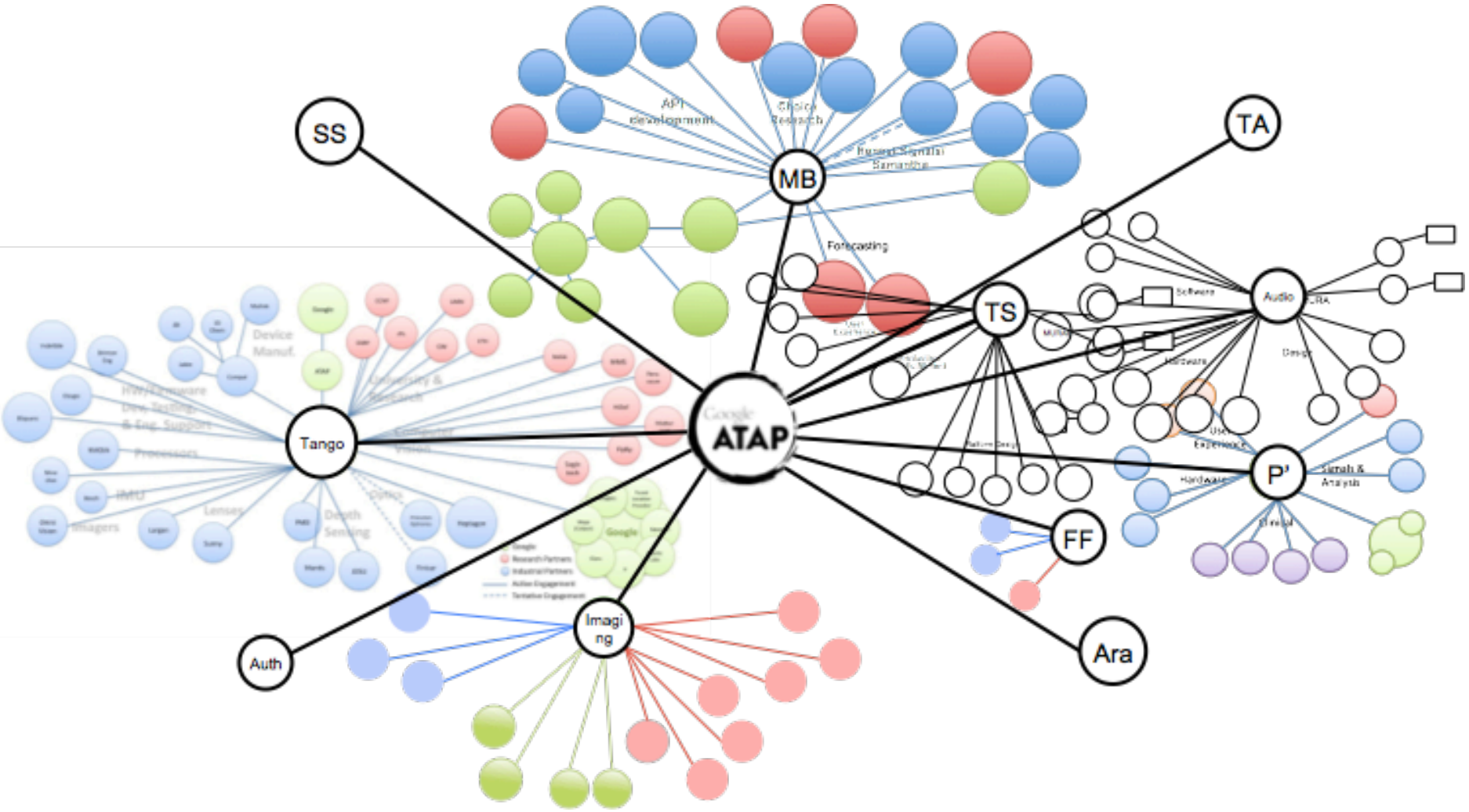
Enabling a Future of Authentication in Mobile Centric World

