MIGRATING BIOMETRICS TO MOBILE SCENARIOS: PERFORMANCE AND USABILITY EVALUATION

R. Sanchez-Reillo, R. Blanco-Gonzalo, O. Miguel-Hurtado, A. Mendaza-Ormaza

University Group for Identification Technologies (GUTI)

Carlos III University of Madrid

http://guti.uc3m.es

http://idtestinglab.uc3m.es

rsreillo@ing.uc3m.es

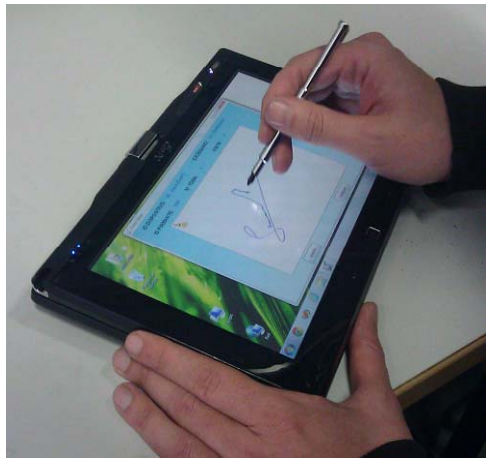
OUTLOOK

- Introduction
  - Capture Devices
  - Sample Illustrations
- Performance Evaluation
  - Specifications
  - Parameter Study towards Interoperability
  - Conclusions (Performance) – 1
  - Skilled Forgeries
  - Conclusions (Performance) – 2
- "Usability" Evaluation (Scenario Evaluation)
  - Specifications
  - Intra-device and Intra-scenario
  - Scenario Inteoperability
  - Expected Interoperability Deployment
  - Conclusions (Usability)

INTRO: HANDWRITTEN SIGNATURE

- Handwritten Signature:
  - Off-line (static) signature
  - **On-line (dynamic) signature**
    - Recommended for non-forensic applications
- Motivation for using Handwritten Signature:
  - The de-facto authentication method in many operational scenarios
  - Accepted by users of any age and cultural level
  - New capture devices are cheap and high quality
  - Touch sensitive screens can be used as capture device
    - Already in the hands of the final user ➔ "costless" biometrics
- Motivation for this study
  - Is there any impact on the capture device technology?
  - Is there any impact on the "item" used to sign (stylus/finger)?
  - Is there any impact on the position of the user when signing?

INTRO: CAPTURE DEVICES (STYLUS)



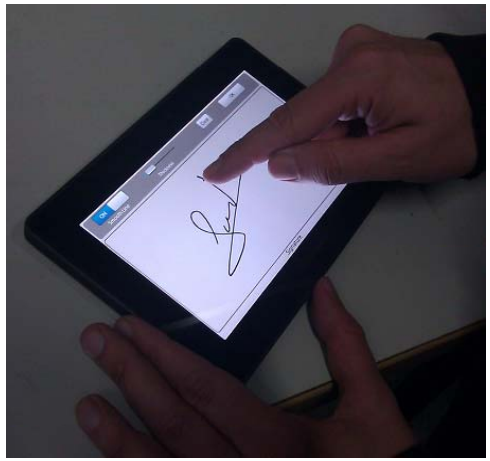
- Asus Eee PC Touch T101MT
- Tablet PC
- Mixed
- 10.1”

- Wacom STU-500
- Peripheral
- 5”

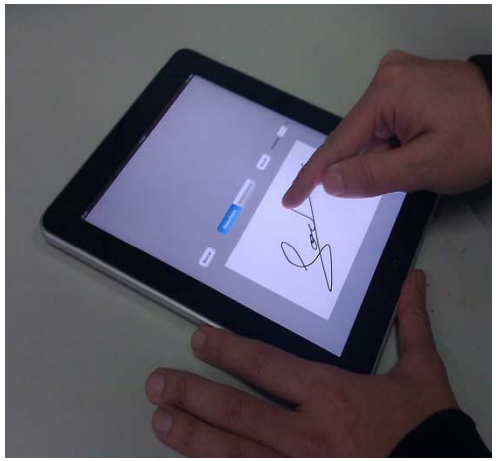
- Samsung Galaxy Note (Note-S)
- Smartphone
- Mixed
- 5.3”

- Wacom Intuos 4
- Peripheral
- 10.81”

INTRO: CAPTURE DEVICES (FINGER)

