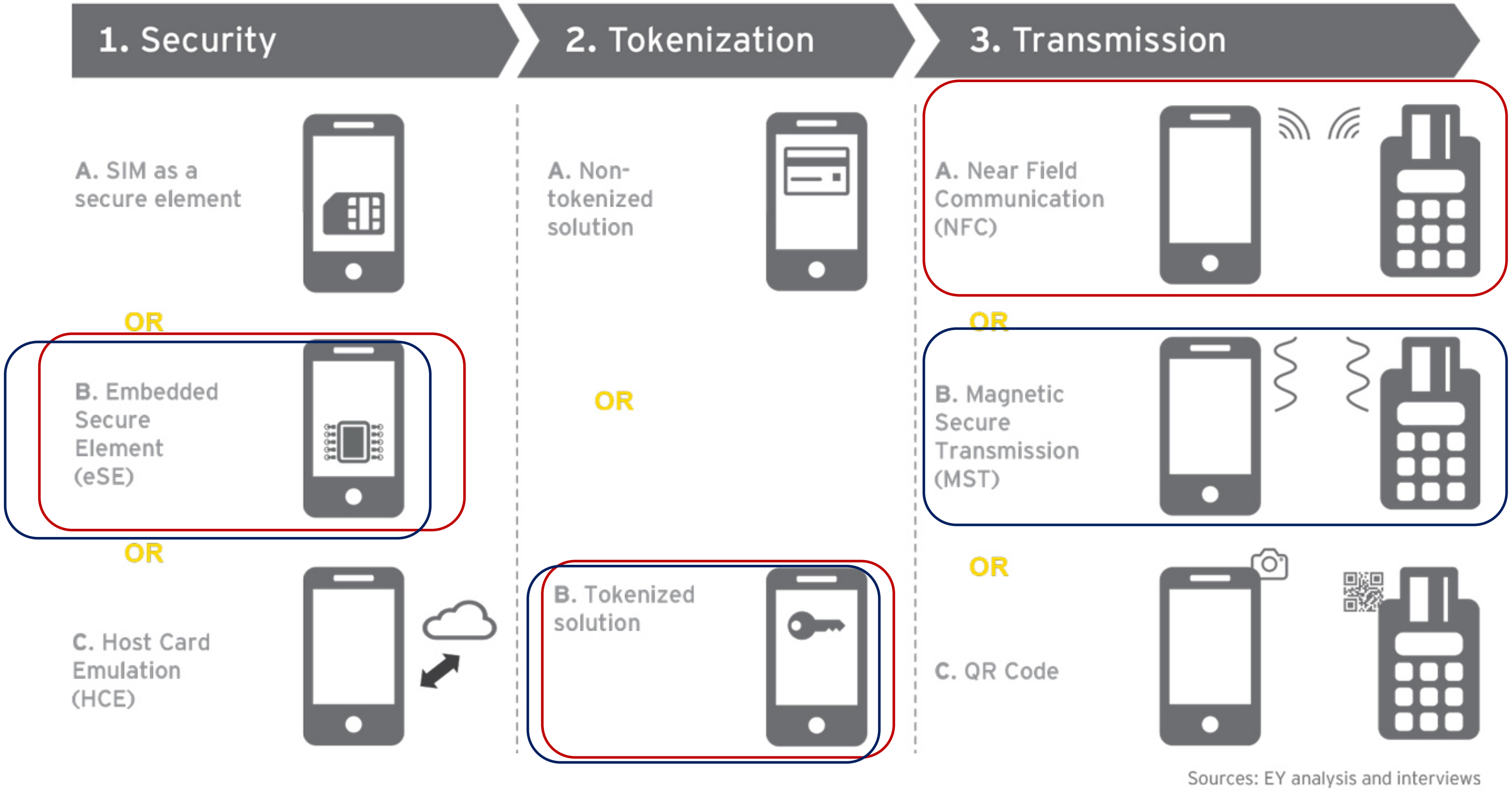
I.	The RFID Technology
II.	Risks of RFID Technology
III.	The NFC Technology
IV.	Impacts on merchants
V.	Impacts on customers
VI.	Impacts on issuers

HEC PARIS

Annex: the three steps of mobile payments

Mobile payments can be decomposed in three steps, each with different solutions

Overview of mobile payment solutions



Apple Pay

Samsung Pay

Source: http://www.ey.com/Publication/vwLUAssets/ey-mobile-payment-war-of-wallets-nov-2015/$FILE/ey-mobile-payment-war-of-wallets-nov-2015.pdf

The NFC Technology

Near Field Communication Technology allows safer payment solutions

Overview of the NFC Technology

What is NFC?

RFID

Based on RFID communication between two electronic devices

High Frequency

Based on the HF RFID Technology (13.56 MHz)

Near Field

Very short distances for safety and technical reasons (less than 4cm)

Source: *http://www.centrenational-rfid.com/introduction-au-nfc-article-132-fr-ruid-17.htm* /

Note: Icon made by Smashicons, Freepik from www.flaticon.com

Annex: the other transmission technologies

Alternatives to NFC for mobile payments

What are the alternatives and why are they different?

Magnetic Secure Transmission	**QR Code**
Used by Samsung Pay only **Allows to pay on all POS equipped for magstripe payment (less restrictive than NFC)** **Less secure than NFC and need to give the phone to the merchant (POS with magstripe are often behind the counter)**	**Can be used on all phones but only with POS displaying a QR Code** **Less convenient (need to open the camera) and less safe than standard NFC**

NFC is the most efficient transmission solution and even if MST has the advantage to work with more POS it is only seen as a transition before moving to standard NFC

Source: [*http://www.ey.com/Publication/vwLUAssets/ey-mobile-payment-war-of-wallets-nov-2015/$FILE/ey-mobile-payment-war-of-wallets-nov-2015.pdf*](http://www.ey.com/Publication/vwLUAssets/ey-mobile-payment-war-of-wallets-nov-2015/$FILE/ey-mobile-payment-war-of-wallets-nov-2015.pdf)

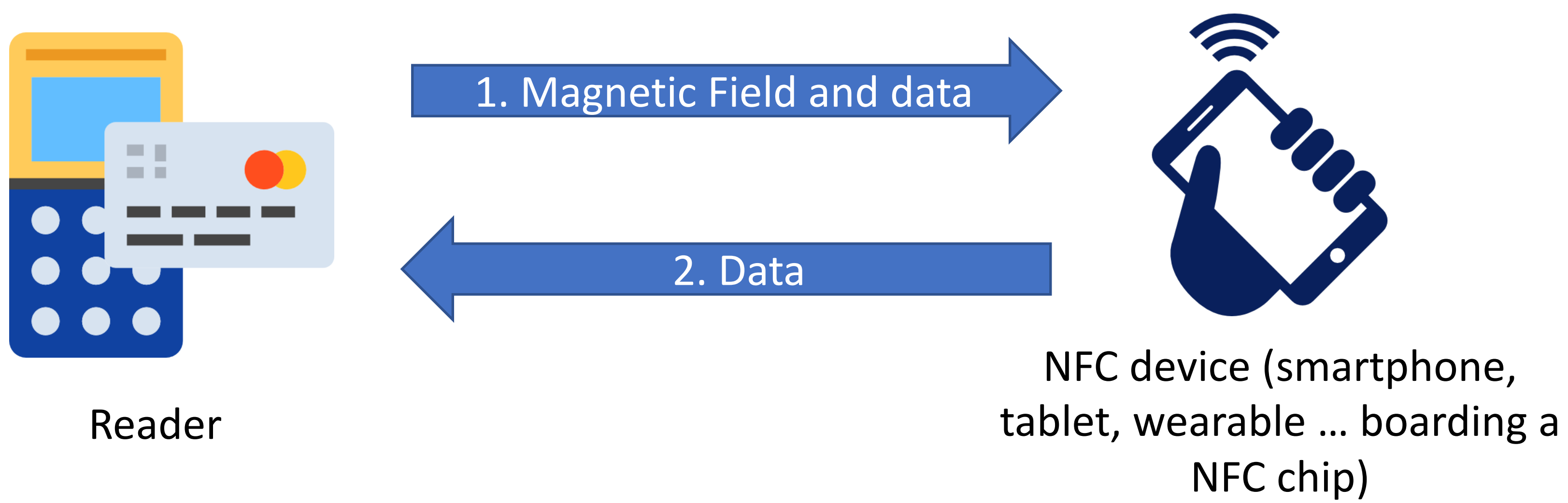
The NFC Technology

The different ways to use NFC Technology

NFC Technology can be used in three different ways

Reader	Emulator	Peer-to-Peer
The NFC device plays the role of the interrogator/reader Used in mobile POS solutions: seller can cash you using mobile phone	The NFC device plays the role of the RFID tag, by emulating a credit card when needed Used by Apple Pay for example	The NFC device can be simultaneously sender or receiver Less common because too slow, used by Android Beam (Air Drop's opponent)

Focusing on emulation of credit card



Source: http://www.centrenational-rfid.com/introduction-au-nfc-article-132-fr-ruid-17.html

Note: Icon made by Freepik from www.flaticon.com

The NFC Technology

Why is payment through NFC safer than standard RFID?

The reasons that makes NFC payment safer

Emulation	ODCVM	Tokenization
Credit card is only emulated when needed thanks to HCE (Host Card Emulation) technology or to eSE (embedded Secured Element) Data are not accessible at any time like for RFID tags	Mobile security like PIN or even biometric identification (Face ID, Touch ID) reduces the risks of use against your will These are called ODCVM (On Device Cardholder Verification Methods)	Credit card numbers are replaced by a token (random series of numbers) that can ONLY be detokenized by the Point of Sale during the transaction The ability to detokenize is given by the issuers of the cards to a restrictive list of people

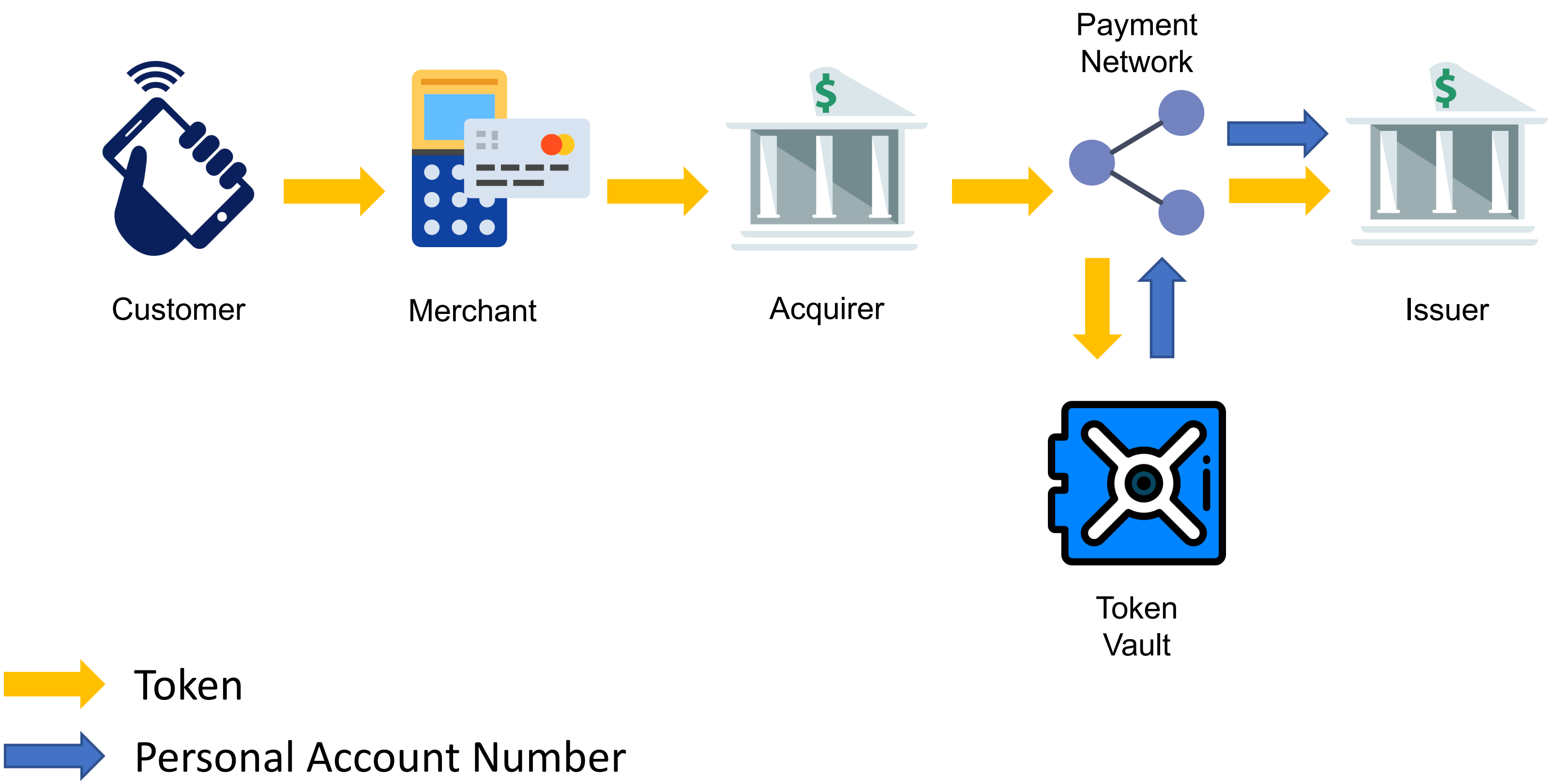
Thanks to ODCVM, you are not restricted to the usual Transaction Value Limit that you have on contact less credit cards

All these reasons makes NFC technology safer and more convenient to use than standard contact less credit cards

Annex: tokenization

How does tokenization work? (1/2)

General principles of tokenization



Source: http://www.contactlesspaymentcards.com/whatistokenizationpayments.php

Note: Icon made by Freepik, RoundIcons, Smashicons from www.flaticon.com

Annex: tokenization

How does tokenization work? (2/2)

General principles of tokenization

Tokenization	• **The Personal Account Number (PAN) is replaced by a token** • **The token** is a randomly generated number furnished by the **Token Service Provider (TSP)** • This token is sent by the customer to the merchant
Token Vault	• After going through the acquirer and the payment network, the token arrives to the **token vault** • The **token vault** is hold by the **Token Service Provider** • The TSP will send back the token and the corresponding PAN to the payment network
Issuer	• The payment network will then send **the token and the PAN** to the **issuer** • If both matches, the **issuer will authorize the transaction at the POS of the merchant**

Source: *http://www.emv-connection.com/downloads/2016/06/Contactless-2-0-WP-FINAL-June-2016.pdf*

Annex: emulation

Two major ways to emulate the card are used today

What are the differences between these two ways?

embedded Secured Element (eSE)	Host Card Emulation (HCE)
Used by Apple Pay* and Samsung Pay The card data are stored directly on the phone, the Secured Element is on the device That does not mean that the real data is stored in all the cases, eSE can be used with tokenization for more safety. Apple Pay only stored the corresponding token *Apple Pay actually uses a mix between HCE and eSE for more safety	Used by Android Pay Nothing is stored on the device The data are stored in the cloud, the Secured Element is in the cloud.