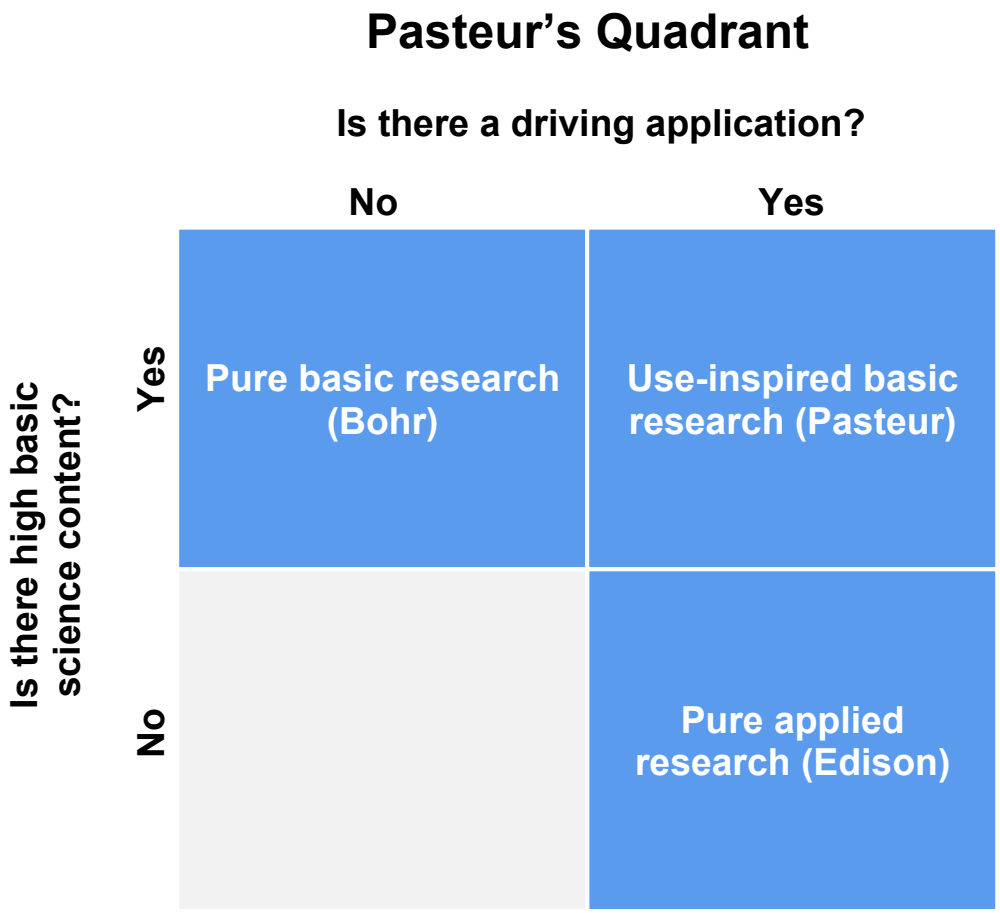
Deepak Chandra
dchandra@google.com

About ATAP



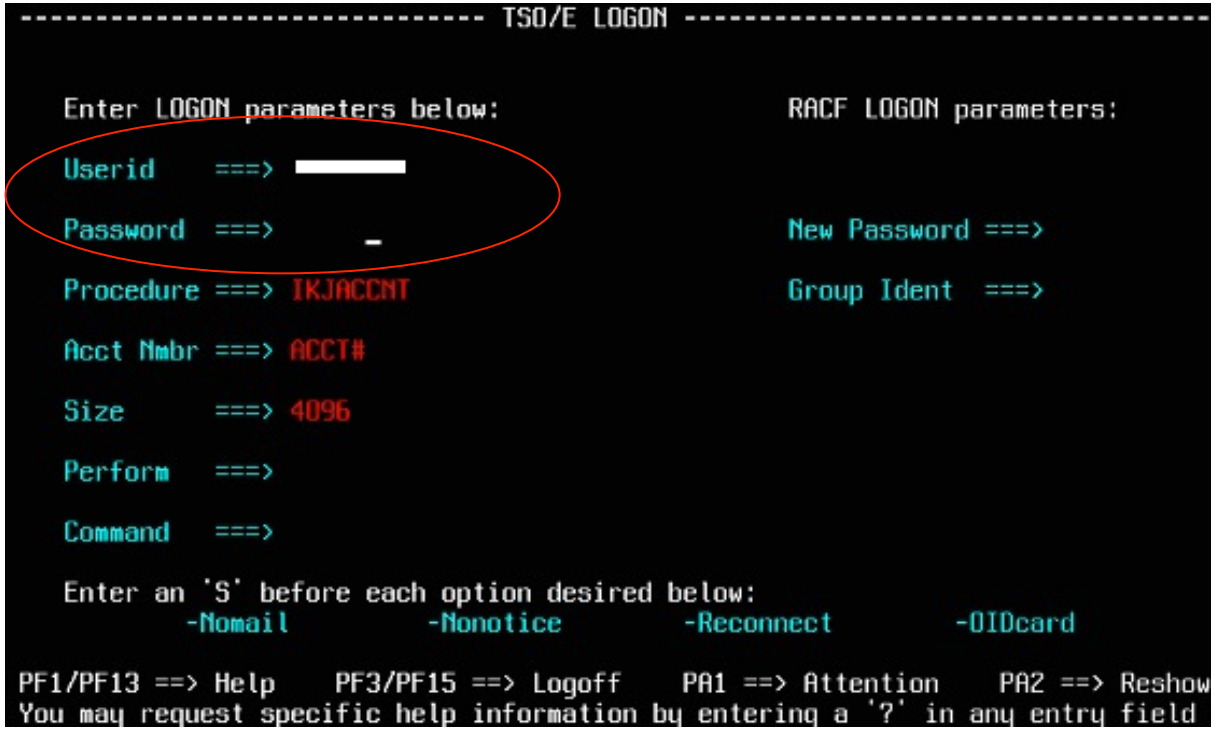
Google

Research in Pasteur's Quadrant

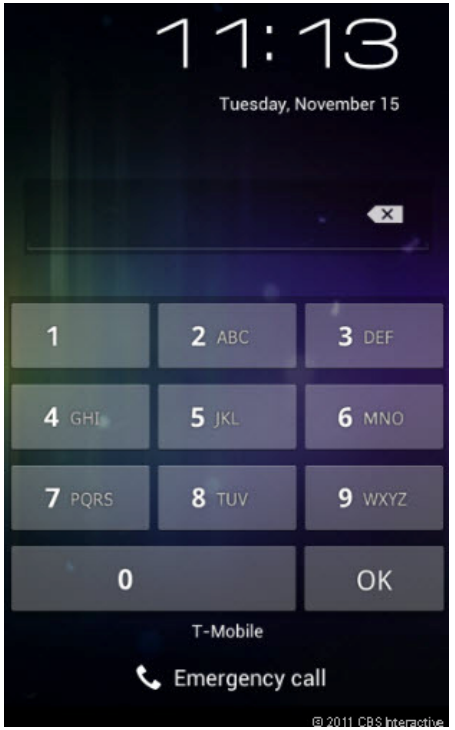


Donald E. Stokes

40 years: Authentication remained the same



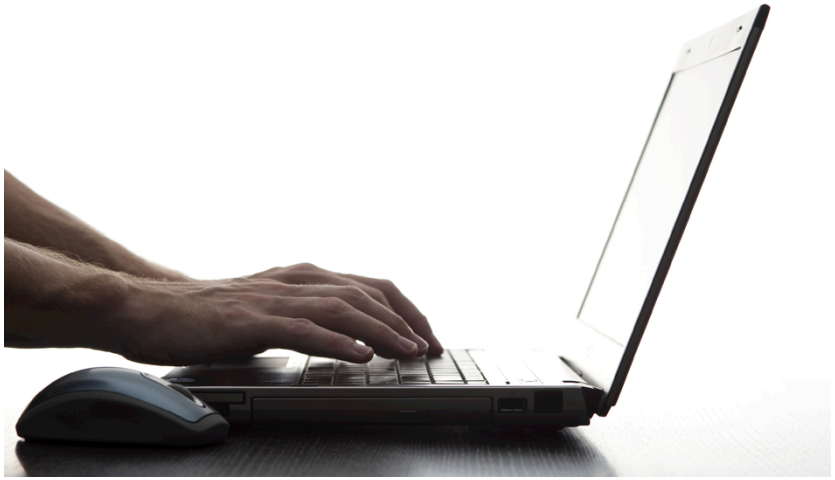
Login Screen of RACF (1976)



Login Screen
Android (2015)

Google

Password Overload



25
online passwords[2]

39
times/day[3]

70%
users forget password once a month[1]

2.3
seconds per unlock[4]

1 Symantec survey with 1028 adults in March, 2012
2 Large-scale study of web password habits: Florencio et al. Proceedings of the 16th international conference on World Wide Web

3. Motorola handset data for phones activated between 10/01 and 10/03/2012) Timeframe: Week 3 through Week 6 after activation.
4. Internal ATAP survey 2012.

Google

Authentication is too cumbersome

76%

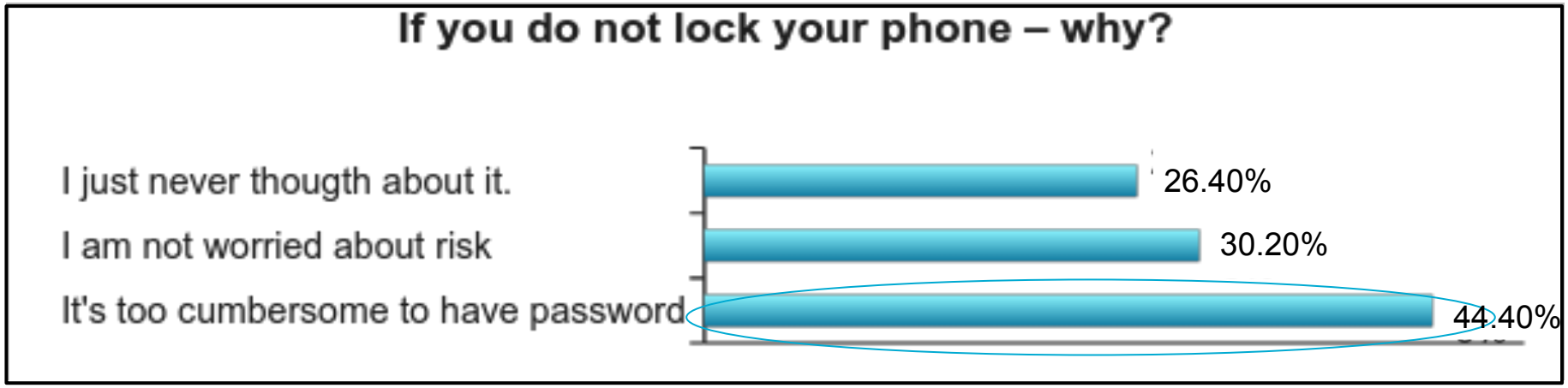
attacks due to weak passwords

2013 Data Breach Investigations Report: Verizon RISK study

53%

users do not lock their phones.

source: confident technologies, 2011



source: confident technologies, 2011

Google

How are we trying to solve it?

Complex password policies.