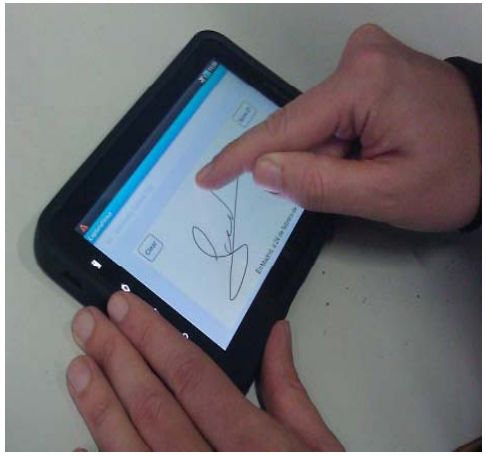


- BlackBerry Playbook
- Tablet
- Capacitive
- 7”



- Apple iPad
- Tablet
- Capacitive
- 9.7”

- Samsung Galaxy Note (Note-F)
- Smartphone
- Mixed
- 5.3”



- Samsung Galaxy Tab
- Tablet
- Capacitive
- 7”

PERFORMANCE EVALUATION

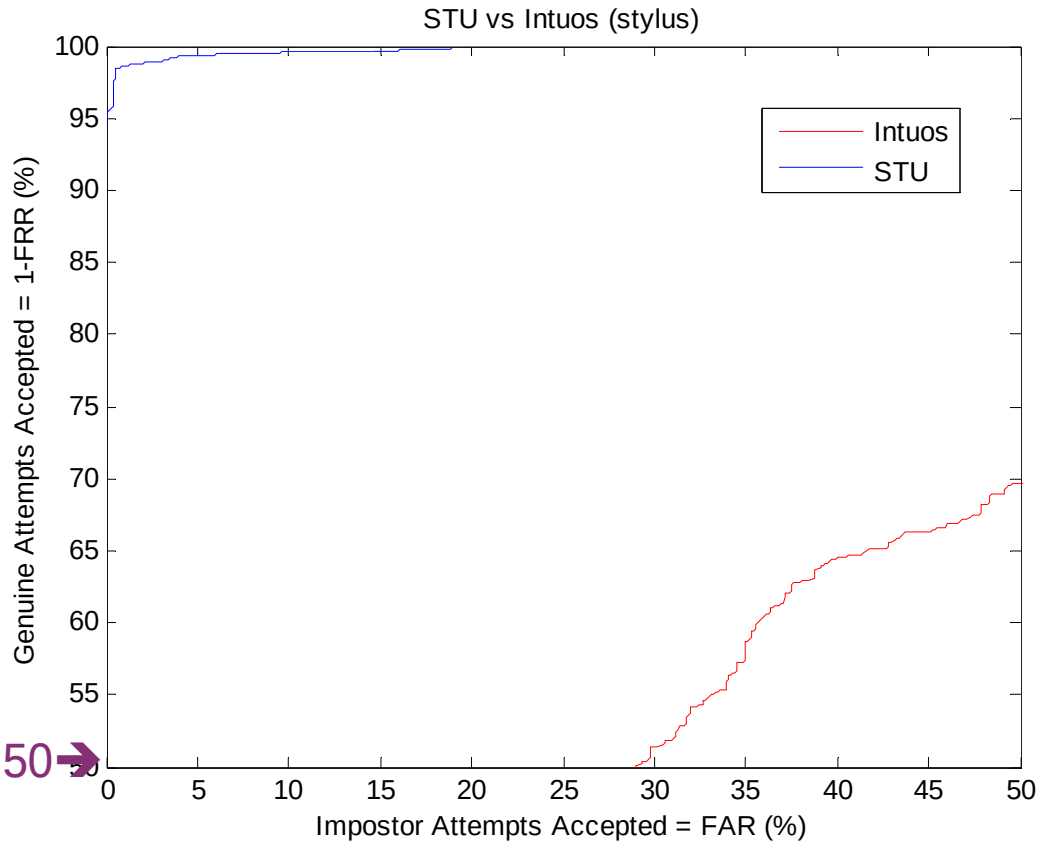
SPECIFICATIONS

- All devices
- Crew: 11 people
  - Age: 24 – 39 years old
  - Other data: Familiar with the technology. Engineers.
- Sessions: 3
- Signatures/session: 20
- All users' real signatures
- Skilled forgeries:
  - Advanced knowledge about the signature to forge
  - 2 x 10 forgeries/user