HCE is globally less secured and provides less privacy to the user

However, HCE is more easily scalable and doesn't require any hardware on the phone (allowing to sell cheaper phones)

Source: http://www.ey.com/Publication/vwLUAssets/ey-mobile-payment-war-of-wallets-nov-2015/$FILE/ey-mobile-payment-war-of-wallets-nov-2015.pdf

Thesis plan

The development of new payment solutions from contactless to Apple Pay

I.	**The RFID Technology**
II.	**Risks of RFID Technology**
III.	**The NFC Technology**
IV.	**Impacts on merchants**
V.	**Impacts on customers**
VI.	**Impacts on issuers**

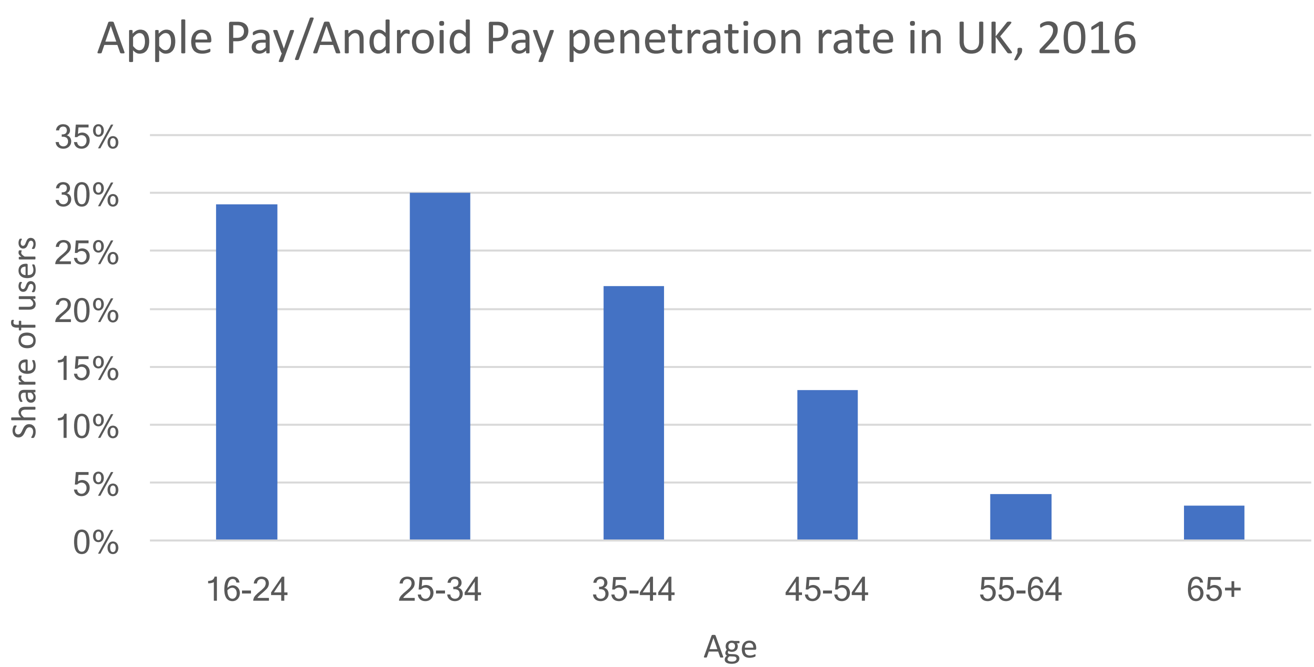
HEC PARIS

Impacts on merchants

Contactless payment in general has had several impacts on merchants

Customer Experience

- **Offer better customer experience**
- **More different ways to pay**
- Seducing **young / tech-friendly** (and high income) population with technologically advanced way of payments (Apple Pay, Google Pay, Android Pay etc.)

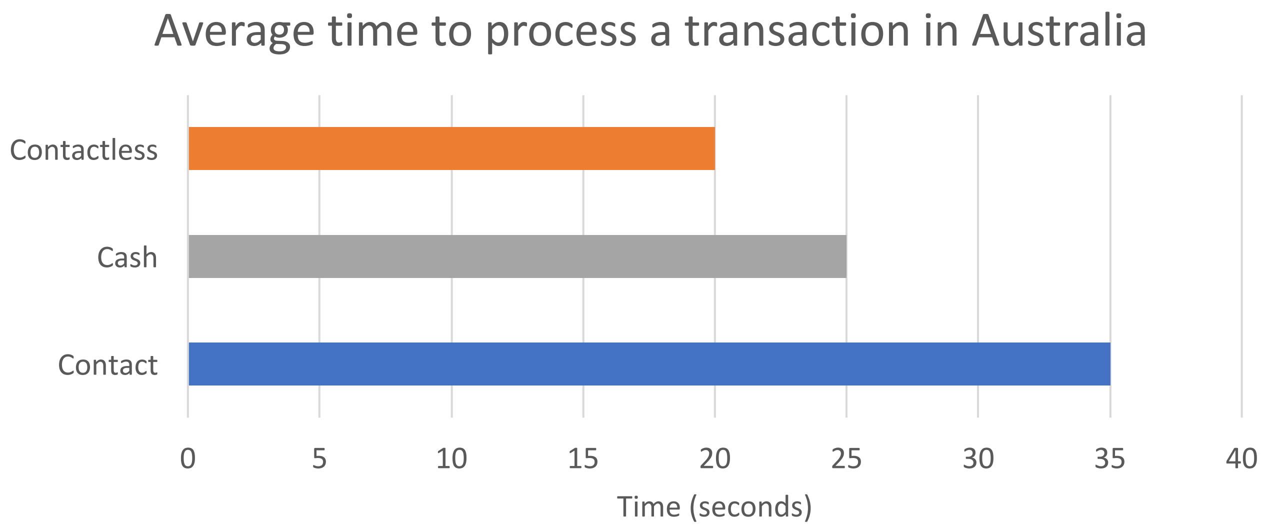


Source: *"Contact less payment in the UK", Statista*

Impacts on merchants

Contactless payment in general has had several impacts on merchants

Time	
	• **Gain in time on small transactions because no need to enter PIN** • **Time saved** by reducing the risks of **dysfunction during the payments** (unreadable card) • **ODCVM** allows to gain time also on biggest transactions



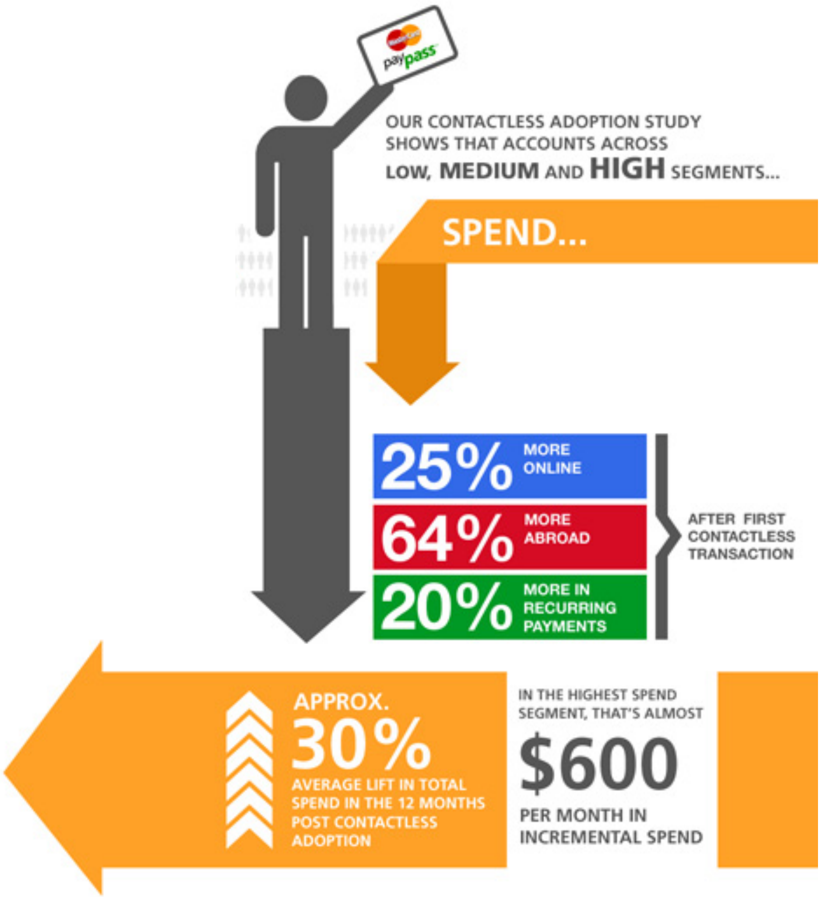
Source: https://www.rba.gov.au/publications/rdp/2014/pdf/rdp2014-14.pdf

Impacts on merchants

Contactless payment in general has had several impacts on merchants

Transactions

- Increase in **number of transactions** (thanks to a decrease in time spent by transaction)
- Increase **amount spent** by customers



LOW	• $ 499
MEDIUM	• $500 - $1,499
HIGH	• $1,500 and more

Spending per months with the card

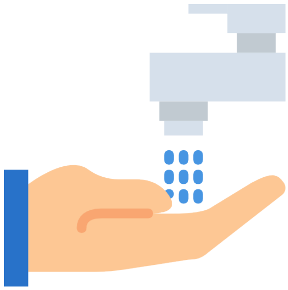
Impacts on merchants

Contactless payment in general has had several impacts on merchants

Cash Handling	• Reduce **cash handling** • Reduce **risks** (loss, robbery) **and time wasting** • Solve **hygiene issues** (restaurants)

More germs on £1 coin than on a toilet seat

Only 1 out of 5 European admit washing their hands after handling money



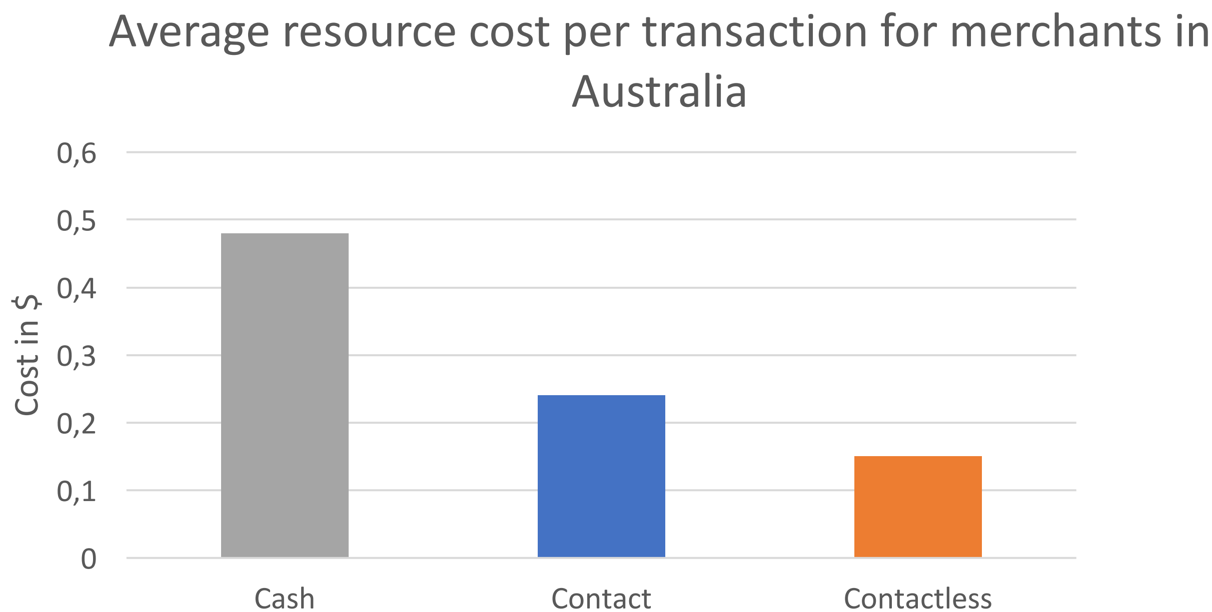
Source: http://www.dailymail.co.uk/sciencetech/article-2621500/Dirty-cash-Bank-notes-contain-26-000-bacteria-half-Britons-wash-hands-handling-them.html

Note: Icon made by Freepik, Smashicons from www.flaticon.com

Impacts on merchants

Contactless payment in general has had several impacts on merchants

Costs	• Decrease in **cost per transaction** • 2 seconds change in **time spent per transaction** also reduces costs by $0.01 • Cash is a **more expensive payment solution**



Source: https://www.rba.gov.au/publications/rdp/2014/pdf/rdp2014-14.pdf

Annex: the cost of cash (1/2)

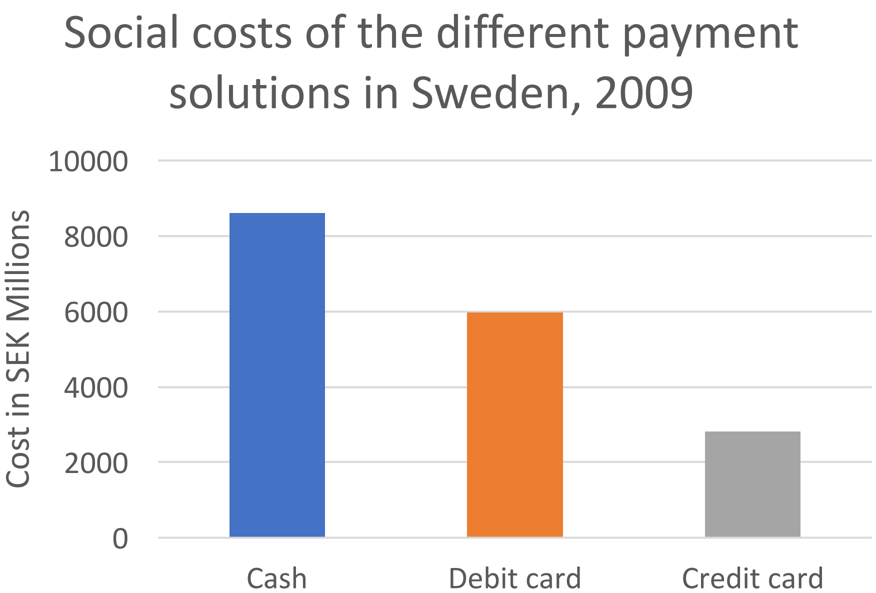
Why is cash more costly?

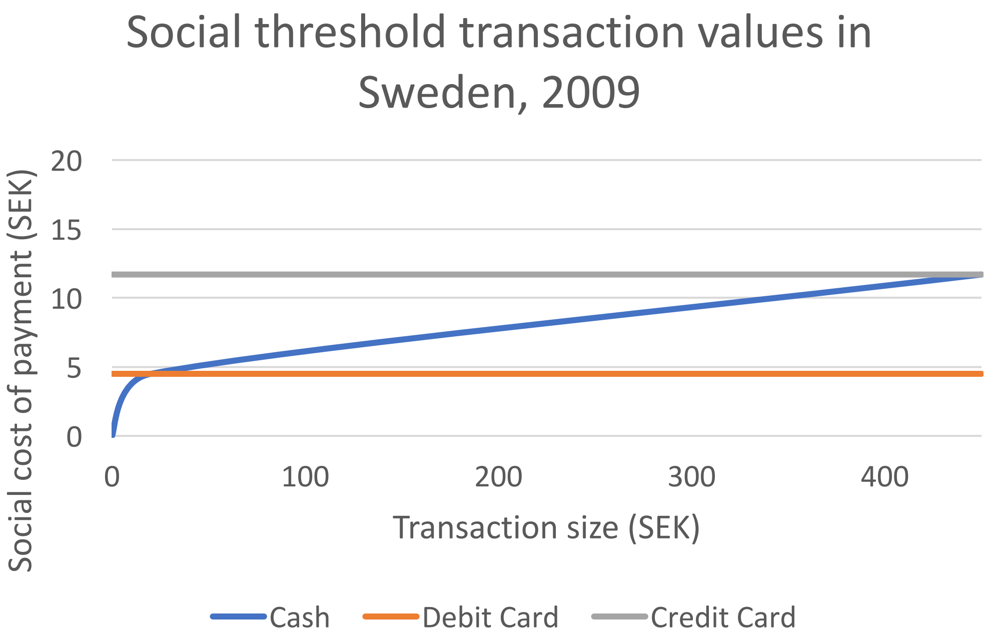
Central Bank	• Production of note and coins • Vault keeping • Interest rates • Destruction, sorting, …
Cash-in-Transit (CIT)	• Collection, transport • Cash handling • Security
Banks	• ATM withdrawals • Deposits • Fees paid to CITs
Retail Sector	• Fees paid to CITs and banks • Time lost (during transaction and in back-office) • Maintaining cash register, paper rolls, …
Consumers	• Fees paid to banks • Time lost • Interests lost

Source: *https://www.econstor.eu/bitstream/10419/81925/1/71906726X.pdf*

Annex: the cost of cash (2/2)

Cash compared to other payment solutions





- ***It is less costly on a social point of view to use debit card instead of cash for transactions above SEK 4.5 (€ 0.42)***

Source: ***https://www.econstor.eu/bitstream/10419/81925/1/71906726X.pdf***

Note: Icon made by Freepik from www.flaticon.com

Thesis plan

The development of new payment solutions from contactless to Apple Pay

I.	**The RFID Technology**
II.	**Risks of RFID Technology**
III.	**The NFC Technology**
IV.	**Impacts on merchants**
V.	**Impacts on customers**
VI.	**Impacts on issuers**