Longer passwords

DoD enforces a min of 15 char password

(?=.*[A-Z]) (?=.*[a-z])(?=.*[!@#$&*])(?=.*[0-9])

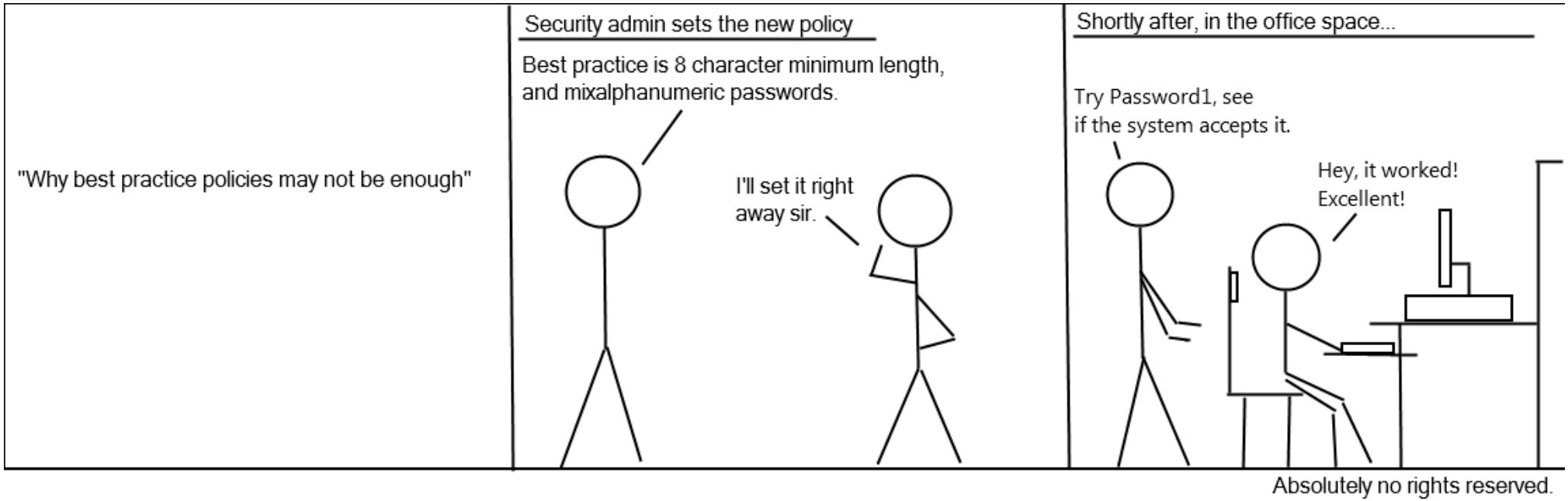
Numbers, characters, and cases. Most companies enforce subset of the rules include Google,,

Expiration

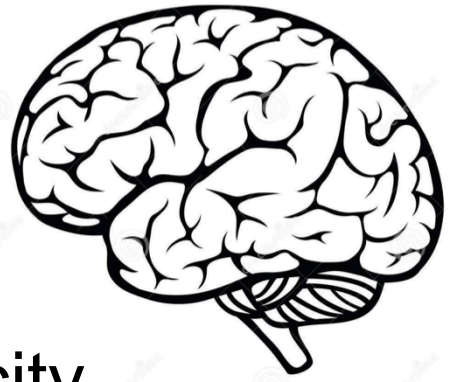
DoD expires passwords every day.

Federal Information Processing Standards Publications(FIPS) 1993 recommends the adoption of complicated passwords that include a mixture of numbers, punctuation symbols and upper and lowercase letters

…. and really not solving it



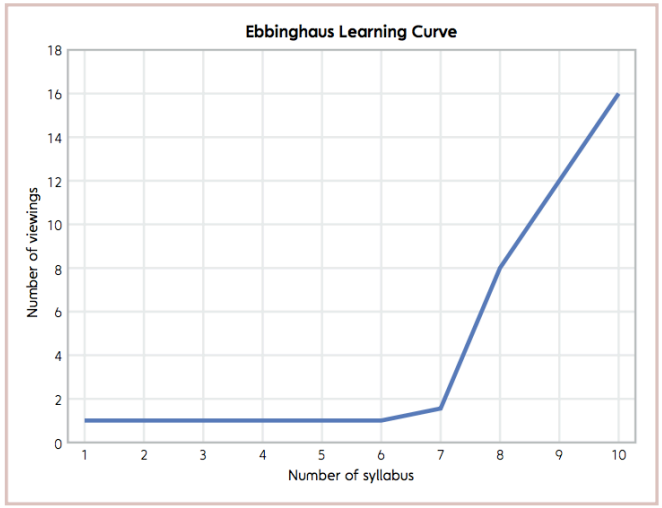
… and we are limited by memory capacity



Memory Capacity

FIGURE 5.1 The Ebbinghaus Learning Curve The graph shows how many times Ebbinghaus needed to look at lists of different lengths before he could recall them perfectly. The sharp elbow in the curve at 7 items reflects the maximum he could store in STM. This discontinuity holds for any list as long as the individual items are unrelated.

Source: Based on Ebbinghaus, 1885/1913, Figure 6



Number Seven Plus or Minus Two: Miller et.al in 1956 concluded that the capacity of STM was really between 5 and 9 meaningful items or chunks of information for the typical adult.

Precezewski and Fisher (1990) findings indicate mixing letters and digits within one chunk was more difficult to recall than just having letters or digits make up the chunk

Memory Interference

Phenomenon that occurs due to the negative interaction of similar information in recall.

The existence of multiple password-account pairs and the fact that many accounts require frequent change of the password are the two basic reasons for causing proactive interference.

Adams & Sasse, 1999 suggests 4-5 distinct password is what an average user can remember.

Pilar et.al looked at memory limitations due to age and education on password. They found the single biggest factor that co-related with

Gaw and Felten report that users in their lab study tried an average of 2.43 passwords before a correct login

No more than 7 char passwords and 5 distinct ones.

Zero Memory Authentication

Knowledge

"Something you know"

Ownership

"Something you have"

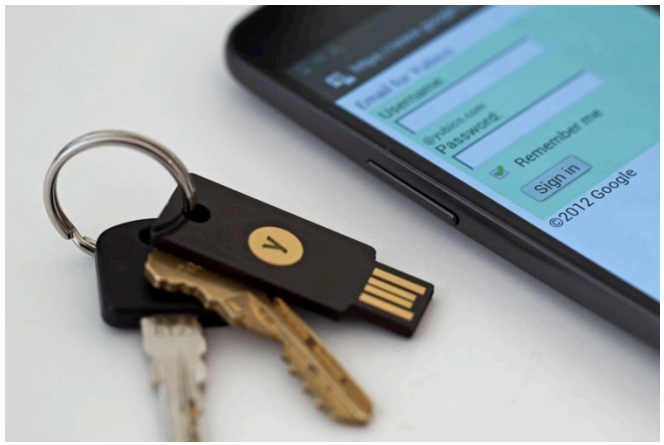
Biometrics

"Something unique to you"