- Target: Obtain a preliminary idea on the different parameters that may affect going towards mobility
- Algorithm used:
  - DTW-based
  - Only X and Y signals (no pressure!)
  - Enrolment with the 3 first signatures (not the "typical best" 5)
  - Genuine Signatures (Random Forgeries): EER = 1.8% (MCYT)
  - Skilled Forgeries: EER = 7.6% (MCYT)

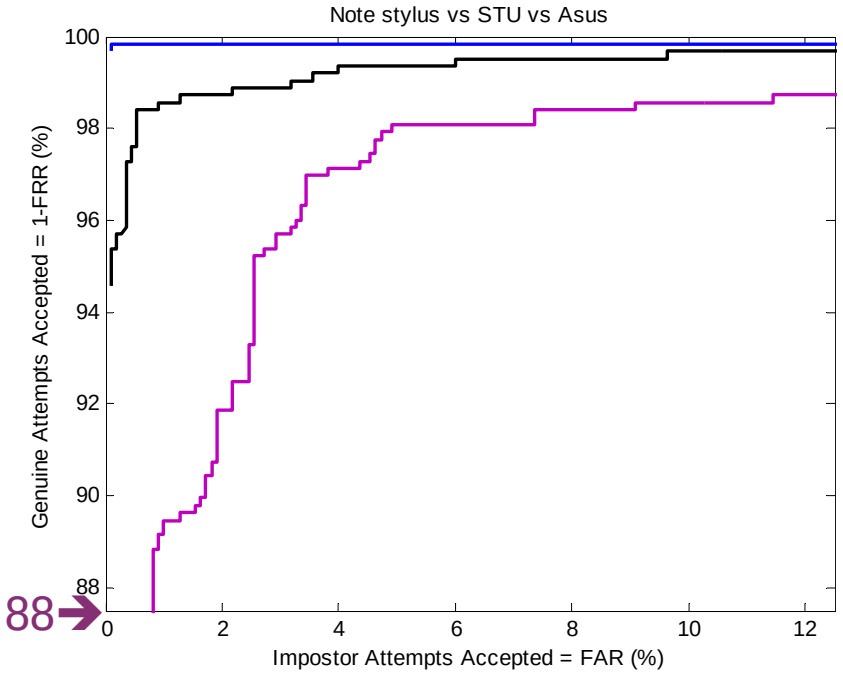
DOES VISUAL FEEDBACK MATTERS?

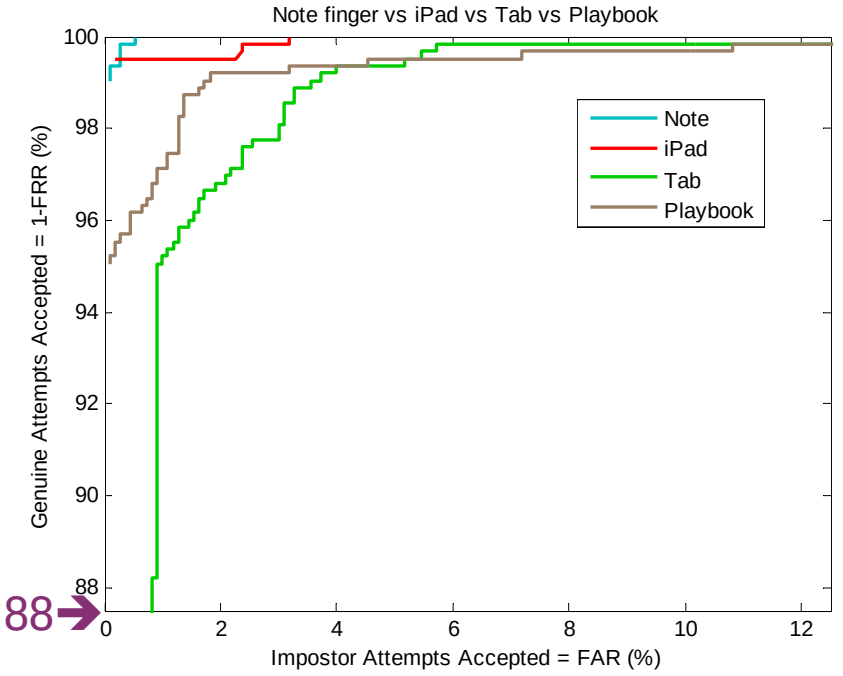
- Yes!
  - STU (visual fb, EER=1,27%) vs. Intuos (non-visual fb)
  - EERs for different users:
    - 31.79%
    - 12.14%
    - 28.03%
    - 35.04%
    - 19.05%
    - 47.18%
    - 40.17%
    - 33.16%
    - 56.07%
    - **22.90%**
    - 12.14%



DOES TECHNOLOGY MATTERS?