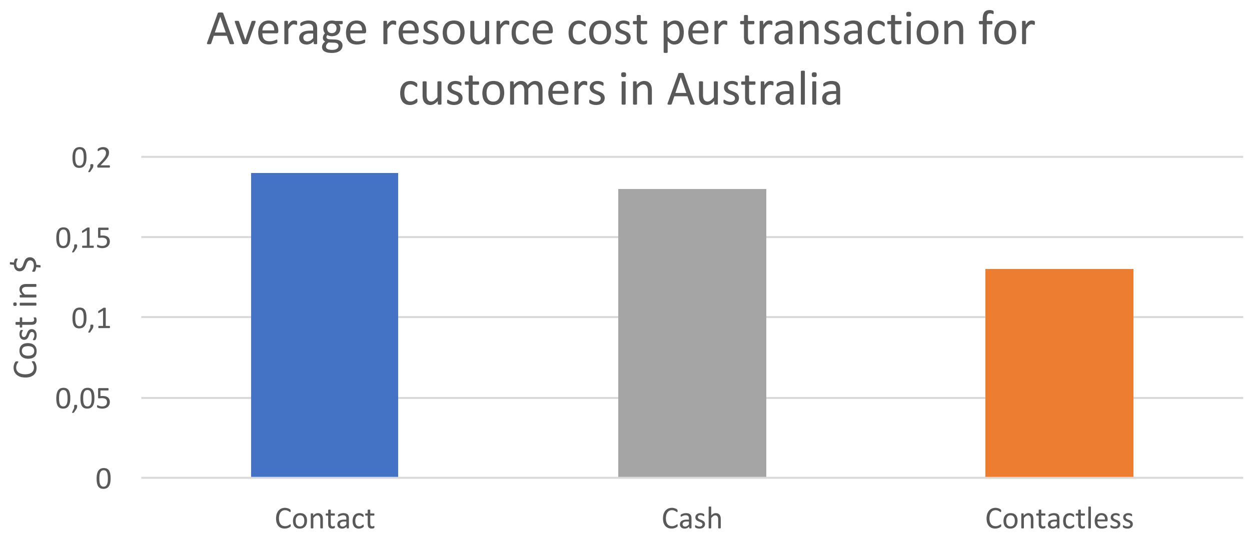
Impacts on customers

Positive impacts also for customers

Similar impacts as for merchants

Time	Cash	Costs
Customers benefit from gains in time as they wait less when at the shop	Customers do not need to carry cash anymore	It is also more cost effective for customers to pay with contactless credit cards

Focusing on costs



Source: https://www.rba.gov.au/publications/rdp/2014/pdf/rdp2014-14.pdf

Note: Icon made by Freepik from www.flaticon.com

Impacts on customers

More positive impacts linked to NFC

More safety …

- ***Thanks to ODCVM, credit cards cannot be used by offenders even if the NFC device get stolen / lost***
- ***Your card number can also not be memorized by the merchant***

… and improvement of expenses management

- ***Apps can be linked to your NFC device to improve management and keep tracks of your expenses***
- ***No need for paper bills so good ecological impact***

***Mobile Banking** in general also had a huge impact on **developing countries** (as we saw in the introduction). Indeed, it helped unbanked people to **get access to financial services**. The spectrum of impact goes from mobile transfer (with M-Pesa in Kenya for example) to opening and accessing on a phone to a bank account from rural places (MyBank in China).*

Source: https://fin.plaid.com/articles/mobile-innovation-financial-inclusion

Thesis plan

The development of new payment solutions from contactless to Apple Pay

I.	The RFID Technology
II.	Risks of RFID Technology
III.	The NFC Technology
IV.	Impacts on merchants
V.	Impacts on customers
VI.	Impacts on issuers

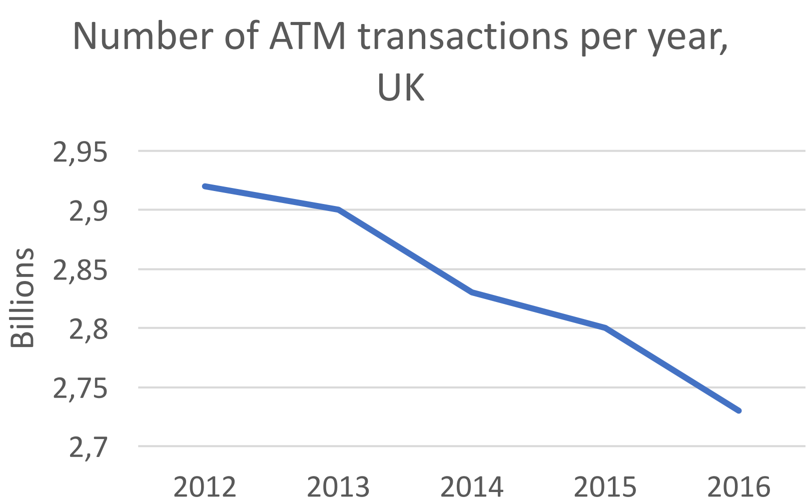
Impacts on issuers

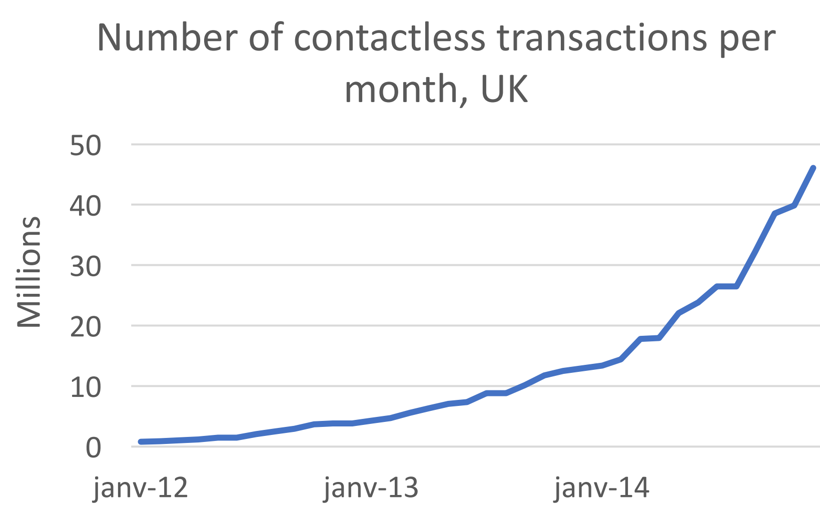
Positive impacts also for issuers

Similar impacts as for merchants and customers

Transactions	Cash	Safety
Issuers benefit from the increase in number and value from transaction	Allows to bank to access to small transactions that were usually paid in cash	Issuers also benefit from the increase in safety as they have less money to reimburse to customers victim of frauds

Focusing on cash





Source: ***https://www.theguardian.com/money/2018/feb/19/peak-cash-over-uk-rise-of-debit-cards-unbanked-contactless-payments*** ***and*** ***https://www.telegraph.co.uk/news/shopping-and-consumer-news/11456380/The-end-of-cash-as-we-know-it.html***
Note: Icon made by Freepik from www.flaticon.com

Annex: consumer protection laws

Issuers are liable for consumers in most of the fraud cases

Consumer protection in two major cases

Consumer Fraud Liability	• **Popular in the US** • **Most of the time customers face zero liability (meaning that all the costs are reimbursed by the issuers) in case of fraudulent or unauthorized use of their cards**
Chargeback	• **Protection against merchants provided by issuers** • **In case of non delivery of the article or delivery of an unsatisfactory article the consumer can be reimbursed by the issuer of his card**

Consequences on payment forms

- ***Credit/Debit cards provide huge advantages to customers that will need to be extended on mobile payment to make the payment solution competitive***
- ***Fraud in general is very costly to issuers, providing incentives to them to reinforce safety of payment solutions***

Source: *Goldman Sachs Global Investment Research : The Future of Finance Part 2*

Annex: decrease in deception rates

Data from the Victoria Police in Australia

Deception

Definition from the Victorian Government Crime Statistics Agency: "Offences involving a dishonest act or omission carried out with the purpose of deceiving to obtain a benefit or avoid a disbenefit. This includes: Forgery and counterfeiting; Possess equipment to make false instrument; Obtain benefit by deception; State false information; Deceptive business practices; Professional malpractice and misrepresentation; Other deception offences"(1)

Statistics

	2012	2013	2014
Number of Deception Offences	3890	5539	5292
Number of contactless transactions	37.2m	149.9m	352m
Percentage of offences against transactions	0.01%	0.004%	0.002%

- ***The number of contactless transactions doubled between 2013 and 2014, whereas the number of deception offences slightly decreased***
- ***Percentage of offences against transactions has been divided by 5 between 2012 and 2014, suggestion that contactless is a safer way to pay***

Source: *http://www.canberra.edu.au/cis/storage/Contactless%20payments.pdf*

Note: *(1) Page 7*

Conclusion

The change in way people pay

Advantages of contactless and NFC …		... pushed by regulations
Costs	**Less costly on a global point of view than cash**	***Banks:*** *Lots of them do not accept cash deposit anymore, specially true for deposits into other customers' accounts (BofA, Wells Fargo, HSBC, Chase)* *Banks also have extended KYC requirements*
Safety	**Safer than any other way of payment (particularly true for NFC with ODCVM)**	***Governments:*** *Cash payment limits are reduced years after years (went from €3,000 to €1,000 in France in 2015)*

- **However, contactless payment has still low penetration rate with elder and less educated people (Bank of Sweden study) and with lower household income (Statista, UK, 2016)**
- **Therefore, cash is still widely used for several reasons**

Source: *http://www.dotecon.com/assets/images/DotEcon-Costly-cash-report-v5-3.pdf, "Contactless Payment in the UK", Statista and Goldman Sachs Global Investment Research : The Future of Finance Part 2*
Note: *(1) As of January 7th 2018*

Why cash is still widely used?

Looking at all the advantages of contactless payments, cash should disappear

Lack of transparency on costs linked to cash …

- ***Huge lack of transparency on the costs:***
  - Time
  - Storage
  - Counting
- ***ATM costs are paid by all customers, not only users, making the global cost of cash higher***

… and lack of trust in contactless

- ***35% of men and 29% of women said they don't pay with contactless because they don't trust it (UK, 2016)***
- ***53% of the users think that there is a risk that someone could steal information (UK, 2016)***

Cash is the fundamental of the "Shadow Economy"

- ***Last reason why cash is still widely used and also the one that justify the most why we should fight against cash payments***
- ***Cash payment is "most of the time" asked for illegal reasons (avoiding tax payments for example)***

2010 study from AT Kearney

ATKearney

-0.78

Correlation between size of the shadow economy of a country and number of electronic transactions per inhabitant

"increasing electronic payment by 10% can lead to decline in the size of the shadow economy by up to 5%"(1)

HEC PARIS

Source: *"Contactless Payment in the UK", Statista and http://www.dotecon.com/assets/images/DotEcon-Costly-cash-report-v5-3.pdf*

Note: *(1) "Costly Cash: a synthesis of international evidence on the cost of making payments", DotEcon, (p.44) http://www.dotecon.com/assets/images/DotEcon-Costly-cash-report-v5-3.pdf*

Annex: Bitcoin, the payment solution of the future? (1/2)

Can Bitcoin become the way of payment of the future?

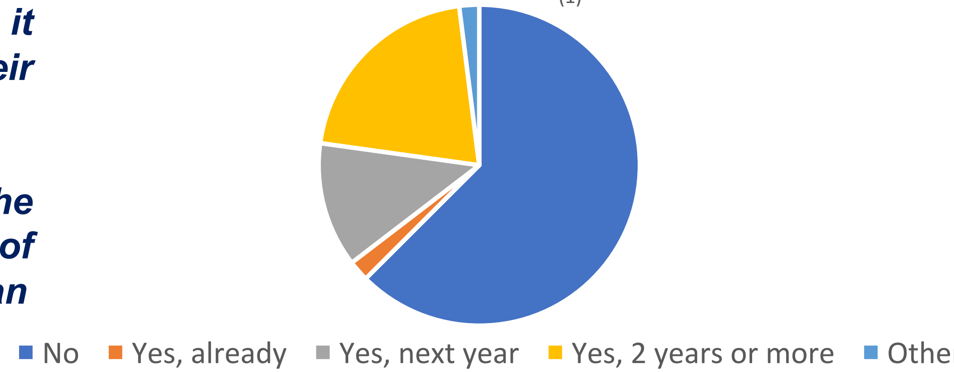
Advantages

Transparency	**Bitcoin promises more transparency thanks to the distributed ledger, it could therefore help to restore trust in some industries (digital advertising)**
Fees	**Thanks to Bitcoin merchants are likely to face less fees (merchants discount) during a transaction. This is notably linked to the fact that you do not need a payment network anymore. On top of that it should create competition between the acquirers and issuers, hopefully resulting in a decrease in fees charged**
Clearing Time	**Bitcoin should also be able to reduce significantly clearing time. Indeed, there would be no need to go through the ACH (Auto Clearing House) in the US for example. Whereas it can take up to three days to receive the funds with the ACH, it takes 10min with Bitcoin (if the merchant does not ask for the Bitcoin amount to be converted back into fiat currency)**

Adoption

- ***A few well know firms already accept bitcoins: Expedia, Dell, … For most of them it still represents a low percentage of their revenue (0.2% for Overstock.com in 2014).***
- ***Bitcoin payment restricted in China by the People's Bank of China, even if 80% of Bitcoin exchanged volume is made with yuan***

"Do you plan to enable the acceptance of Bitcoin for your merchants?"(1)



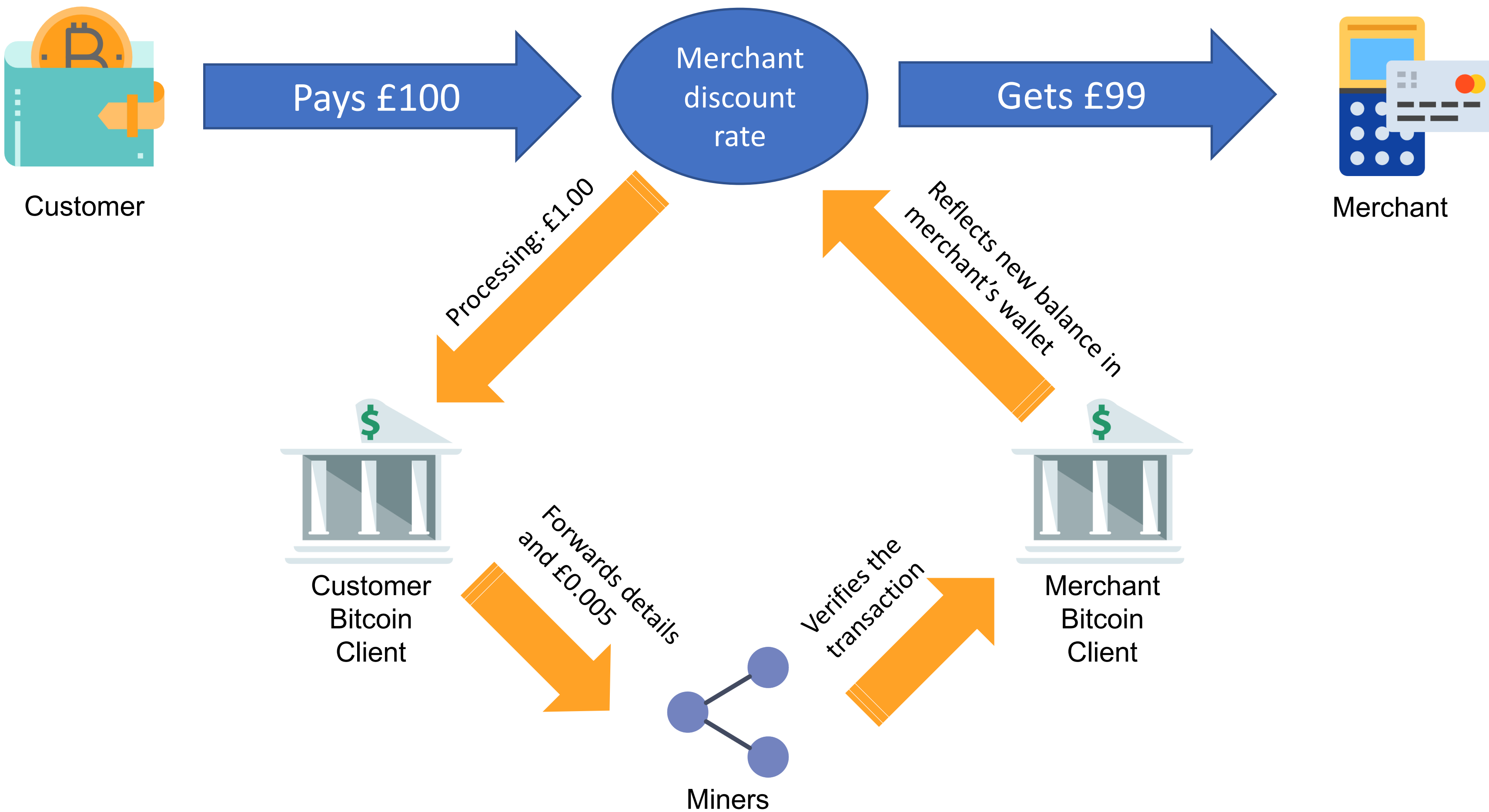
Source: *Goldman Sachs Global Investment Research : The Future of Finance Part 2*

Note: *(1) Charts based on survey conducted with the Electronic Transactions Association*

Annex: Bitcoin, the payment solution of the future? (2/2)

The Bitcoin payment ecosystem

What are the fees and the actors when you use Bitcoin?