Ownership Authentication

Motorola Skip

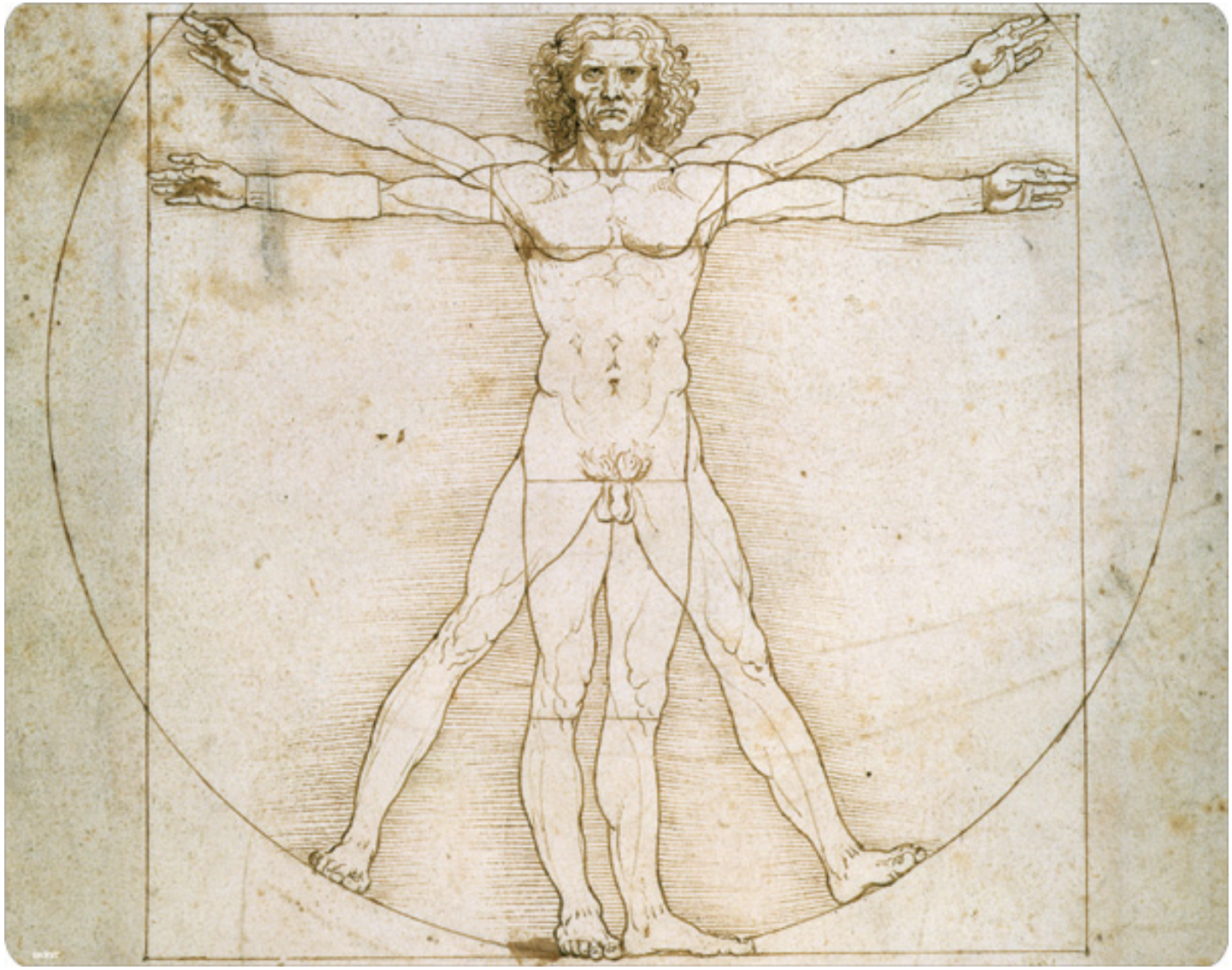
NFC Rings



YubiKey

Digital Tattoo

Can Biometrics Solve the problem?