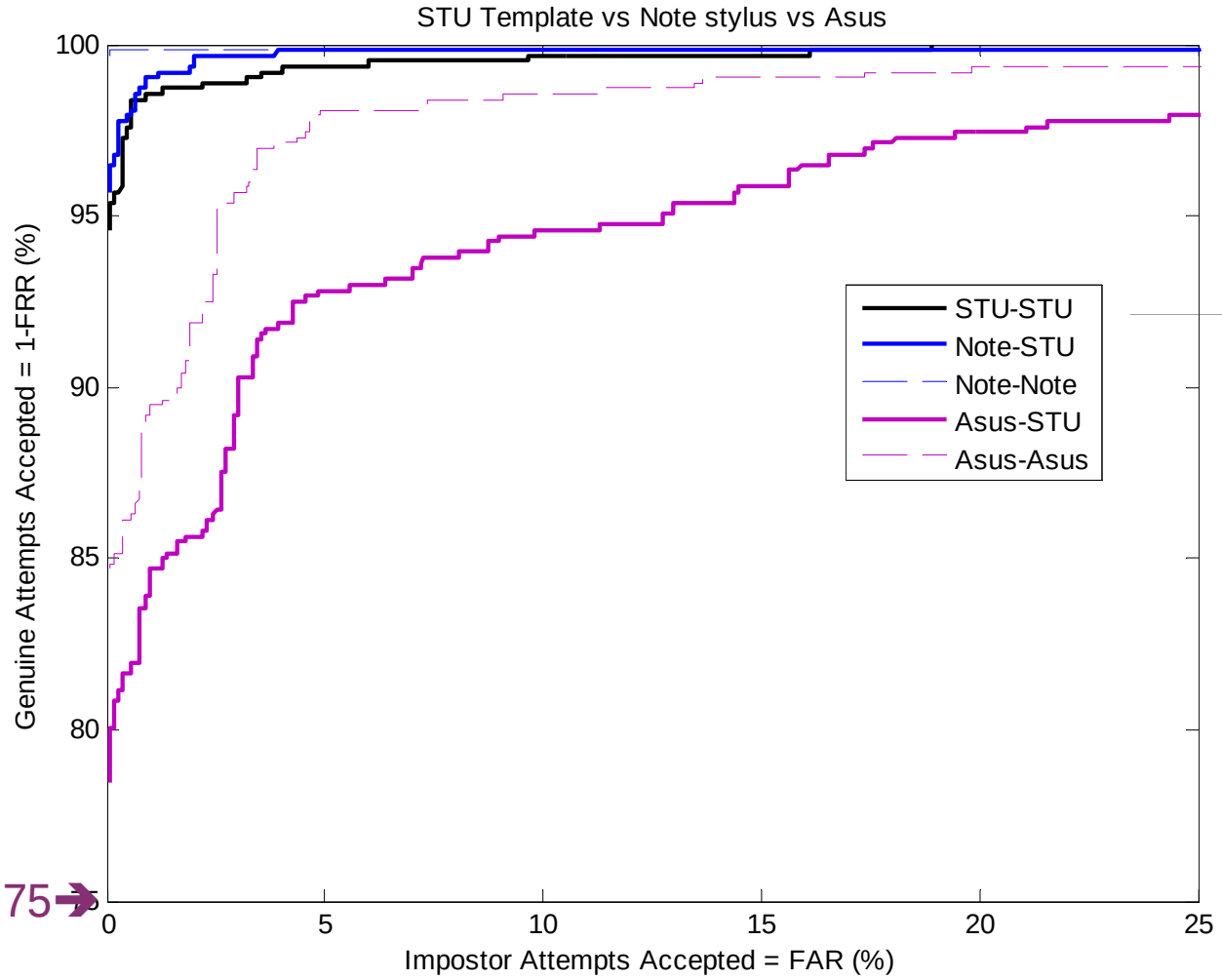
- Technology = Capture technology + capture size + O.S.
- Technology matters:
  - With Stylus (EER):
    - Note-S = 0.17; STU = 1.27; ASUS = 3.48
  - With finger (EER):
    - Note-F = 0.29; iPad = 0.47; Playbook = 1.39; Tab = 2.38
- Not really sure about which aspect of technology makes the dependency:
  - Not found a relationship with size (from a minimum size of the Note) or O.S.
  - About O.S.: which version? Which implementation?