Source: *Goldman Sachs Global Investment Research : The Future of Finance Part 2*

Note: Icon made by [Freepik](), Becris, RoundIcons, Smashicons from [www.flaticon.com]()

Disclaimer

Disclaimer

This presentation has been prepared for informational and educational purposes only. Although the information contained in this presentation has been obtained from sources which the authors believes to be reliable, it has not been independently verified and no representation or warranty, express or implied, is made and no responsibility is or will be accepted by the authors as to or in relation to the accuracy, reliability or completeness of any such information.

Opinions expressed herein reflect the judgement of the authors as of **[May 2018]** and may be subject to change without notice if the authors become aware of any information, whether specific or general, which may have a material impact on any such opinions.

The information of this presentation is not intended as and does not constitute investment advice or legal or tax advice or an offer to sell any securities/tokens to any person or a solicitation of any person of any offer to purchase any securities/tokens.

The authors will not be responsible for any consequences resulting from the use of this presentation as well as the reliance upon any opinion or statement contained herein or for any omission.

This presentation is confidential and may not be reproduced (in whole or in part) nor summarised or distributed without the prior written permission of the authors.

IV. Appendix

Bibliography (1/5)

Main sources

"La technologie RFID/NFC", Samia Bouzefrane

https://cedric.cnam.fr/~bouzefra/cours/CoursNFC_Bouzefrane_Decembre2013.pdf

"Le CNRFID, de l'innovation au déploiement de solutions RFID et NFC"

http://www.centrenational-rfid.com/fonctionnement-dun-systeme-rfid-article-17-fr-ruid-17.html

"Costly cash: a synthesis of international evidence on the cost of making payments", DotEcon

http://www.dotecon.com/assets/images/DotEcon-Costly-cash-report-v5-3.pdf

"The Evolution of Payment Costs in Australia", The Reserve Bank of Australia

https://www.rba.gov.au/publications/rdp/2014/pdf/rdp2014-14.pdf

"The Truth about Contactless Payments", Center for International Safety

http://www.canberra.edu.au/cis/storage/Contactless%20payments.pdf

"Contactless EMV Payments: Benefits for Consumers, Merchants and Issuers", Smart Card Alliance

http://www.emv-connection.com/downloads/2016/06/Contactless-2-0-WP-FINAL-June-2016.pdf

"The Future of Finance Part 2: redefining "The Way we Pay" in the next decade", Goldman Sachs

Bibliography (2/5)

Secondary Sources

"How to read a contactless credit card such as Visa paywave or MasterCard paypass", NFC Admin

http://www.nfc.cc/2012/04/02/android-app-reads-paypass-and-paywave-creditcards/

"MasterCard: Contactless cards can deliver 30% spending lift"

https://www.nfcworld.com/2012/05/04/315479/mastercard-contactless-cards-can-deliver-30-percent-spending-lift/

"New MasterCard Advisors Study on Contactless Payments Shows Almost 30% Lift in Total Spend Within First Year of Adoption", MasterCard

https://newsroom.mastercard.com/press-releases/new-mastercard-advisors-study-on-contactless-payments-shows-almost-30-lift-in-total-spend-within-first-year-of-adoption/

"Millions of Barclays card users exposed to fraud", Benjamin Cohen

https://www.channel4.com/news/millions-of-barclays-card-users-exposed-to-fraud

"Consumers turn to contactless as usage surges", The UK Cards Association

http://www.theukcardsassociation.org.uk/wm_documents/05022015%20Contactless%20spending%202014%20-%20FINAL.pdf

"Are contactless payments safe?", Vladislav Biryukov

https://www.kaspersky.com/blog/contactless-payments-security/9422/

"Contactless payment card theft: How is the data stolen – and what can I do to protect myself?", James Rush

https://www.independent.co.uk/news-14-1/contactless-payment-card-theft-how-is-the-data-stolen-and-what-can-i-do-to-protect-myself-10409319.html

Bibliography (3/5)

Secondary Sources

"Contactless credit cards: Convenience versus security?", Heng Loong Cheong, Edward Chaterton, Louise Crawford

https://www.dlapiper.com/en/singapore/insights/publications/2015/10/contactless-credit-cards/

"Why contactless pickpocketing is impossible?", Gemalto

https://www.gemalto.com/brochures-site/download-site/Documents/documentgating/fs-why-contactless-pickpocketing-impossible.pdf?webSyncID=079d946a-00b7-01bb-740d-9c2445a59040&sessionGUID=64b82cf1-9134-4e28-890b-666ea54cd4a9

"Necessity is the mother of invention: Mobile innovation in financial inclusion", Napala Pratini

https://fin.plaid.com/articles/mobile-innovation-financial-inclusion

"La norme EMV", Samia Bouzefrane

https://cedric.cnam.fr/~bouzefra/cours/Cartes_Bouzefrane_EMV_nov2009.pdf

"EMV", Wikipedia

https://en.wikipedia.org/wiki/EMV#North_America

"How does RFID reader reads a passive tag?", Aniruddha Sarkar

https://electronics.stackexchange.com/questions/334419/how-does-rfid-reader-reads-a-passive-rfid-tag

"Can blockchain rescue digital advertising?", Charlie Stewart

https://www.businesslive.co.za/redzone/news-insights/2017-09-13-can-blockchain-rescue-digital-advertising/

Bibliography (4/5)

Secondary Sources

"The cost of consumer payments in sweden" Segendorf Björn, Jansson Thomas,

https://www.econstor.eu/bitstream/10419/81925/1/71906726X.pdf

"HCE and Tokenization for Payment Services", GSMA

https://www.gsma.com/digitalcommerce/wp-content/uploads/2014/11/GSMA-HCE-and-Tokenisation-for-Payment-Services-paper_WEB.pdf

"The end of cash as we know it", Laurence Dodds

https://www.telegraph.co.uk/news/shopping-and-consumer-news/11456380/The-end-of-cash-as-we-know-it.html

"Contactless Payments, A History", Alice Chen

https://www.payfirma.com/payments-101/contactless-payments-a-history/

"RFID Pickpockets – Stop'em with RFID Blocking Gear", Beth Williams

https://www.corporatetravelsafety.com/safety-tips/rfid-pickpockets/

"What is Tokenization Payments?"

http://www.contactlesspaymentcards.com/whatistokenizationpayments.php

"Mobile payment: war of the wallets", EY

http://www.ey.com/Publication/vwLUAssets/ey-mobile-payment-war-of-wallets-nov-2015/$FILE/ey-mobile-payment-war-of-wallets-nov-2015.pdf

Bibliography (5/5)

Secondary Sources

"Dirty money: There are more germs on a £1 coin than on a TOILET SEAT", Sarah Griffiths

http://www.dailymail.co.uk/sciencetech/article-2621500/Dirty-cash-Bank-notes-contain-26-000-bacteria-half-Britons-wash-hands-handling-them.html

"Revealed: cash eclipsed as Britain turn to digital payments", The Guardian

https://www.theguardian.com/money/2018/feb/19/peak-cash-over-uk-rise-of-debit-cards-unbanked-contactless-payments

"Contactless payment in the United Kingdom", Statista

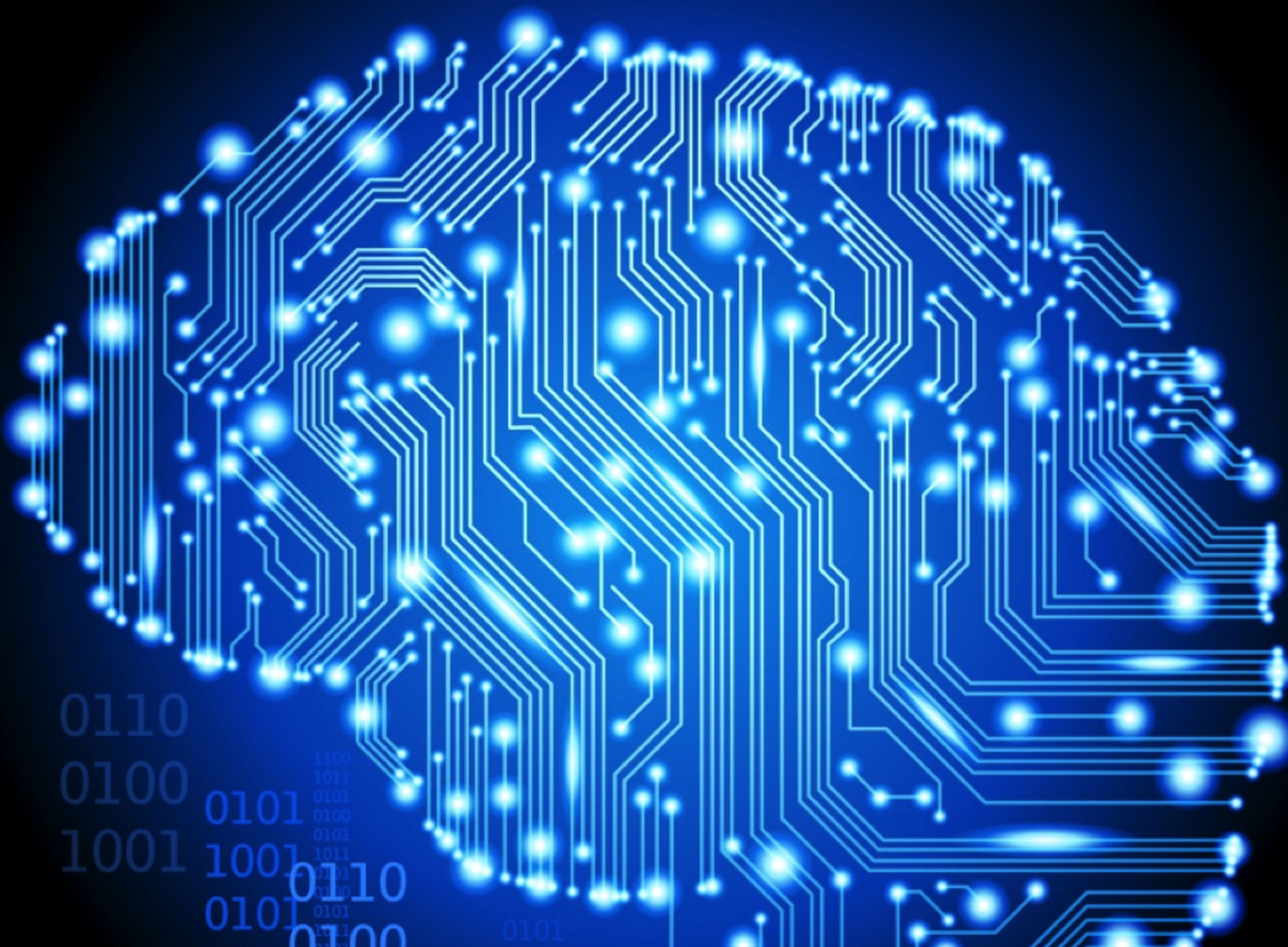
Visuals

Picture on the first page is from: "Payment solutions for businesses in the US", DelawareAgency

https://delawareagency.com/payment-solutions-for-businesses-in-the-u-s/

All logos are from Flat Icon

https://www.flaticon.com



Machine Learning impacts in the Lending Space

Master Thesis 2018

Louis Marty, Damien Mossuz, Matteo Screnci

HEC PARIS

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	Overview of the Fintech activities in the Lending space
II.	The traditional lending market : a focus on credit-scoring for individuals
III.	What are the main innovation revolutionizing the credit-scoring activities?
1.	Access to alternative data
2.	Application of machine learning algorithm to credit-scoring
V.	What does it mean for individuals?

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	**Overview of the Fintech activities in the Lending space**
II.	**The traditional lending market : a focus on credit-scoring for individuals**
III.	**What are the main innovation revolutionizing the credit-scoring activities?**

- **1.** **Access to alternative data**
- **2.** **Application of machine learning algorithm to credit-scoring**

V.	**What does it mean for individuals?**

HEC PARIS

Overview – FinTech companies in the lending space

Recent innovation has triggered a high level of activity in the lending space

World Economic Forum highlighted three main driving forces

Mass P2P Lending	Alternative Adjudication	Lean and Automated Process
"P2P services were growing quickly, reaching a significant number of customers across the globe" (1)	***"New ways to measure and track credit worthiness were being developed"*** (1)	***"Automation was transforming adjudication and loan origination"*** (1)

The FinTech Lending space has experienced a high level of activity

JAMES – The Artificial Intelligence for Credit Risk

AVANT

Funding Circle

affirm

Kabbage

zest finance

Source : Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

(1) Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

Overview – FinTech Companies in Lending Space

The World Economic Forum has identified three main disruption segments

Three main disruption segments

1.	**New adjudication techniques have significantly expanded access to credit for underbanked, "thin-file" and subprime customers**
2.	**Individual and small-business borrowers expect their lender to deliver the seamless digital origination and rapid adjudication pioneered by leading fintechs**
3.	**Non-financial platforms are emerging as an important source of underwriting data and a point of distribution for credit**

Nevertheless, FinTech faces strong competition from banks

4.	**Funding economics put marketplace lenders at a cost disadvantage compared to traditional banks, raising questions about the model's sustainability**

In the following presentation, we are going to focus on the innovation in terms of adjudication techniques (namely alternative data and modelling)

Source : Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I. **Overview of the Fintech activities in the Lending space**

II. **The traditional lending market : a focus on credit-scoring for individuals**

III. **What are the main innovation revolutionizing the credit-scoring activities?**

- **1.** **Access to alternative data**
- **2.** **Application of machine learning algorithm to credit-scoring**

V. **What does it mean for individuals?**

HEC PARIS

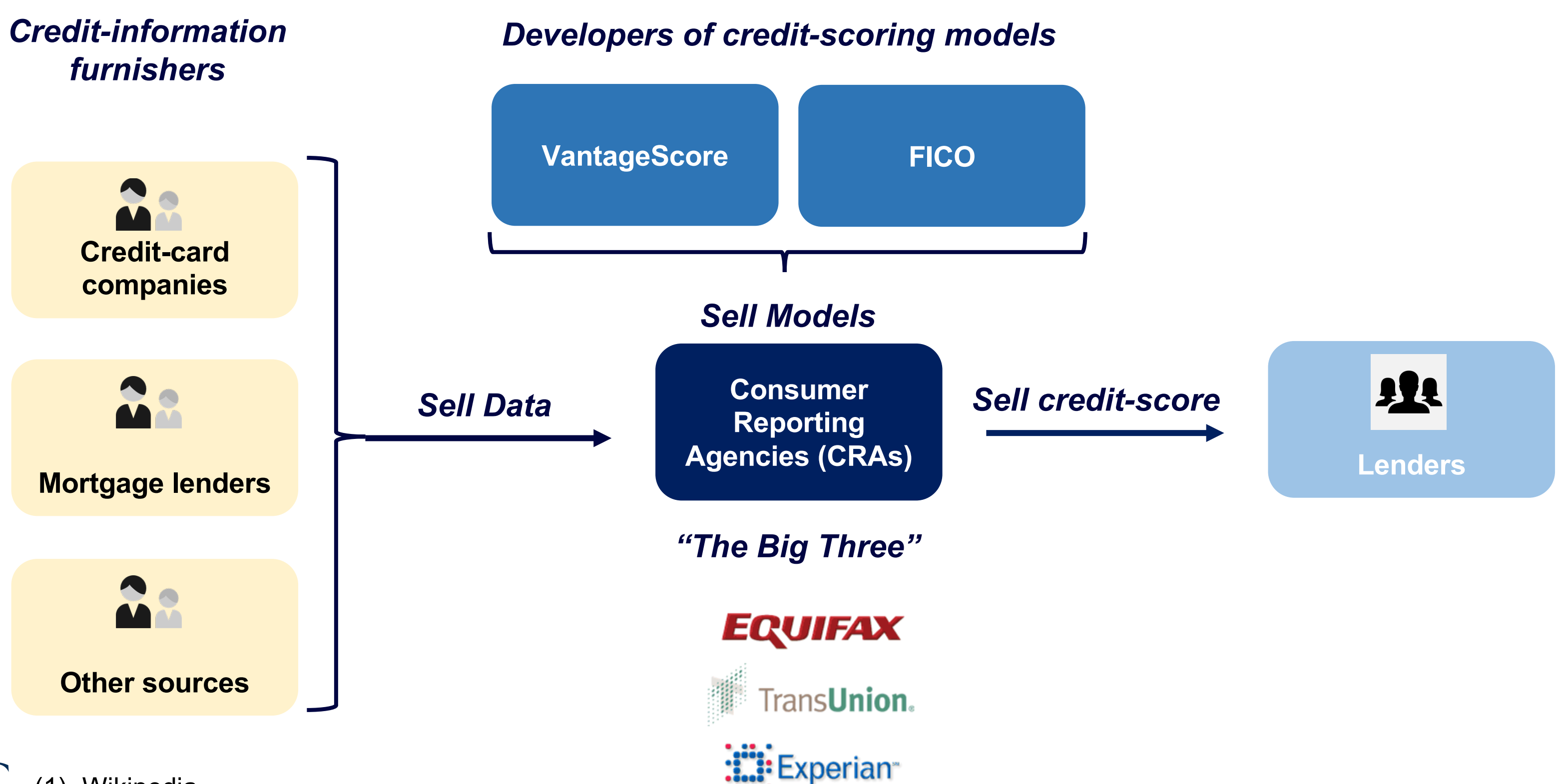
The traditional lending market : focus on credit scoring

So far credit-scoring have been dominated by the “Big Three”

What is a credit-score?