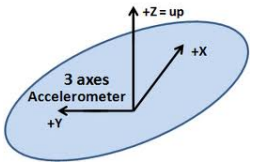
Your phone knows a lot about you

Camera

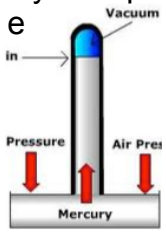
Microphone

Magnetometer

Gyroscope



Accelerometer



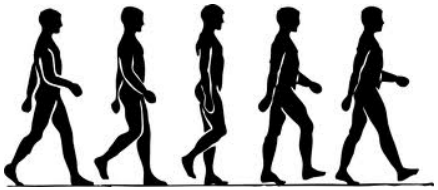
Barometer

GPS

Proximity Sensor

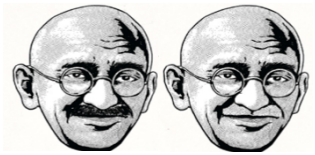
Gender

Color of Skin

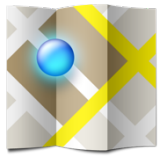


Gait

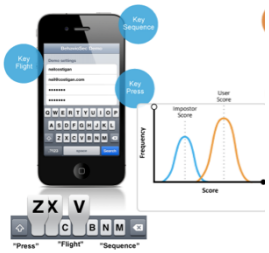
Color of hair



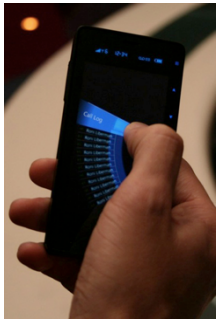
Presence and color of facial hair



Location patterns



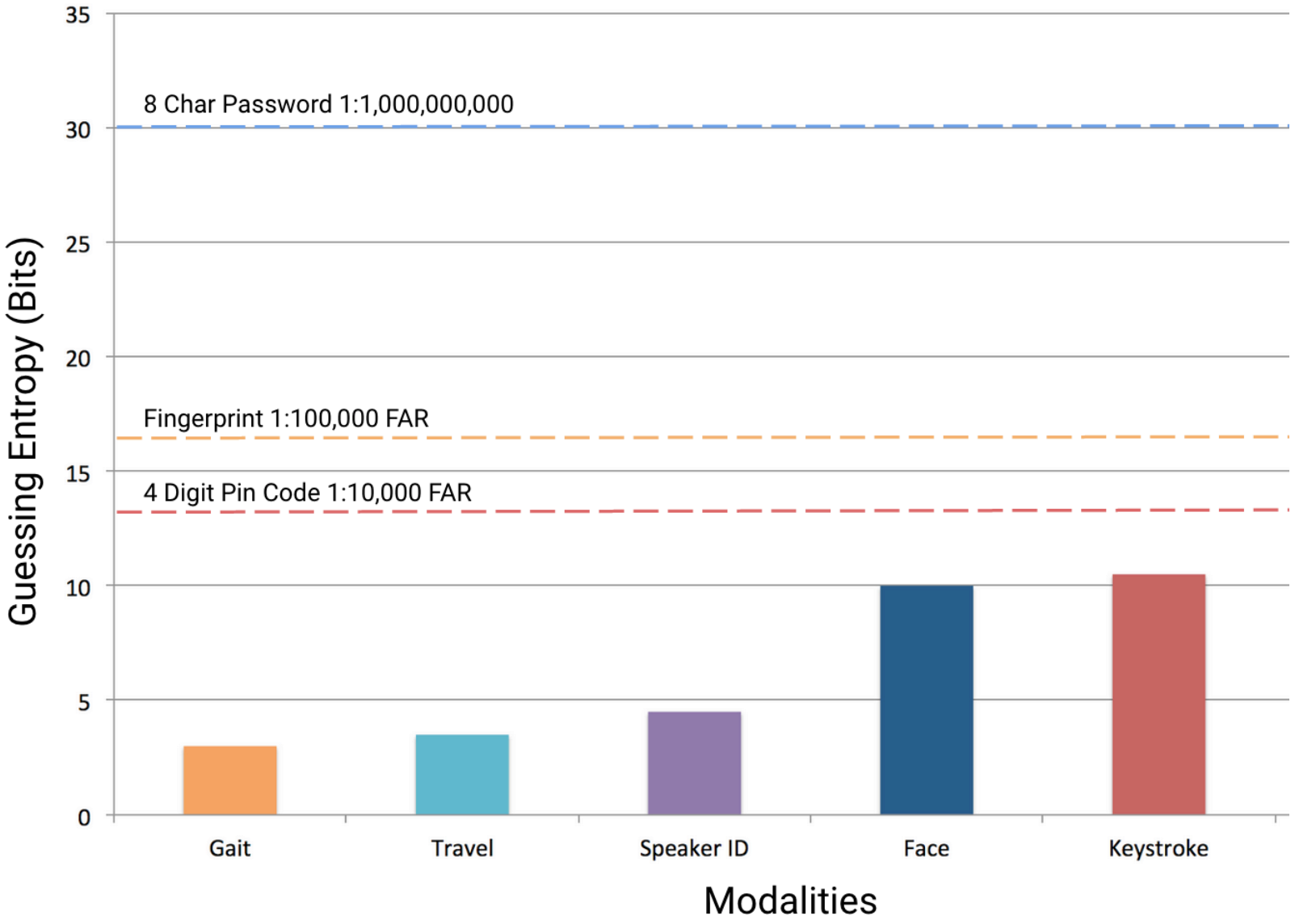
Typing and swipe patterns



Handedness

Google

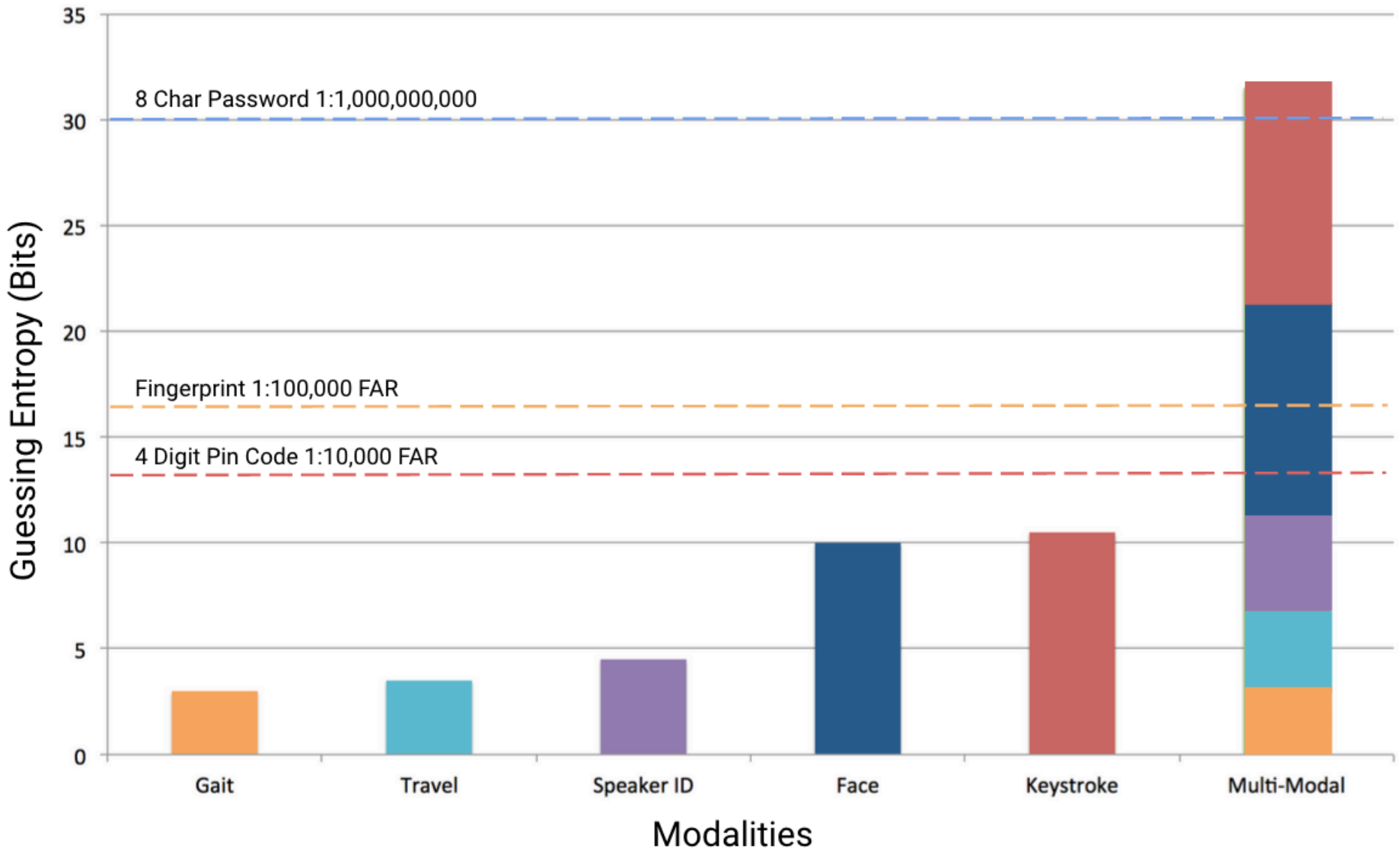
Individual modalities don't cut it



Gait: Benchmarking the performance of SVMs and HMMs for Accelerometer-base Biometric Gait Recognition. 2011. CASED, Germany.
Keystroke: Gunetti, University of Torino, 2002.
Face: Android ICS face unlock data 2011.

Can we combine them to get stronger authentication?

Multiple modalities can add up to strong security

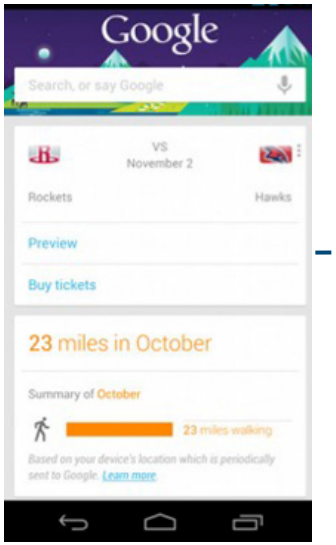


Gait: Benchmarking the performance of SVMs and HMMs for Accelerometer-base Biometric Gait Recognition. 2011. CASED, Germany.
Keystroke: Gunetti, University of Torino, 2002.
Face: Android ICS face unlock data 2011.