“A credit score is a “summary of a person’s apparent creditworthiness that is used to make underwriting decisions” as well as to “predict the relative likelihood of a negative event, such as a default on a credit obligation”(1)

The traditional value-chain in credit-scoring



(1) Wikipedia

Source : https://cache-content.credit.com/wp-content/uploads/2014/04/3_Major_Credit_Reporting_Bureaus_Experian_Equifax_Transunion_Logos.png

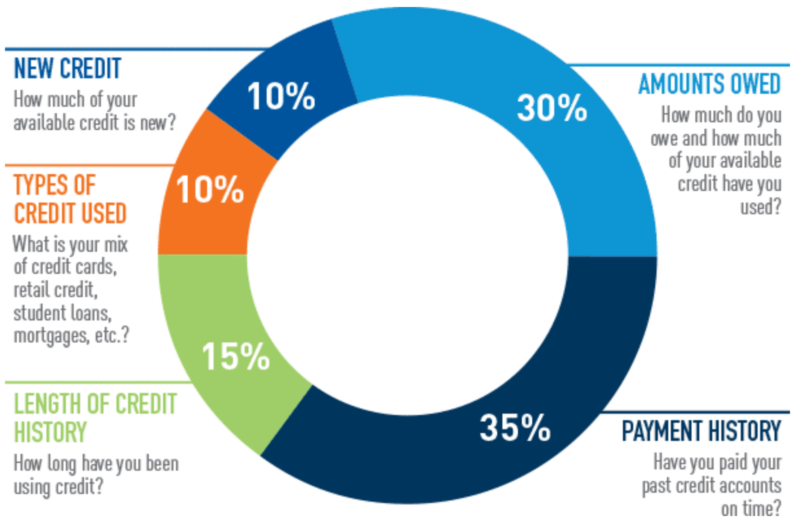
The traditional lending market : focus on credit scoring

What are the current methods to assess your credit-score?

FICO score is currently the reference in the industry

- FICO score has been founded by Fair Isaa Corporation and is today one of the most wildly used credit score in the USA
- Individuals are **scored between 300 and 850**; the higher your score the lower your risk. Therefore, your score has a direct impact on your ability to substract a loan and on the rate at which you are able to borrow
- Though the exact formula remains a trade secret, is it is **basically an average of several data input** (cf. below) with predefined weight for each
- FICO score is basically saying **"You should pay your bills on time and not incur too much debt; the rest are details."**

Five main data are inputed in FICO model …



… resulting in a score assessing creditworthiness

CREDIT SCORE CATEGORY

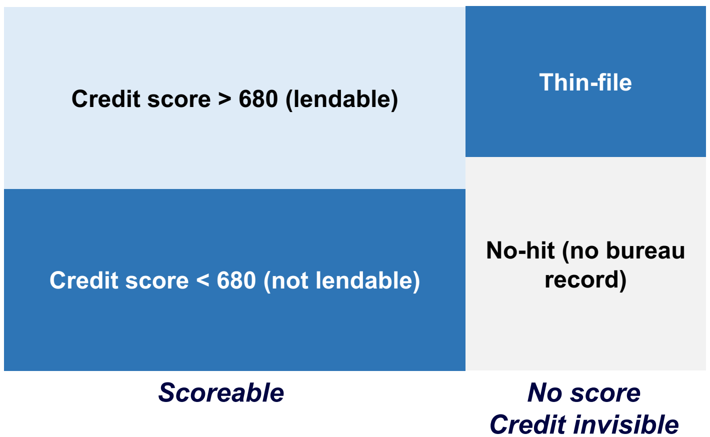
CATEGORY	RANGE
Excellent	750 & Above
Good	700 - 749
Fair	650 - 699
Poor	550 - 649
Bad	550 & Below

Source : Investopedia, https://www.creditsesame.com/blog/credit/credit-score-range-for-experian-transunion-equifax/

The traditional lending market

Flaws in the credit score prevent roughly 50m people from credit access

A large part of the US population is "unscorable" resulting in its exclusion from the credit market



- **Full-file (180–190 million)**: Borrower has a credit file with sufficient recent tradeline data to generate a traditional credit score
- **Thin-file (25–35 million)**: Borrower has credit file but with insufficient and/or outdated tradeline data to generate a traditional credit score
- **No-hit (20–25 million)**: Credit bureau has no information/file on the person at all

A system in closed-loop

- Variables **focus on credit history rather than directly on the capability to pay back debt** : for instance employment history or asset owned is not taken into account in FICO score
- Therefore, critics have risen highlighting that current score results in an exclusion from the credit access based on the previous exclusion to the credit access

Credit reports data is affected by inaccuracy

- In the US, **26% of customers had inaccuracies in their credit report in 2013**
- *Example : "TransUnion repeatedly reported the bad debts of "Judit L. Uption' on the report corresponding to "Judy Thomas*"

Source : Oliver Wyman point of View : Alternative Data and the Unbanked. Peter Carroll, Saba Rehmani-

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	Overview of the Fintech activities in the Lending space
II.	The traditional lending market : a focus on credit-scoring for individuals
III.	What are the main innovation revolutionizing the credit-scoring activities?
1.	Access to alternative data
2.	Application of machine learning algorithm to credit-scoring
V.	What does it mean for individuals?

HEC PARIS

Innovation in the Lending Space

Innovation is driven by both access to alternative data and new models

Three main sources of innovation

New sources of data	New sources of data have emerged for use in adjudicating credit, such as social and mobile data for individuals, and payments or accounting data for businesses,. While this data has had limited effectiveness in improving the underwriting of established customers, it has proven to be valuable for "thin-file" borrowers (with insufficient credit bureau history) and small businesses
Using Data more Effectively	Incumbent lenders are looking to their existing stores of data to bolster their underwriting models, especially for underbanked customers? However, that data is often unstructured and siloed, making it difficult to be put to use. To address these challenges, incumbents are investing heavily in data transformation, automation and new analytics
More Agile Credit Models	New entrants improve on their credit models using short iteration cycles, while incumbents are constrained to making adjustments much more slowly. This lag in implementing best-in-class methodologies provides new entrants a temporary competitive advantage in understanding the credit risk of underbanked and "thin-file" customers, especially as new sources of data become available

Caveats

Lack of Credit Cycles	While credit models have improved since the financial crisis, many alternatives approaches were developed following the crisis, making it unclear how alternative models for subprime customers ill fare over the full like of the next macro-credit cycle

Source : Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

Innovation in the Lending Space

What is the modelling and scoring process for credit-scoring with ML tools?

Set creditworthiness
Definition & Identify examples

Step 1 : defining the problem and specifying the target variable

Collect Data from Disparate Source

- **Social Network data**
- **Applicant data**
- **Public records**
- **Financial data**
- **Reports from Data brokers**

Transform Data into Usable Form

Select important « Features »

Step 2 : gathering and transforming the data

Combined features into « Metavariables »

Socring using multiple models

Calculate final « Ensemble Score »

Step 3 : developing a final model through analysis of training data and feature selection

HEC PARIS

Source : Credit Scoring in the era of big data, Mikella Hurley and Julius Adebayo. Yale Journal of Law and Technology

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	Overview of the Fintech activities in the Lending space
II.	The traditional lending market : a focus on credit-scoring for individuals
III.	What are the main innovation revolutionizing the credit-scoring activities?
1.	Access to alternative data
2.	Application of machine learning algorithm to credit-scoring
V.	What does it mean for individuals?

Alternative data

FinTech companies use non-traditional data to assess the creditworthiness

Alternative data relies on five main sources

- Social network data
- Applicant data
- Public records
- Financial data
- Reports from Data brokers

Key requirements to ensure reliability of new data(1)

Coverage	*Ex : 90% US adults use a cell phone*
Specificity	*Ideally consumer-specific resources vs modelled data*
Accuracy and timeliness	*Avoid inaccuracies in data*
Predictive power	*Ensure signaling power*
Orthogonality	*Should improve predictability of the current score with trad. data*
Regulatory compliance	*E.g.Fair Credit Reporting Act, Equal Credit Opportunity Act*

Main variables integrated in new credit-score models(1)

- *Utilities (gas, water, electricity)*
- *Telecom (TV, mobile, broadband)*
- *Rent*
- *Property/asset record: including value of owned assets*
- *Public records : beyond the limited public records information already found in standard credit reports*
- *Alternative lending payments (e.G., payday, instalment loan, rent-to-own, buy-here-pay-here auto loans, auto title loans): including both on-time and derogatory payment data*
- *Demand deposit account (DDA) information : including recurring payroll deposits and payments, average balance, ect.*

Negative Data | Positive Data

HEC PARIS

Source :)Oliver Wyman point of View : Alternative Data and the Unbanked. Peter Carroll, Saba Rehmani-
(1)Oliver Wyman point of View : Alternative Data and the Unbanked. Peter Carroll, Saba Rehmani-

Alternative data

A few examples of FinTech using alternative data

LexisNexis - RiskView	Residential stability, asset ownership, life-stage analysis, property deeds and mortgages, tax records, criminal history, employment and address history, lines and judgments , ID verification, professional licensure
FICO – Expansion Score	Purchase payments plans, checking accounts, property data, public records, demand deposit account records, cell and landline utility bill information, bankruptcy, liens, judgments, membership club records, debit data, and property asset information
ZestFinance	Major bureau credit reports and thousaud of other variables including financial information, technology usge, and **how quickly a user scrolls through terms of service**
KrediTech	**Location data** (e.g. GPS), **social graphing** (likes, friends, locations, posts), **behavioral analytics** (movement and duration on a webpage), **e-commerce shopping behavior**, device data (apps installed, operating systems)
Earnest	Current job, salary, education history, balances in savings or retirement accounts, **online profile data** (e.g. LinkedIn), and credit card information
Demyst Data	Credit scores, occupation verification, fraud checks, employment stability, work history, and online social footprint

Source : Credit Scoring in the era of big data, Mikella Hurley and Julius Adebayo. Yale Journal of Law and Technology

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	**Overview of the Fintech activities in the Lending space**
II.	**The traditional lending market : a focus on credit-scoring for individuals**
III.	**What are the main innovation revolutionizing the credit-scoring activities?**
1.	**Access to alternative data**
2.	**Application of machine learning algorithm to credit-scoring**
V.	**What does it mean for individuals?**

HEC PARIS

How to assess a credit model ? (1/2)

Preliminary introduction : a credit model as a classification model

What is a credit model?

- A credit model aims at **assessing whether someone can be lent to.** Basically, a credit model aims at classifying a group of people between two class either “good” borrowers or “bad borrowers”
- A credit model is then a two-class classification model (**binary classification**) : “classify the elements of a given set into two groups on the basis of a **classification rule**”(1)
- The class prediction for each instance is often made **based on a continuous random variable X**, which is a score
- Given a **threshold parameter** T, the instance is classified as “positive“ if X > T or “negative otherwise

How to assess the results of a binary classification?

		True condition	
		True	False
Predicted condition	True	True Positive (Power)	False Positive (Type I error)
	True	False negative (Type II error)	True negative

The True Positive Rate (TPR) and False Positive Rate (FPR) as key metrics to assess a credit model

$$TPR(T) = \int_T^{\infty} f_1\,(x)dx \qquad FPR(T) = \int_T^{\infty} f_0\,(x)dx$$

$$With\ f_1\ the\ distribution\ of\ the\ signal\ in\ case\ of\ True\ condition\ andh\ f_0\ the\ in\ case\ of\ False\ condition$$

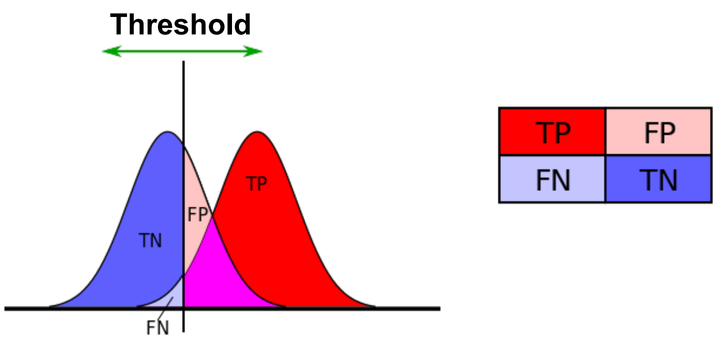
Source : Wkkipedia

(1) Wkkipedia (Binary classification)

How to assess a credit model? (2/2)

ROC curve is a way to compare the accuracy of different classification models

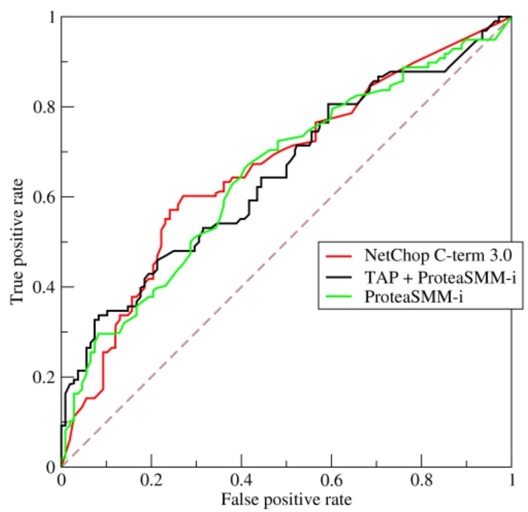
ROC Curve construction



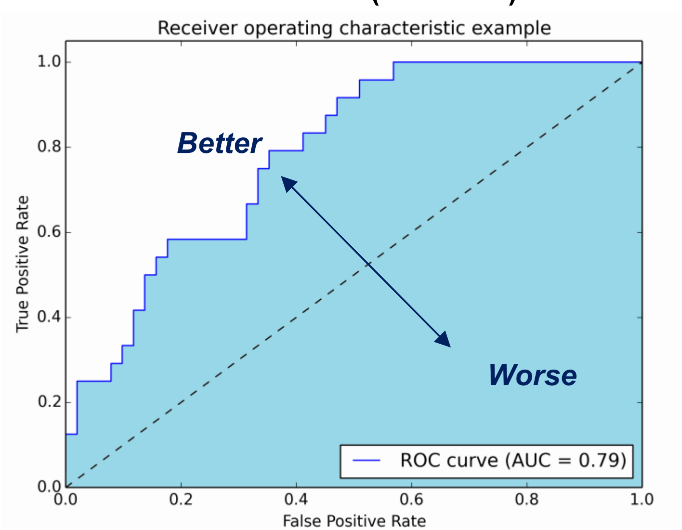
ROC Curve is used to compare models

- ROC curve is a **plot of TPR as a function of FPR with the threshold as a varying component**
- ROC curve shape is determined by how much overlap between the two distributions
- The straight line represents the ROC curve of a random predictor, thus the far from the straight line the better is the model
- From the ROC curve, one can compute the AUROC (Area Under Curve of Receiving Operating Characteristic) so as to compare models with a simple metric

ROC Curve



ROC Curve and AUROC (blue area)



Source : Wkipedia, By Sharpr - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=44059691
By BOR at the English language Wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=10714489 ;
https://stats.stackexchange.com/questions/132777/what-does-auc-stand-for-and-what-is-it ;

Binary logistic regression (1/2)

Basic model presentation

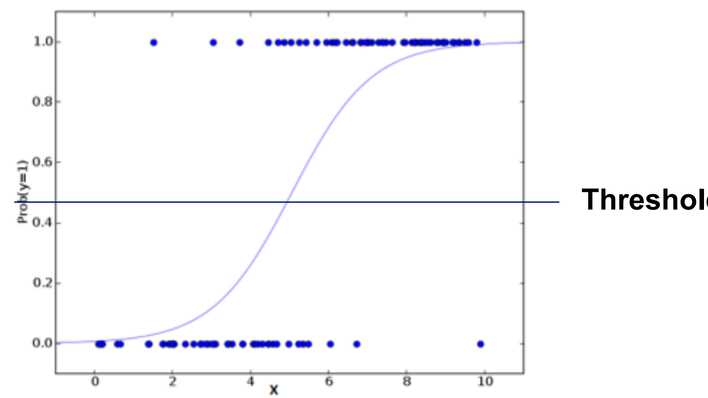
Basic introduction to the binary logistic regression

- **Model** : the probablity of Y being egual to 1 is assume to be distributed as a logistic law :

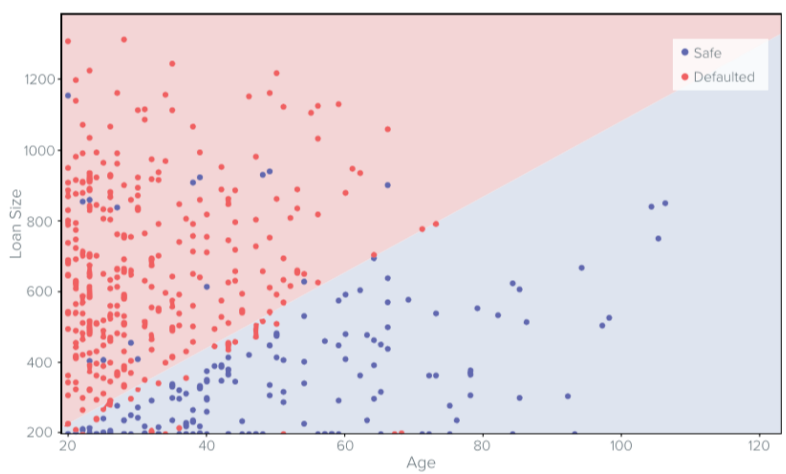
$$P(Y = 1 \mid X) = \frac{1}{1 + exp^{-\beta^T X}}$$

- **Parameters** : β is optimized based on a defined cost function
- **Decision boundary :** Once you have found the optimal parameters, you need to define a decision boundary to classify your data as either « good » or « bad : if p > 0,5 then good if p < 0,5 then bad

Plot of a logistic distribution



An example of a logistic regression with credit data by James



James Example :

- Variables : Loan size and Age

$$P(Y = 1 \mid X) = \frac{1}{1 + exp^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

- Decision boundary : $P(Y = 1 \mid X) = 0,5$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

Color dot refers to True condition vs background color to predicted condition

HEC PARIS

Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

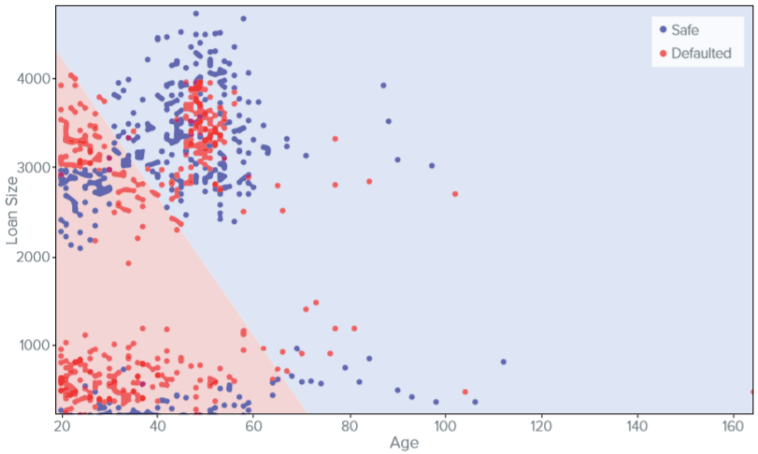
Binary logistic regression (2/2)

Conclusion & Limitations

Advantages	**Disadvantages**
• Only for linear relations • **Easy to understand and to interpretate**	• Does not manage non-linear relations • Issues in case of high correlation between variables

Logistic regression does not handle multicollinearity in the data by James

Model results plot (loan size in fonction of age)



- Logistic regression does not handle properly multicollinearity among variables leading in « unrealiable and unstable estimates of regression coefficients »

Color dot refers to True condition vs background color to predicted condition

Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Decision tree algorithm (1/6)

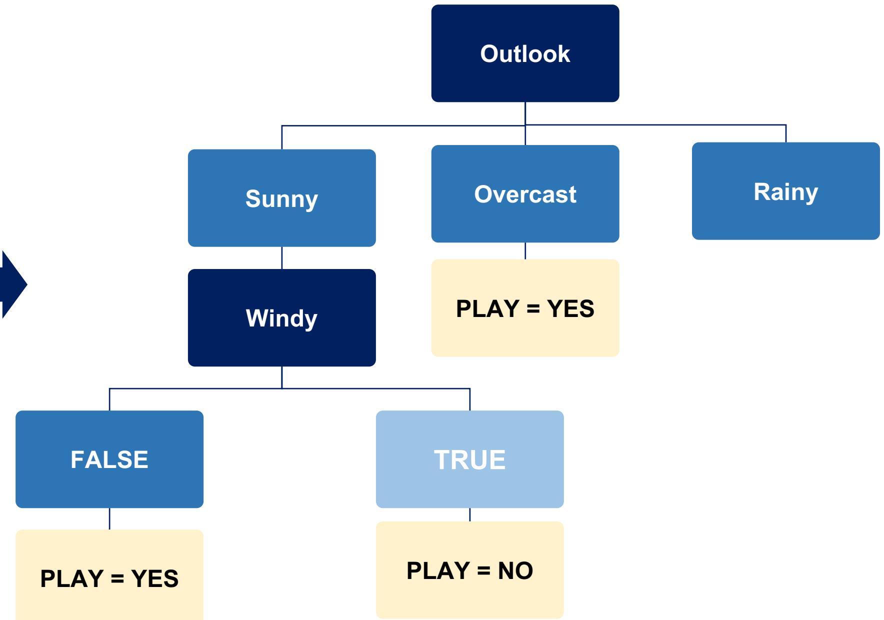
Introduction to decision tree algorithm

Decision tree is a non-parametric model aiming at predicting a target variable from defined predictors

« Goal of a decision tree is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the historical data »

Predictors | **Target**

Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
www.saedsayad.com/decision_tree.htm

Decision tree algorithm (2/6)

The decision rule is based on entropy and information gain metrics

Entropy is used to measure homogeneity of a sample

- Entropy using the frequency table of one attribute

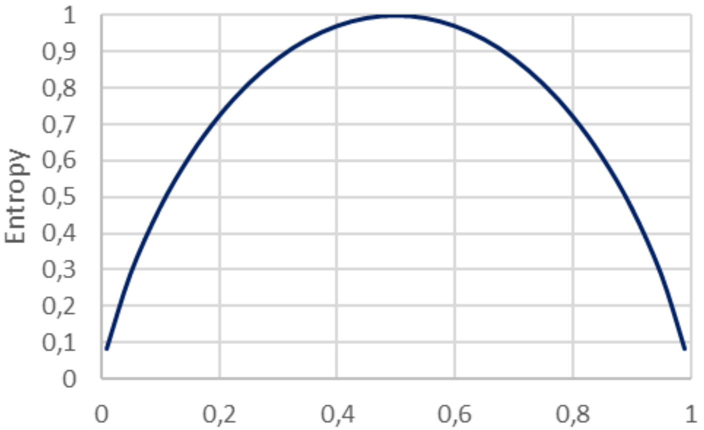
$$E(S) = \sum_{i}^{c} - p_i log_2 p_i$$

- Entropy using the frequency table of two attributes

$$E(T,X) = \sum_{c \in X}^{c} P(c)E(C)$$

Example : entropy curve of one attribute

$$Entropy = -plog_2p \ - q\ log_2q$$



$$Gain(T,X) = Entropy(T) - Entropy\ (T,X)$$

Information gain will be used as the decision rule to split the database

Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
www.saedsayad.com/decision_tree.htm

Decision tree algorithm (3/6)

A step-by-step example (1/3)

Step 1 : Calculate the entropy of the target

Play Golf	
Yes	No
9	5

$$Entropy(PlayGolf) = Entropy(5{,}9)$$
$$= Entropy(0{,}36{,}0{,}64)$$
$$= -0{,}36 log_2 0{,}36 \ -0{,}64 log_2 0{,}64$$
$$= 0{,}94$$

Step 2 : Split the dataset by attribute and calculate the entropy and information gain for each branch

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

$$E(PlayGolf, Outlook)$$
$$= P(Sunny) * E(3{,}2) + P(Overcast) * E(4{,}0) + P(Rainy) * E(2{,}3)$$
$$= \left(\frac{5}{14}\right) * 0{,}971 + \left(\frac{4}{14}\right) * 0{,}0 + \left(\frac{5}{14}\right) * 0{,}971$$
$$= 0{,}693$$

$$G(PlayGolf, Outlook) = E(PlayGOlf) - E(PlayGolf, Outlook) = 0{,}94 \ - 0{,}693 = 0{,}247$$

$$G(PlayGolf, Humidity) = \ldots = 0{,}152$$

$$G(PlayGolf, Windy) = \ldots = 0{,}048$$

$$G(PlayGolf, Temp) = \ldots = 0{,}029$$

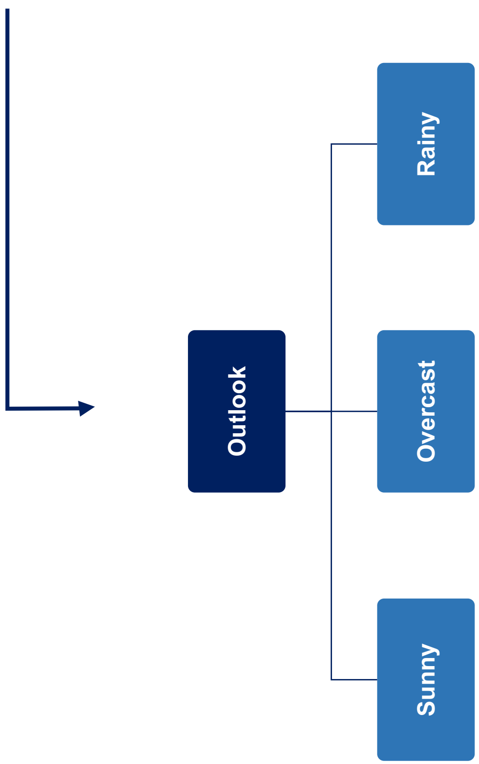
Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
www.saedsayad.com/decision_tree.htm

Decision tree algorithm (4/6)

A step-by-step example (2/3)

Step 3 : Choose the attribute with the largest information gain as the decision node and divide the dataset by branch. Repeat the process.

	Outlook	Temp.	Humidity	Windy
Inf. Gain	0,247	0,152	0,048	0,029



Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

Outlook	Temp.	Humidity	Windy	Play Golf
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

Outlook	Temp.	Humidity	Windy	Play Golf
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No

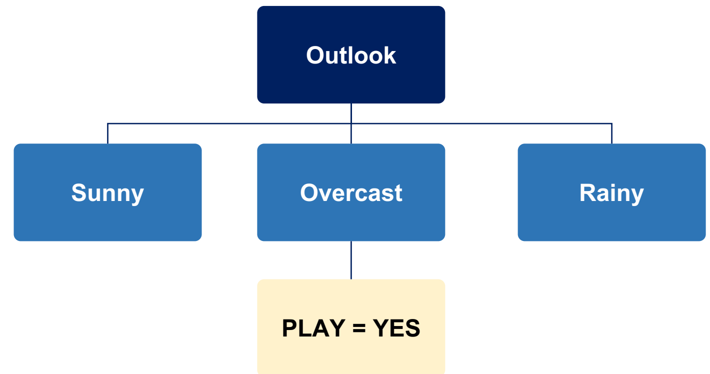
Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
www.saedsayad.com/decision_tree.htm

Decision tree algorithm (5/6)

A step-by-step example (3/3)

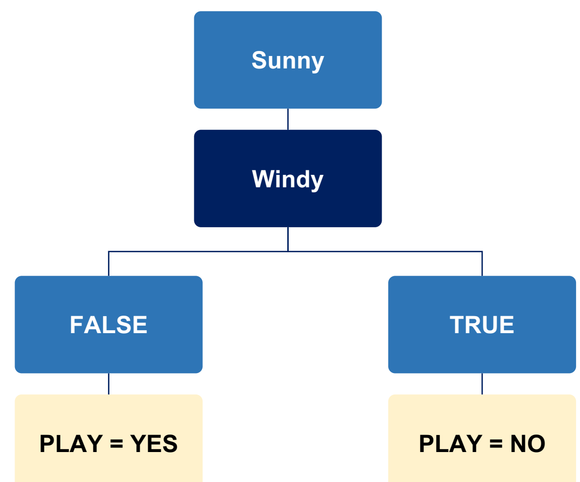
Step 4a : A branch with entropy of 0 is a leaf node

Outlook	Temp.	Humidity	Windy	Play Golf
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes



Step 4b : A branch with entropy >0 needs further splitting (go back to step 1 for the daughter node)

Outlook	Temp.	Humidity	Windy	Play Golf
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No



Source :www.saedsayad.com/decision_tree.htm
Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Decision tree algorithm (6/6)

Conclusion & Limitations

Advantages

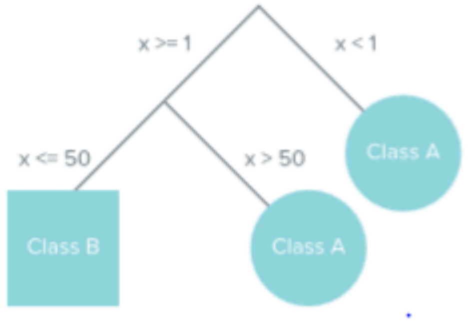
- Handle non linear relationship
- Manage properly outliers

Disadvantages

- Risk of overfitting : low predictive performance

Model results with credit data by James

Decision tree model prediction with two variables (Loan Size and Age)



Color dot refers to True condition vs background color to predicted condition

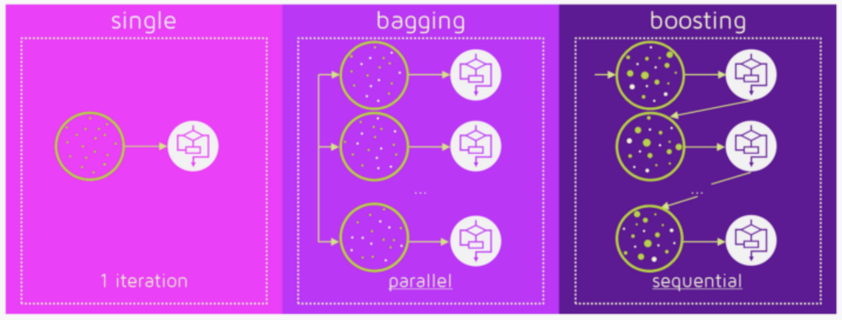
Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
www.saedsayad.com/decision_tree.htm

Introduction to ensemble models

Bagging vs Boosting

Ensemble model : bagging versus boosting

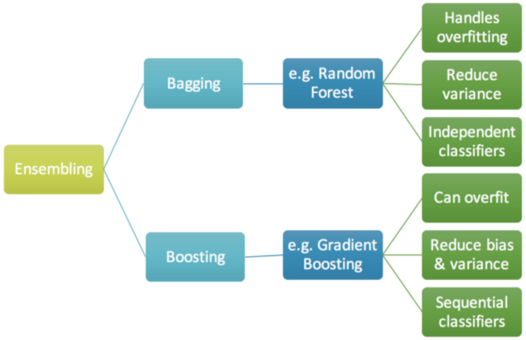
"In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone"[(1)]



Ensemble model : bagging versus boosting

There are two main types of algorithm among ensemble techniques :

- **"Bagging** is a simple ensembling technique in which we build many *independent* predictors/models/learners and combine them using some model averaging techniques. (e.g. weighted average, majority vote or normal average)**"[(2)]**
- **"Boosting** is an ensemble technique in which the predictors are not made independently, but sequentially."[(2)]



We are going to see two examples :
Random Forest ("Bagging") and Gradient boosting ("Boosting")