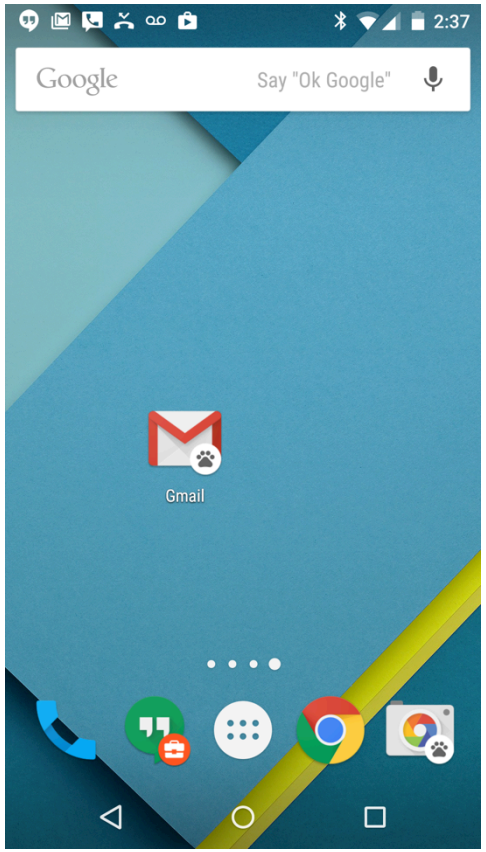
Google

Problems with Lock Screen

4 digit PIN

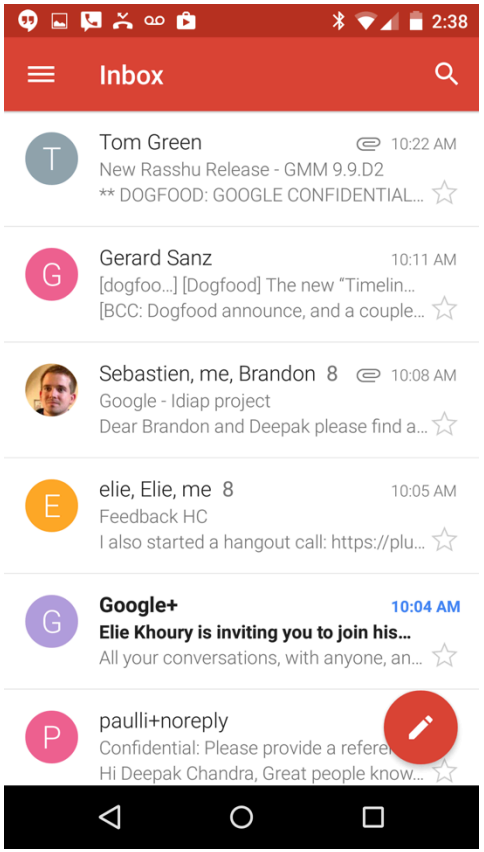


Risk-Based Authentication



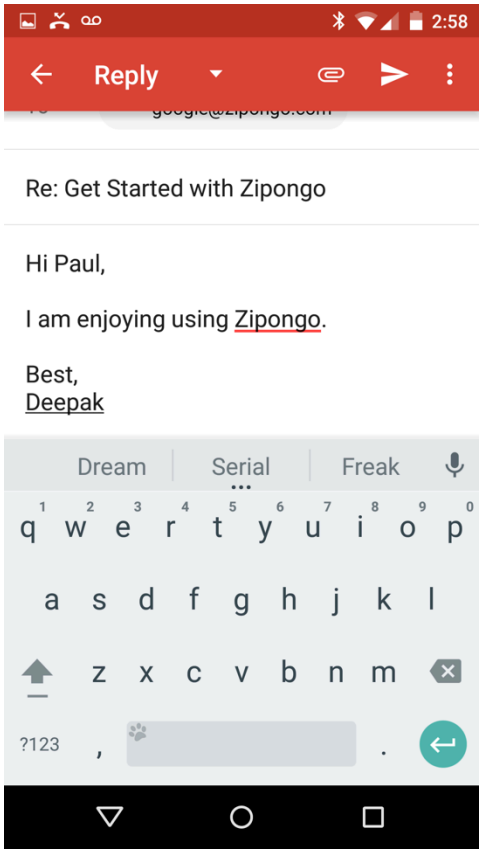
Trust Level: Low
Potential Signals

- Inertial sensors
- GPS/Wifi/Bluetooth
- Ambient voice



Trust Level: Medium
Potential Signals

- Inertial sensors
- GPS/Wifi/Bluetooth
- Ambient voice
- Face
- Swipe



Trust Level: High
Potential Signals

- Inertial sensors
- GPS/Wifi/Bluetooth
- Ambient voice
- Face
- Swipe
- Keystrokes

Google

Authentication as a platform: Beyond securing the phone

Ephemeral Auth

Online Service

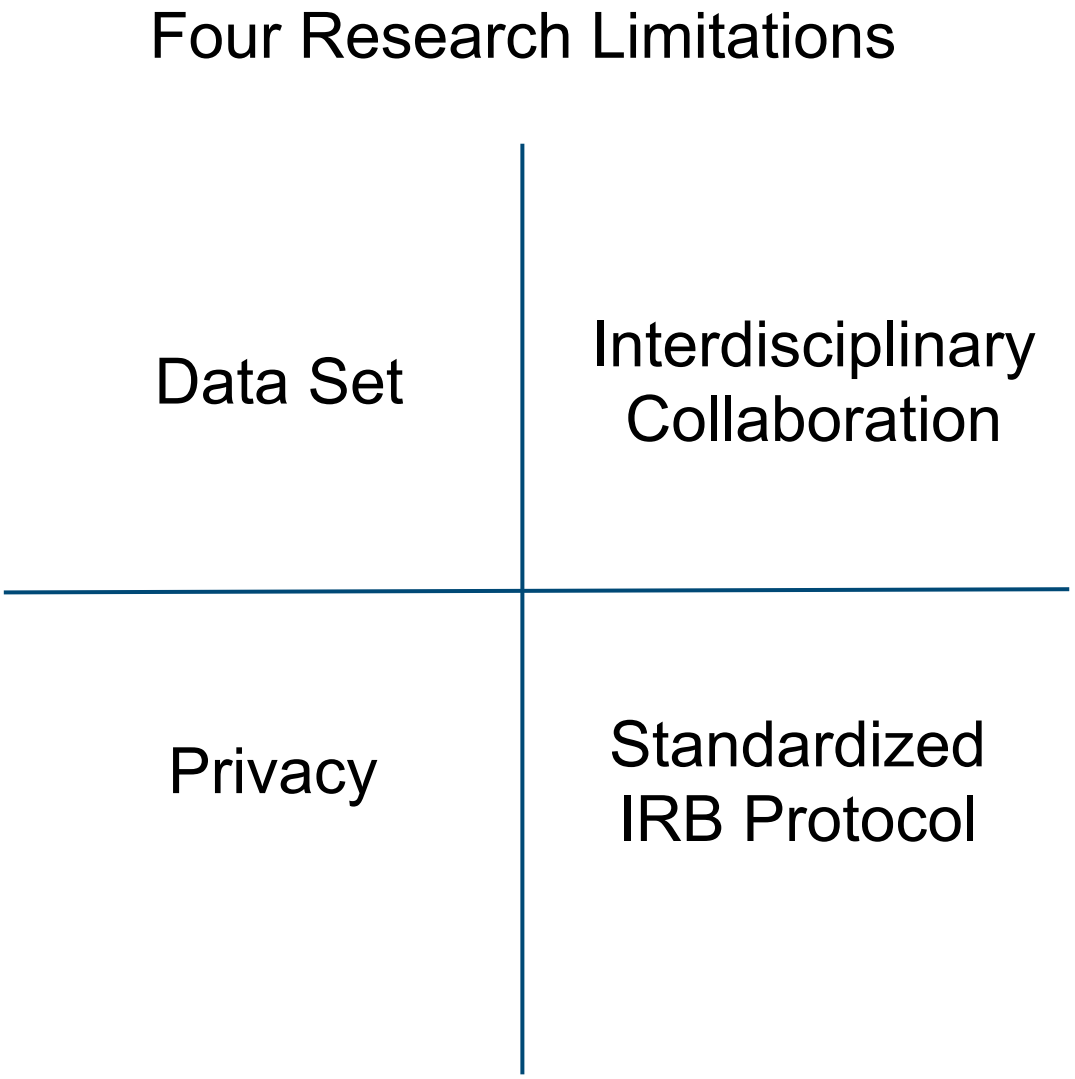
Root of trust for all user authentication needs