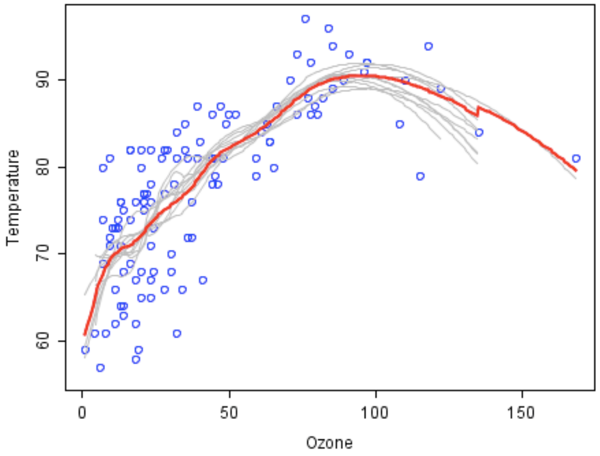
Source : [(1)] Wikipedia ; [(2)] https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d
Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Random Forest Algorithm (1/3)

Basic Model Explanation

Random Forest Algorithm is an Ensemble Method

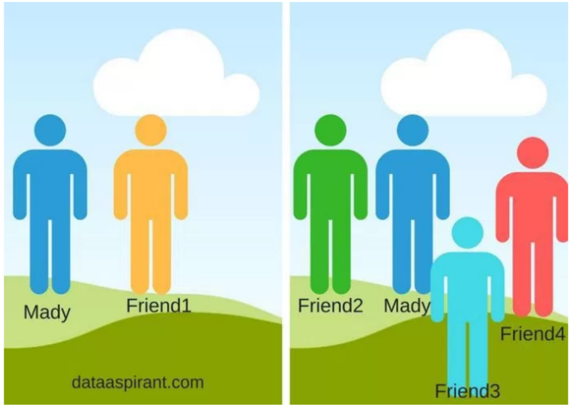
- “**Ensembles** are a **divide-and-conquer approach** used to improve performance. The main principle behind ensemble methods is that a group of “**weak learners” can come together to form a “strong learner**”. “[1]
- *Example on the right : “The data to be modeled are the blue circles. We assume that they represent some underlying function plus noise. Each individual learner is shown as a gray curve. Each gray curve (a weak learner) is a fair approximation to the underlying data. The red curve (the ensemble “strong learner”) can be seen to be a much better approximation to the underlying data.”* [1]



Basic principle of Random Forest Algorithm

“To say it in simple words: Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.” [2]

- **Real-life example** : Mady is planning a two-week trip for her holidays
  - **Decision tree** : She asks her best friend Guy for help : he asks her questions on her previous travel from which he creates a set of rules to advise her a new destination
  - **Random forest algorithm** : she asks several friends of her who asks her random questions and advise a destination. She then decides based on a **majority vote**



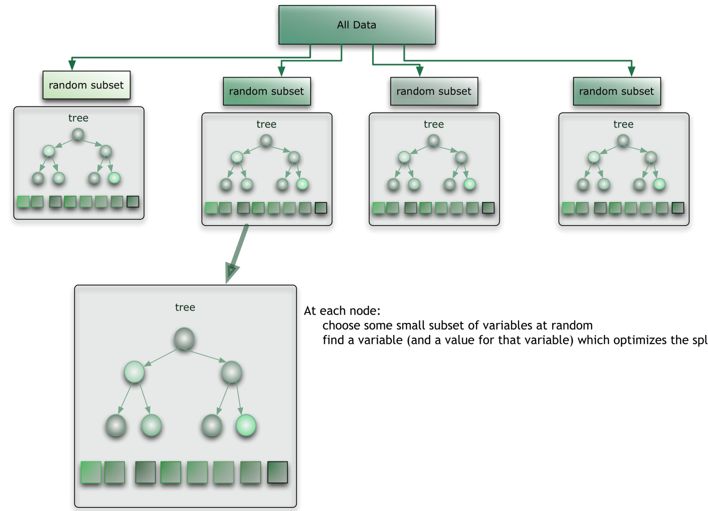
Source : [1]http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics
[2]*https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd*

Random Forest Algorithm (2/3)

A step-by-step introduction to Random Forest Algorithm

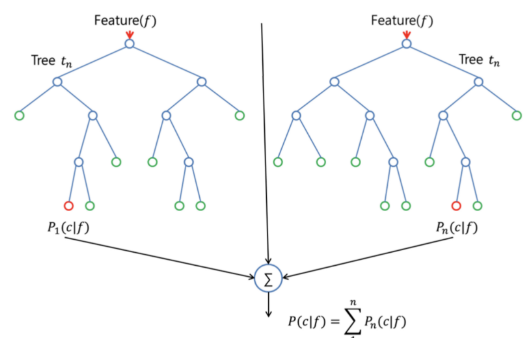
Step 1 : Build multiple decision trees to create a Random Forest

- Create random subset from the whole dataset using a Boostrap Aggregating algorithm ("**Bagging**") in order to build multiple decision trees
- For each node :
  - **Randomly select m** (<< number of predictor variables) **features**
  - Do **binary classification** using a best split approach (cf. Decision trees)
  - Repeat the same process for the next node (with a new random sample of features)



Step 2 : How to perform prediction with a Random Forest Algorithm?

- « To perform prediction using the trained random forest algorithm uses the below pseudocode.
  - Takes the test features and use the rules of each randomly created decision tree to predict the oucome and stores the predicted outcome (target)
  - Calculate the votes for each predicted target.
  - **Consider the high voted predicted target as the final prediction from the random forest algorithm**. »[1]



Source : [1] https://www.hackerearth.com/fr/practice/machine-learning/machine-learning-algorithms/tutorial-random-forest-parameter-tuning-r/tutorial/) ; http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics ; http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/

Random Forest Algorithm (3/3)

Conclusion & Limitations

Advantages

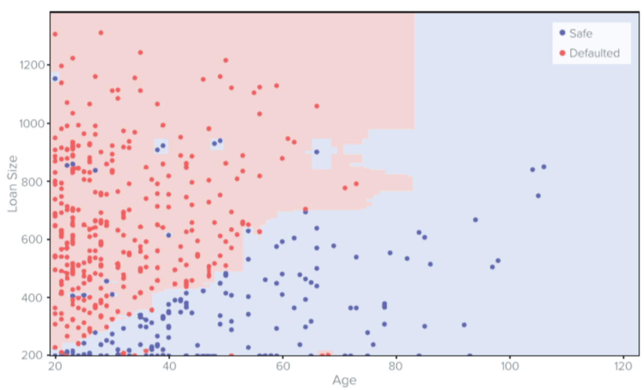
- Works for both **regression and classification**
- Handle **missing values**
- **Don't overfit** if the number of trees is sufficient
- Can **handle categorical values**

Disadvantages

- Due to large number of trees Random Forest is slow and ineffective for real-time predictions
- Trade-off betwee accuracy in prediction and speed of prediction with the number of trees
- Predictive tool not a descriptive tool : hard to understand relationship between data

Model results with credit data by James

Random Forest model prediction with two variables (Loan Size and Age)



Color dot refers to True condition vs background color to predicted condition

Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Gradient Boosting Algorithm (1/6)

What is boosting?

Basic principle of Gradient boosting

« Gradient Boosting is an ensemble technique that is rooted in the concept of Gradient descent. The latter is a first-order optimization algorithm that is usually used to calculate a function's local minimum. ***The idea of boosting came from the idea about whether a weak learner can be modified to become better****. The classifiers are* ***built in a sequential manner*** *and each member of the ensemble is an expert on the errors of its predecessors. »*[(1)]

Preliminary : Gradient Descent Algorithm

- "Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point" [(2)]

Fig. 1 : $$x_i = x_{i-1} - \eta \frac{df(x_{i-1})}{dx}$$

Fig. 2 : $$x_i = x_{i-1} - \eta \nabla F(x_{i-1})$$

Fig. 1 : One dimension

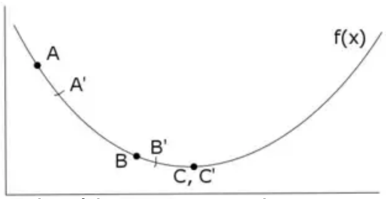
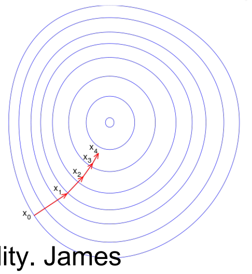


Fig. 2 : Two dimension

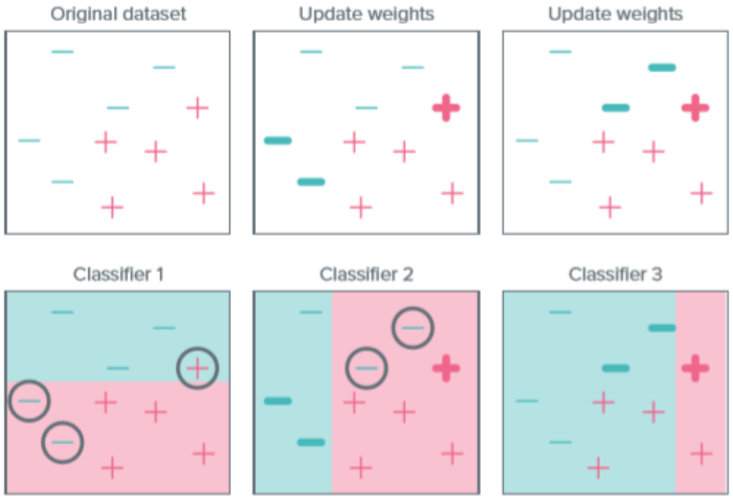


Source : [(1)] Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
[(2)] Wikipedia ; https://www.quora.com/What-is-an-intuitive-explanation-of-Gradient-Boosting

Gradient Boosting Algorithm (2/6)

A classification example : an intuitive approach

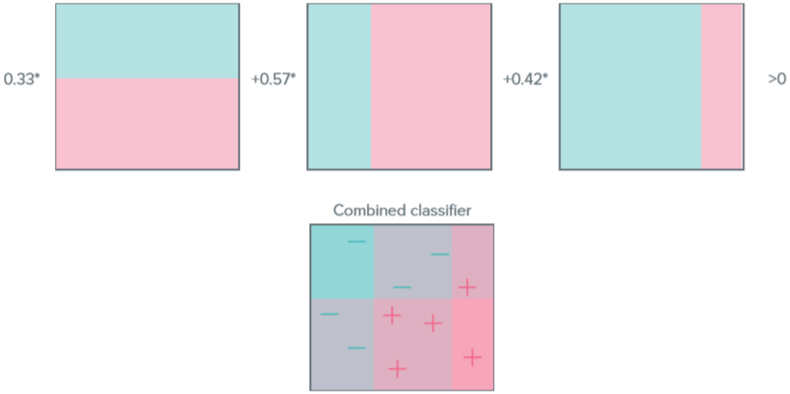
An iterative process to better classify



- Similarly, the idea is to start from a weak learner and iteratively improve it
- From the starting point, more weights is given to the errors of the first predictors, in that way the next predictors is more careful to those points

Model results by James

- Final model is constructed as the sum of the M classifiers each weighted differently
- Adaboost is for instance a very successful algortihm in **face recognition**



Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Gradient Boosting Algorithm (3/6)

A first introduction to a simplified Gradient Boosting algorithm

Step 1 : Fit a model to the data (here decision tree)

$$F_1(x) = y$$

Step 2 : Calculate error residuals

$$e_1(x) = y - F_1(x)$$

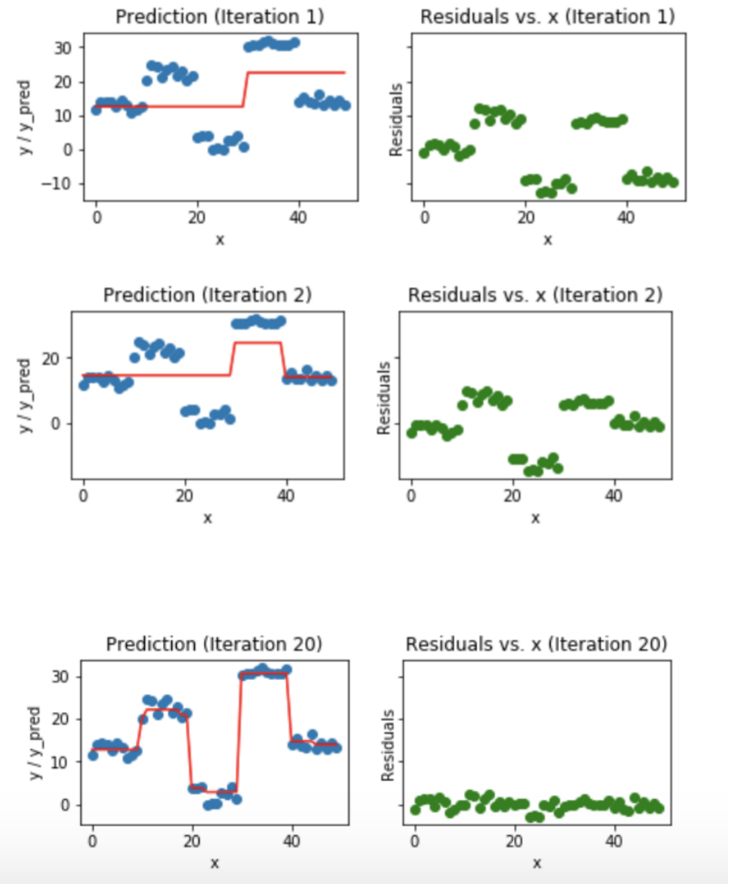
Step 3 : Fit a new model on error residuals as target variables

$$h_1(x) = y - F_1(x)$$

Step 4 : Add the predicted residuals to the previous predictions

$$F_2(x) = F_1(x) + h_1(x)$$

Step 5 : Start the process again



$$F(x) = F_1(x) \rightarrow F_2(x) = F_1(x) + h_1(x) \ldots \rightarrow F_M(x) = F_{M-1}(x) + h_{M-1}(x)$$

Source : http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/
https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d

Gradient Boosting Algorithm (4/6)

Gradient Tree Boosting : a step by step introduction

Step 1 : Initialize by fitting a model to the data based on a defined loss function (for instance square error)

$$F_0(x) = \arg\min_{\gamma} \sum_{i}^{n} L(y_i, \gamma)$$

For m = 1 to M

Step 2 : Compute pseudo residuals

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \qquad for\ i = 1\ \ldots n$$

Step 3 : Fit base learner h_m to pseudo residuals (here tree model) and compute magnitude learner γ_m

$$h_m(x) = \sum_{i=1}^{J_m} b_{jm} \mathbb{I}_{R_{jm}}(x) \qquad \gamma_m = \arg\min_{\gamma} \sum_{i}^{n} (L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Step 3 : Update prediction and go back to step one

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Source : http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/ ; https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d

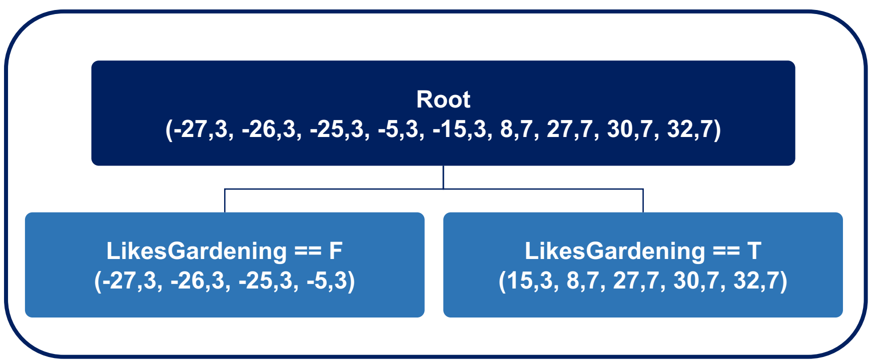
Gradient Boosting Algorithm (5/6)

A simple regression example of Gradient Tree Boosting

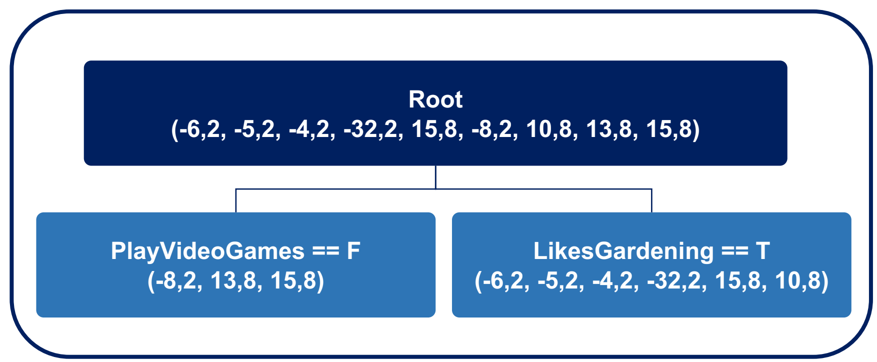
Example :
- Model : Gradient Tree Boosting
- Loss function : mean square error
- Target data : Age
- Predictors : LikesGardening (True or False), PlayVideoGames (True ou False) and LikesHats (True ou False)