Other computers

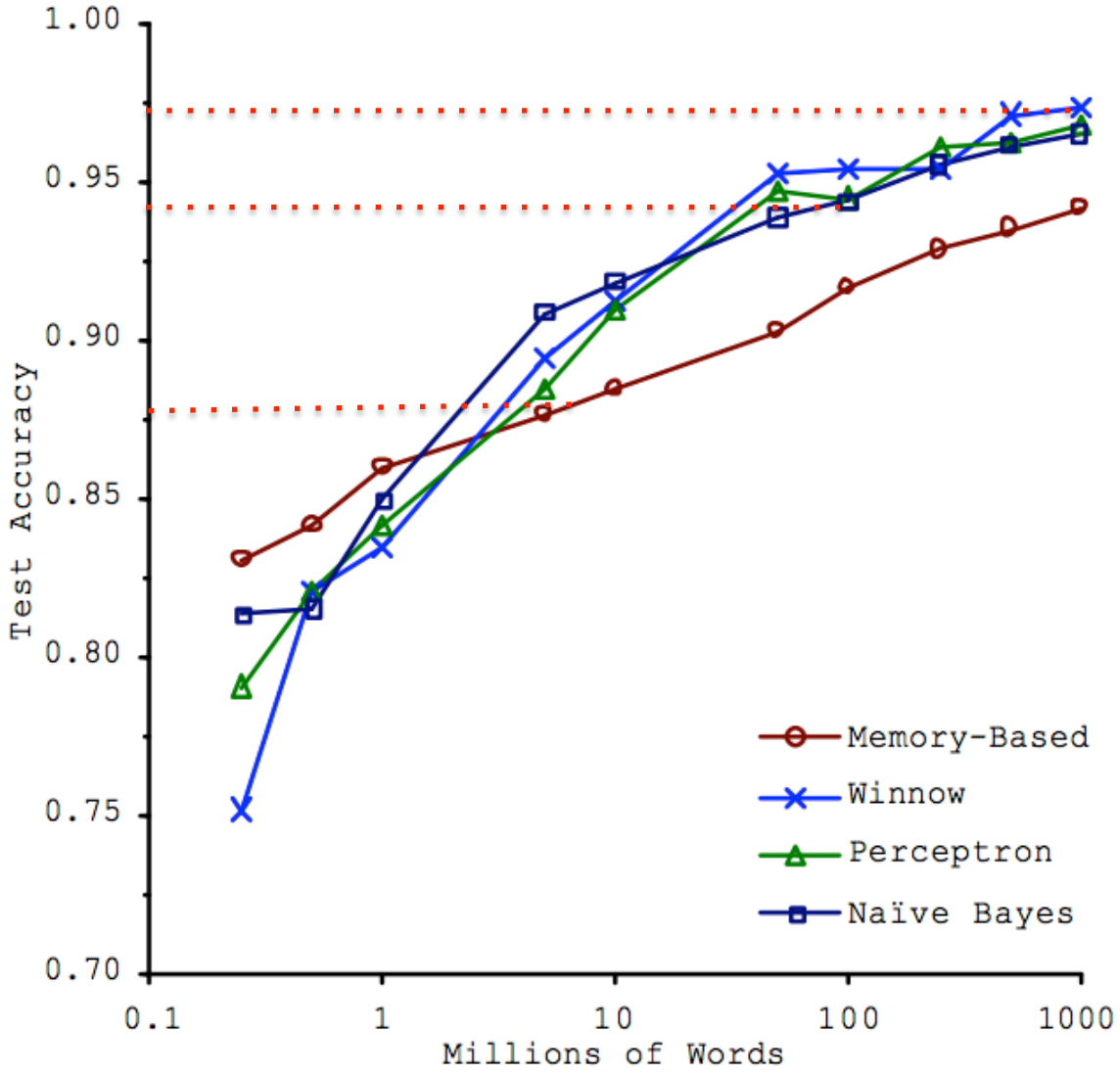
Door locks

Cars

Solving for multimodal authentication



Importance Of Large Datasets



Branco and Brill, 2001

We still work with small Datasets

Lack of Data Set

- **Not comprehensive**: Focused on specific sensors not the full set.
- **Demographically homogenous:** Mostly university students
- **Small**: Data sets are small, typically 20 subjects, large ones have up to 150 subjects
- **Expensive and time consuming:** Time in data collection s/w, subject recruitment and IRB approvals.

Ideal Study Design

- Demographically diverse: age, gender and ethnicity
- Comprehensive: Collect data from all sensors and synchronized so that exact conditions be replicated
- Collect and store data in a way that makes it useful for future studies
- User controls on recording, sharing and deletion

Lots of Signals from a Smartphone

- Front facing camera
- Touch screen
- Key strokes
- Inertial sensors: accelerometer,
- gyroscope and magnetometer
- GPS
- Wifi, BT, and cell tower
- Brightness sensor
- App usage
- Call events

Interdisciplinary Collaboration

Multimodal authentication requires expertise in multiple modalities, however typical research group has expertise in a small number.

Problem in Collaboration

- Groups are geographically disperse
- Privacy consideration limits sharing of the dataset
- No common IRB protocols and privacy regulations differs between states and countries
- No common platform to host datasets and to evaluate and benchmark systems

What is Needed

- Open source platform that is vetted by academic and industry experts
- It can guarantee privacy is subjects in the dataset.
- Automated tools for assessing the privacy sensitivity of queries and to limit the inference about user's information from the output of the queries
- Common IRB protocols and consent forms that can be used internationally by both academia and industry

Interdisciplinary Collaboration Platform

Example: Idiap BEAT

Online platform and framework for standardizing the evaluation of biometric technologies

Consists of:

(1) an online and open platform to transparently and independently evaluate biometric systems against validated benchmarks
(2) protocols and tools for vulnerability analysis
(3) standardized documents for Common Criteria evaluations.

Goals:

(1) Reliability of biometric systems can be measured, leading to an increase in performance
(2) Technology transfer from research to industry will be eased by an interoperable framework
(3) Decision-makers and authorities will be informed about progress in biometrics as the results will have an impact on standards

Designing for Privacy Preserving Systems

On device computation necessary for privacy

Global background model + on device local adaptation

Mobile challenges

Limited

- Power
- Memory
- Compute

Designing systems and algorithms that can adapt to different phone models

- Different set of sensors
- Differences of sensor and calibrations

Mobile opportunities

Lots of training data

Continuous training to account for temporal changes

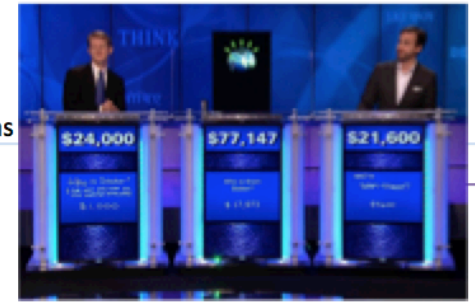
Number and fidelity of sensors is increasing

Isolated Execution Environment (TrustZone) to securely store templates

Privacy Preserving ML

- How do you update the global background models?
- Provably limit the inference about user's information from the global model
- Anonymization ≠ Privacy
  - [Narayanan & Shmatikov !08] identify Netflix users from anonymized records, IMDB.