Age	F0	PseudoResidual0	h0	gamma0	F1	PseudoResidual1	h1
13	40,33	-27,33	-21,08	1	19,25	-6,25	-3,567
14	40,33	-26,33	-21,08	1	19,25	-5,25	-3,567
15	40,33	-25,33	-21,08	1	19,25	-4,25	-3,567
25	40,33	-15,33	16,87	1	57,2	-32,2	-3,567
35	40,33	-5,33	-21,08	1	19,25	15,75	-3,567
49	40,33	8,67	16,87	1	57,2	-8,2	7,133
68	40,33	27,67	16,87	1	57,2	10,8	-3,567
71	40,33	30,67	16,87	1	57,2	13,8	7,133
73	40,33	32,67	16,87	1	57,2	15,8	7,133

h0



h1



Source : http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/

Gradient Boosting Algorithm (6/6)

Conclusion & Limitations

Advantages[1]

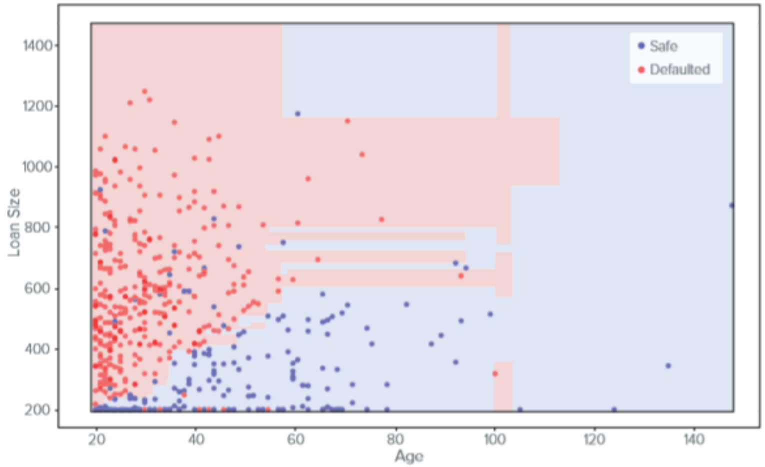
- Handle heteregeneous data very well
- Support different loss functions
- Automatically detects (non-linear) feature interactions

Disadvantages[1]

- Longer training (since it's iterative)
- Requires careful tuning
- Prone to overfitting (however there are strategies to avoid it: good tuning of parameters an d a big number of boosting stages)
- Cannot be used to extrapolate

Model results with credit data by James

Gradient Boosting model prediction with two variables (Loan Size and Age)



Color dot refers to True condition vs background color to predicted condition

Source : [1]https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd
Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

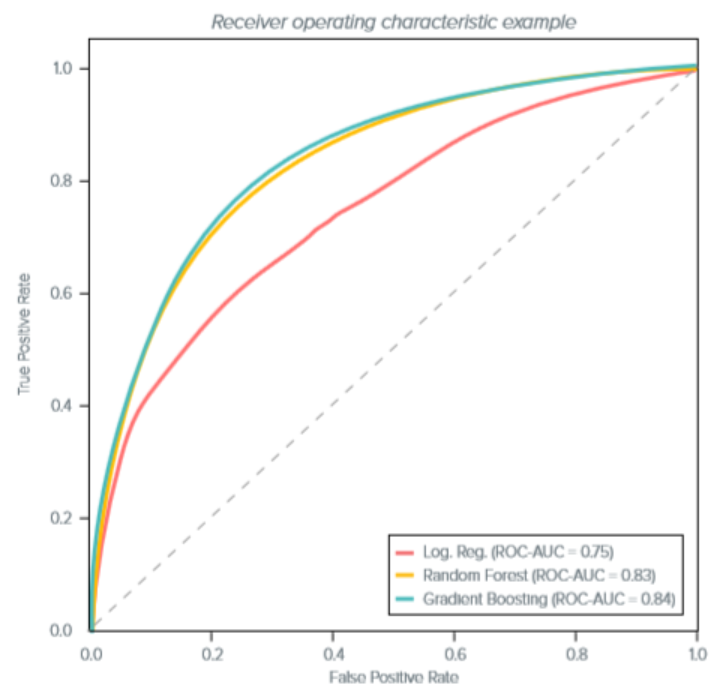
Synthesis : overview of machine learning algorithms

Fintech achieved better risk understanding through machine learning

Credit model overview

	Logistic Regression	Random Forest	Gradient Boosting
Output	PDs	PDs	PDs
Accuracy	Low	High	Very High
Base Classifier	Log. Reg.	Decision Trees	Decision Trees
Data assumptions	Linear Structure	No assumptions	No assumptions
Main advantage	Traditionally considered more intuitive	High performance with reduced overfitting	High performance with reduced overfitting
Main disadvantage	Low accuracy	Cannot be described as an equation	Cannot be described as an equation, slow to train

Quantitative model comparison



New machine learning tools have enabled FinTech in lending to better understand credit risk through a finer modelling

Source : Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James

Thesis plan

Traditional credit-scoring have been disrupted by technological innovations

I.	Overview of the Fintech activities in the Lending space
II.	The traditional lending market : a focus on credit-scoring for individuals
III.	What are the main innovation revolutionizing the credit-scoring activities?

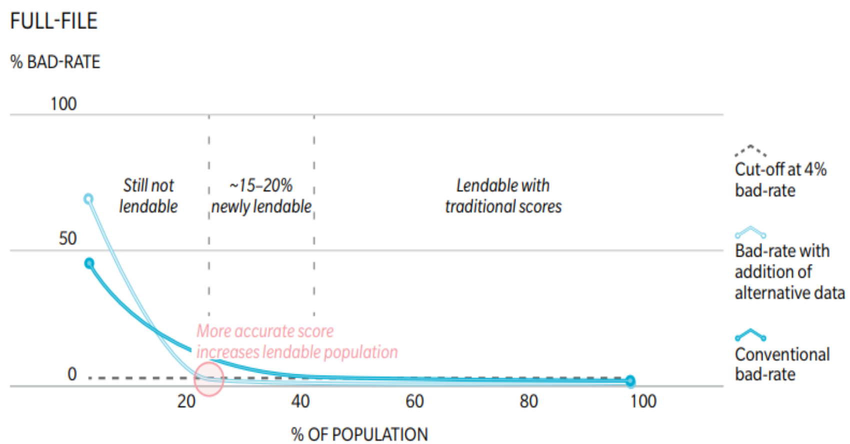
1.	Access to alternative data
2.	Application of machine learning algorithm to credit-scoring

V.	What does it mean for individuals?

Impact for the individuals

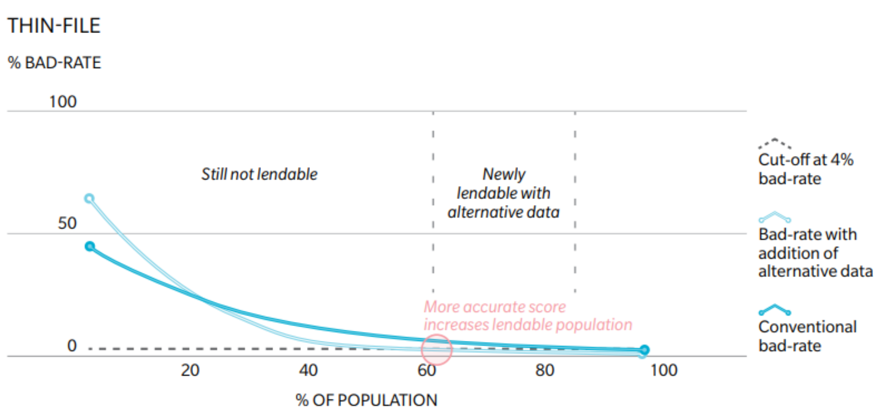
Alternative data would enable to enlarge credit access and offer lower rates

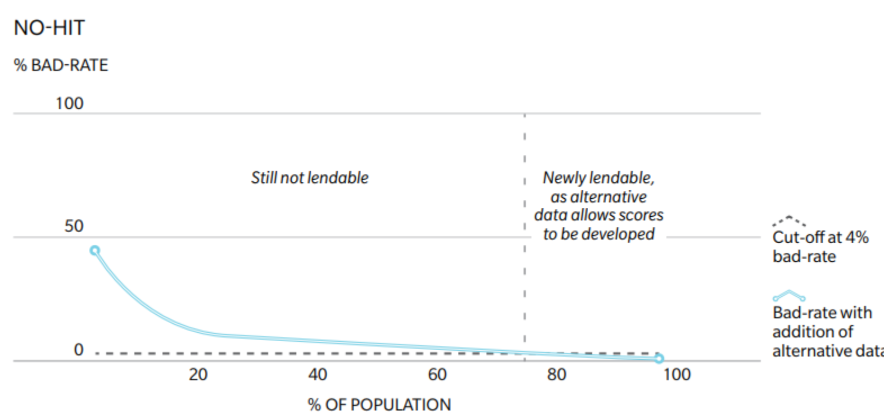
Full-file individuals would be able to borrow at a lower rate



- Alternative data and new adjudication techniques have two main impacts as the risk is better understood
  - **More individuals are granted access to loan**. It impacts no-hit and thin-file customers but also full-file individuals
  - Full-file individuals are able to **borrow at a lower rate**

Thin-file and no-hit individuals would largely benefit from new adjudication techniques and alternative data





Source : Oliver Wyman point of View : Alternative Data and the Unbanked. Peter Carroll, Saba Rehmani-

Impact for the individuals

The World Economic Forum has highlighted several practical evidence

Case studies

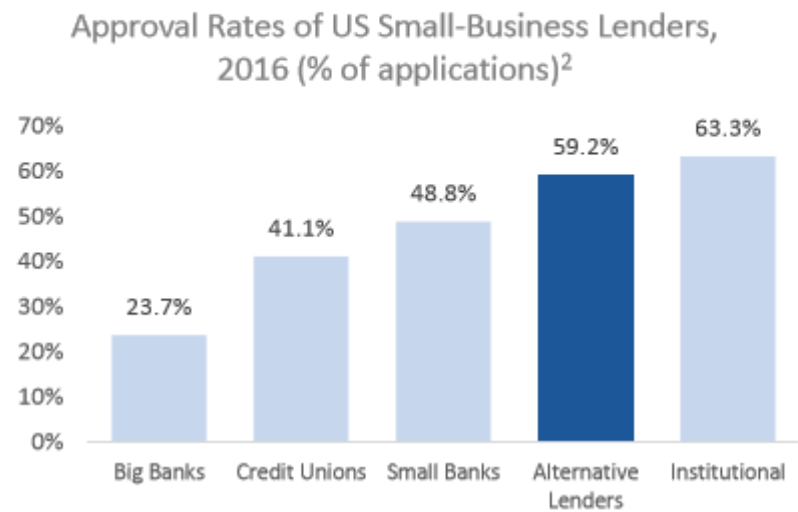
Payday loan alternative

- LendUp is a US direct online lender and financial education company
- It serves subprime customers who lost access to credit after the financial crisis
- Offers lower rates versus payday lenders and decreases rates as the borrower pays back

Alterficial intelligence for underwriting

- ZestFinance provides machine learning underwriting for financial institutions so as to improve pricing
- Investment from Baidu in 2016 : developed a credit scoring platforms for Chinese borrowers based on search data

Quantitative evidence



Increased lending to small businesses through alternative lenders

Key uncertainties

New credit adjudication techniques have proven to be effective, demonstrating strong approval and loss rates

1. How will new credit adjudication methodologies perform during a severe credit contraction?
2. What sources of data will prove to be the most valuable in credit decisions, who will own the data?
3. What techniques and sources of data will regulators deem appropriate to use?

Source : Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

Impact for the individuals

What are the main risks?

Transparency	• Secrecy around algorithms and data usage • Individuals are not able to know the impact on their actions on their credit score
Inaccuracy	• Non-traditional credit-scoring includes more data points • So far, the customers is in charge to prove inaccuracy of the data
Discrimination by proxy	• Models could used several variables as a proxy of race or religion
Could be used to target vulnerable customers	• Big data tools could be used to pursue predatory lending in order to target vulnerable individuals

HEC PARIS

Source : Credit Scoring in the era of big data, Mikella Hurley and Julius Adebayo. Yale Journal of Law and Technology

Disclaimer

Disclaimer

This presentation has been prepared for informational and educational purposes only. Although the information contained in this presentation has been obtained from sources which the authors believes to be reliable, it has not been independently verified and no representation or warranty, express or implied, is made and no responsibility is or will be accepted by the authors as to or in relation to the accuracy, reliability or completeness of any such information.

Opinions expressed herein reflect the judgement of the authors as of **[May 2018]** and may be subject to change without notice if the authors become aware of any information, whether specific or general, which may have a material impact on any such opinions.

The information of this presentation is not intended as and does not constitute investment advice or legal or tax advice or an offer to sell any securities/tokens to any person or a solicitation of any person of any offer to purchase any securities/tokens.

The authors will not be responsible for any consequences resulting from the use of this presentation as well as the reliance upon any opinion or statement contained herein or for any omission.

This presentation is confidential and may not be reproduced (in whole or in part) nor summarised or distributed without the prior written permission of the authors.

IV. Appendix

Bibliography (1/3)

Main sources :

- Oliver Wyman point of View : Alternative Data and the Unbanked. Peter Carroll, Saba Rehmani-
- Machine Learning in Credit Risk Modeling, Effeciency should not come at the expense of explainability. James
- Credit Scoring in the era of big data, Mikella Hurley and Julius Adebayo. Yale Journal of Law and Technology
- Beyond Fintech: A Pragmatic Assessment of Disruptive Potential in Financial Services by the World Economic Forum

Other sources:

- Risk and opportunities in expanding mortgage credit availability through new credit scores, Tom Parrent, George Haman, Quantilytic, LLC. Sponsored by FICO
- Consumer Credit Risk Models via Machine-Learning Algorithms, Amir Khandani, Adlar J. Kim and Andrew W. Lo
- 2017 Fintech 100, Leading Global FinTech Innovators, H2 Ventures, KPMG
- My solution to the Loan Default Prediction Competiton, Josef Feigl
- Scaling Up Affordable Lending: Inclusive Credit Scoring by Henry N and Morris, J. Responsible Finance, Oak Foundation and Coventry University
- Fintech credit : Market structure, business models and financial stability implications by the Committee on the Global Financial System and the Financial Stability Board
- Credit scoring: Case study in data analytics by Deloitte
- Credit scoring using Machine Learning Techniques by Sunil Bhatia, Pratik Sharma, Santosh Hazari, Rohit Burman, Rupali Hande

How FICO Scores Are Calculated https://www.investopedia.com/financial-edge/0212/how-is-fico-calculated.aspx#ixzz5DNvIzQgr
https://www.creditsesame.com/blog/credit/credit-score-range-for-experian-transunion-equifax/
http://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html
https://statisticalhorizons.com/multicollinearity
www.saedsayad.com/decision_tree.htm
https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

HEC PARIS

Bibliography (2/3)

Boosting algorithm :

https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/
http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf
http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/
https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d
https://www.kaggle.com/grroverpr/gradient-boosting-simplified/
https://www.quora.com/What-is-an-intuitive-explanation-of-Gradient-Boosting
https://en.wikipedia.org/wiki/Gradient_descent

https://engineeringblog.yelp.com/2018/01/building-a-distributed-ml-pipeline-part1.html
http://xgboost.readthedocs.io/en/latest/model.html
https://www.displayr.com/gradient-boosting-the-coolest-kid-on-the-machine-learning-block/
http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html
https://towardsdatascience.com/boosting-algorithm-gbm-97737c63daa3
https://datascience.stackexchange.com/questions/9134/gradient-boosting-algorithm-example
http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/
http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf
https://towardsdatascience.com/build-develop-and-deploy-a-machine-learning-model-to-predict-cars-price-using-gradient-boosting-2d4d78fddf09
https://www.analyticsvidhya.com/blog/2015/09/complete-guide-boosting-methods/
https://machinelearningmastery.com/tune-learning-rate-for-gradient-boosting-with-xgboost-in-python/

Bibliography (3/3)

Random Forest Algorithm :
https://en.wikipedia.org/wiki/Random_forest
https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674
https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd
http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/
https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#overview
http://blog.easysol.net/machine-learning-algorithms-2/
https://www.hackerearth.com/fr/practice/machine-learning/machine-learning-algorithms/tutorial-random-forest-parameter-tuning-r/tutorial/
https://www.quora.com/I-need-an-step-by-step-example-for-Random-Forests-Algorithm
https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/
http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics

Page de garde:
https://imarticus.org/wp-content/uploads/2018/01/1_5ZuLCsB1KXEPgHu-qJ8WxQ.png