The Rise of Mobile GPU Computing



IBM Watson Wins
Jeopardy, 2011

Stanley, Winner of 2005
Grand DARPA Challenge

MIPS

1,000,000
100,000
10,000
1,000
100
10
1

1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020

ARM7500FE
40MHz

ARM11
412MHz

ARM Cortex
A8 @ 1GHz

Qualcomm Cortex
A8 @ 1GHz

ARM Cortex A9
Dual @ 1.5GHz

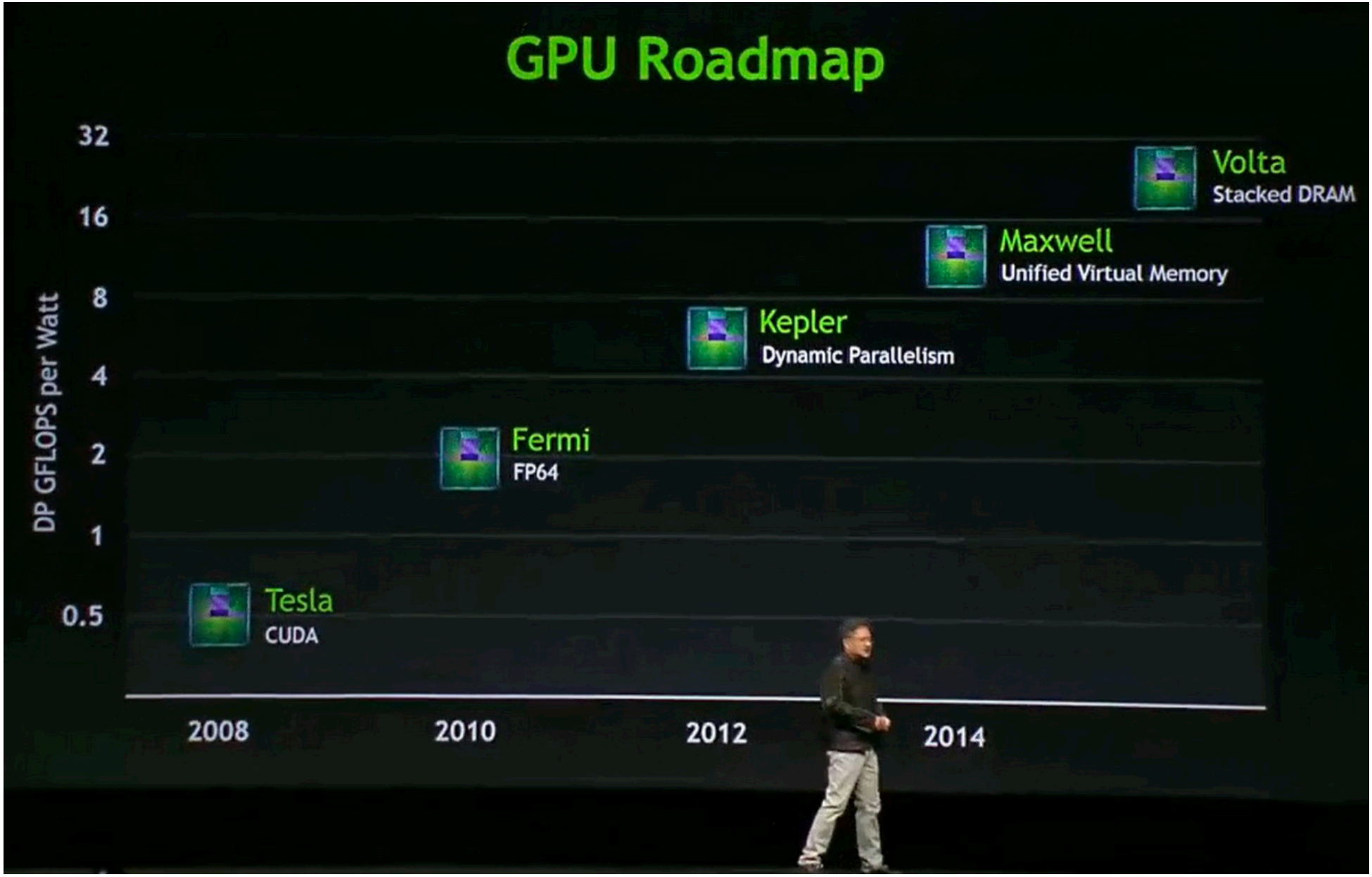
Snapdragon 800
@ 2.3 GHz
+ Adreno 420 GPU

Nvidia Tegra K1
@ 2.5 GHz
+ Kepler GPU

Mobile GPUs

Google

The Rise of Mobile GPU Computing



Google

Solving for a Standardized IRB Protocol

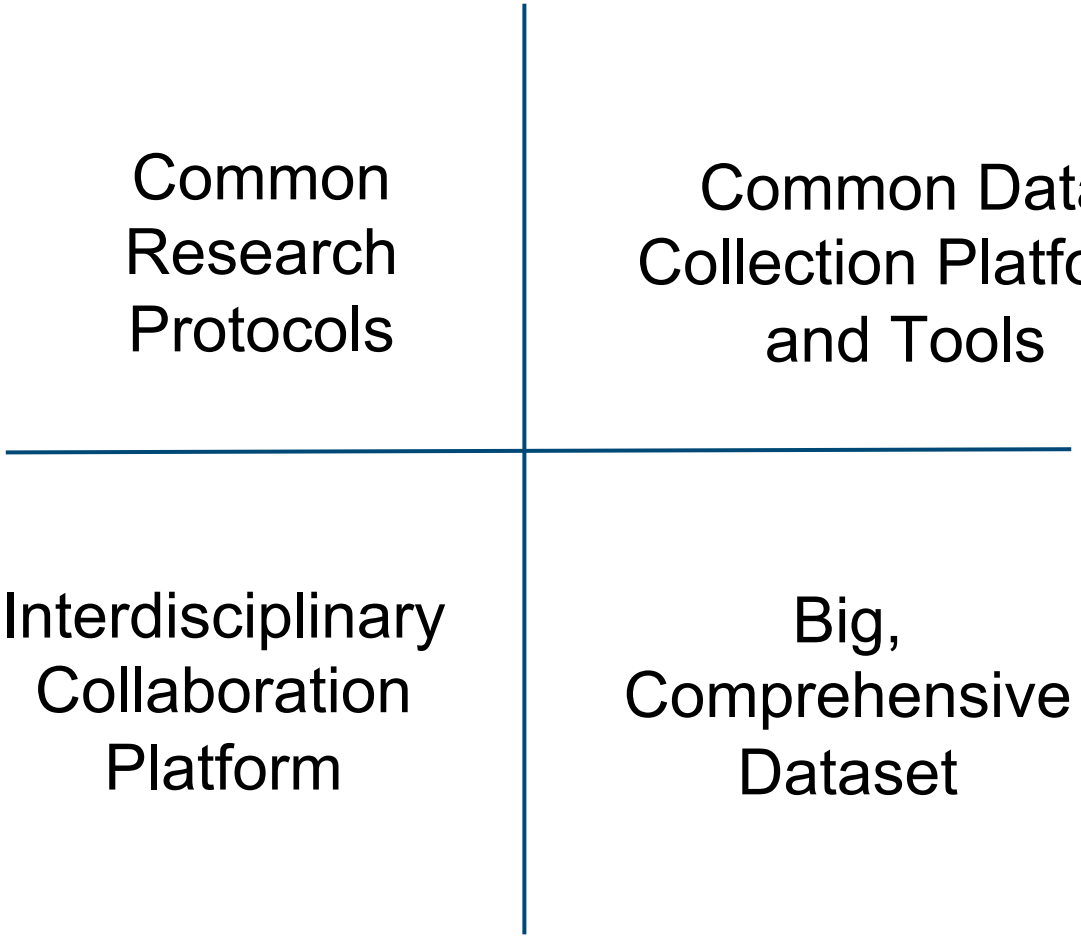
Enable research by making personal sensor data studies easier

- Data sets cannot be collected or shared without extensive IRB approvals due to privacy concerns
- Homebrew protocols are wasted effort and are not consistently sound or comprehensive
- Privacy concerns hamper current research by restraining the usefulness of data sets.
- Individual IRBs may be specialized for the nuances of privacy sensitive sensor data

Standard IRB Protocol

- Best practices guidelines for
  - Consent form to be meaningful
  - User privacy controls:
  - Ability to opt-in and opt-out of studies
  - Security guidelines for collection and storage
- Reference collection software
- Reference research infrastructure
- Endorsement from privacy advocates
- Protocol approved by university IRB so that approvals are minimal for specific studies

What's needed to democratize biometrics research?



What's possible when we democratize biometrics research?

Let’s foster collaboration on data set creation and research

Deepak Chandra
Technical Program Lead, Google ATAP
dchandra@google.com