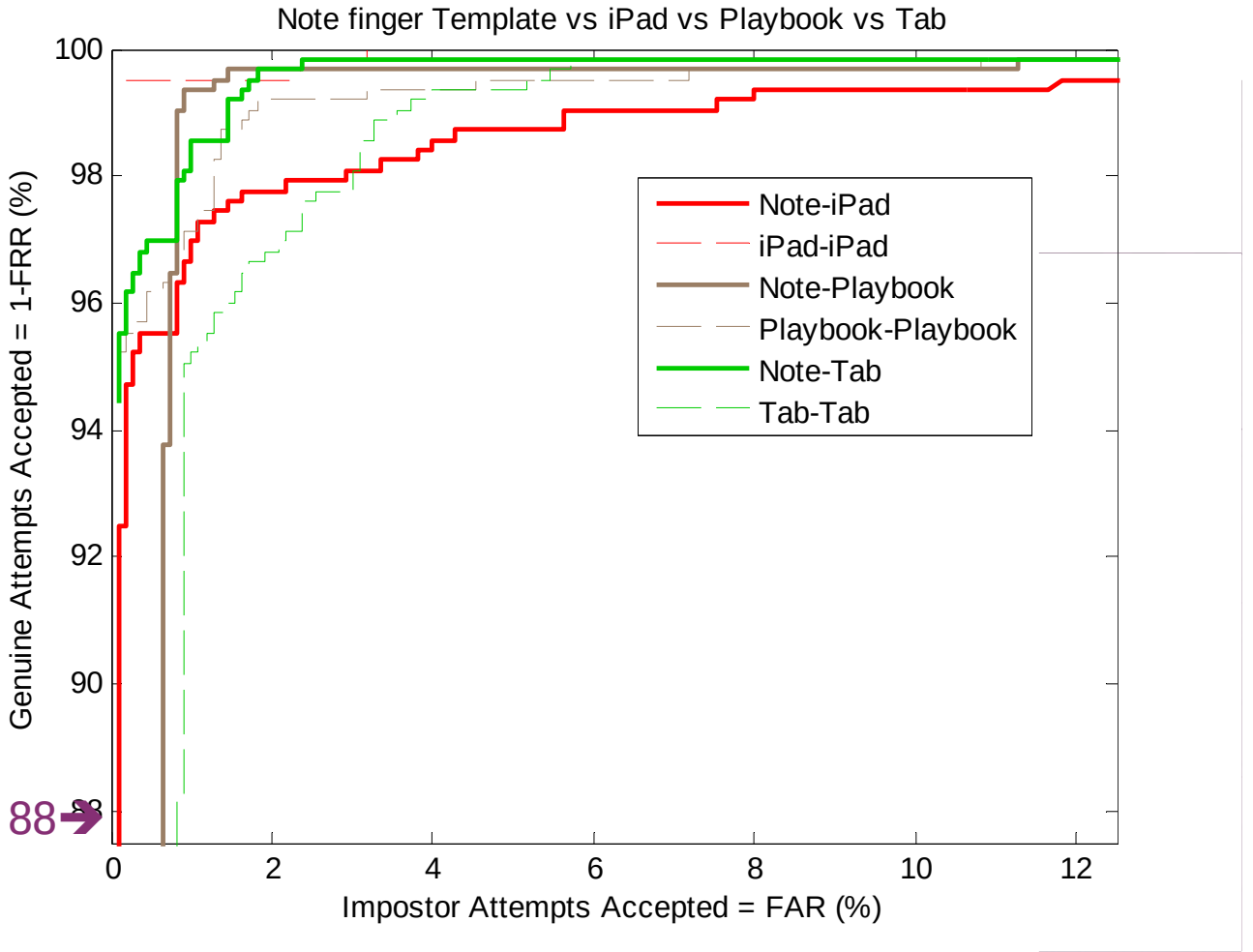
INTEROPERABILITY (STYLUS)

- STU as reference (office scenario)
- RESULTS: lower performance than with it's own pattern, but acceptable:
  - Note:
    - 0.17 ➔ 0.98
  - Asus:
    - 3.48 ➔ 6.84



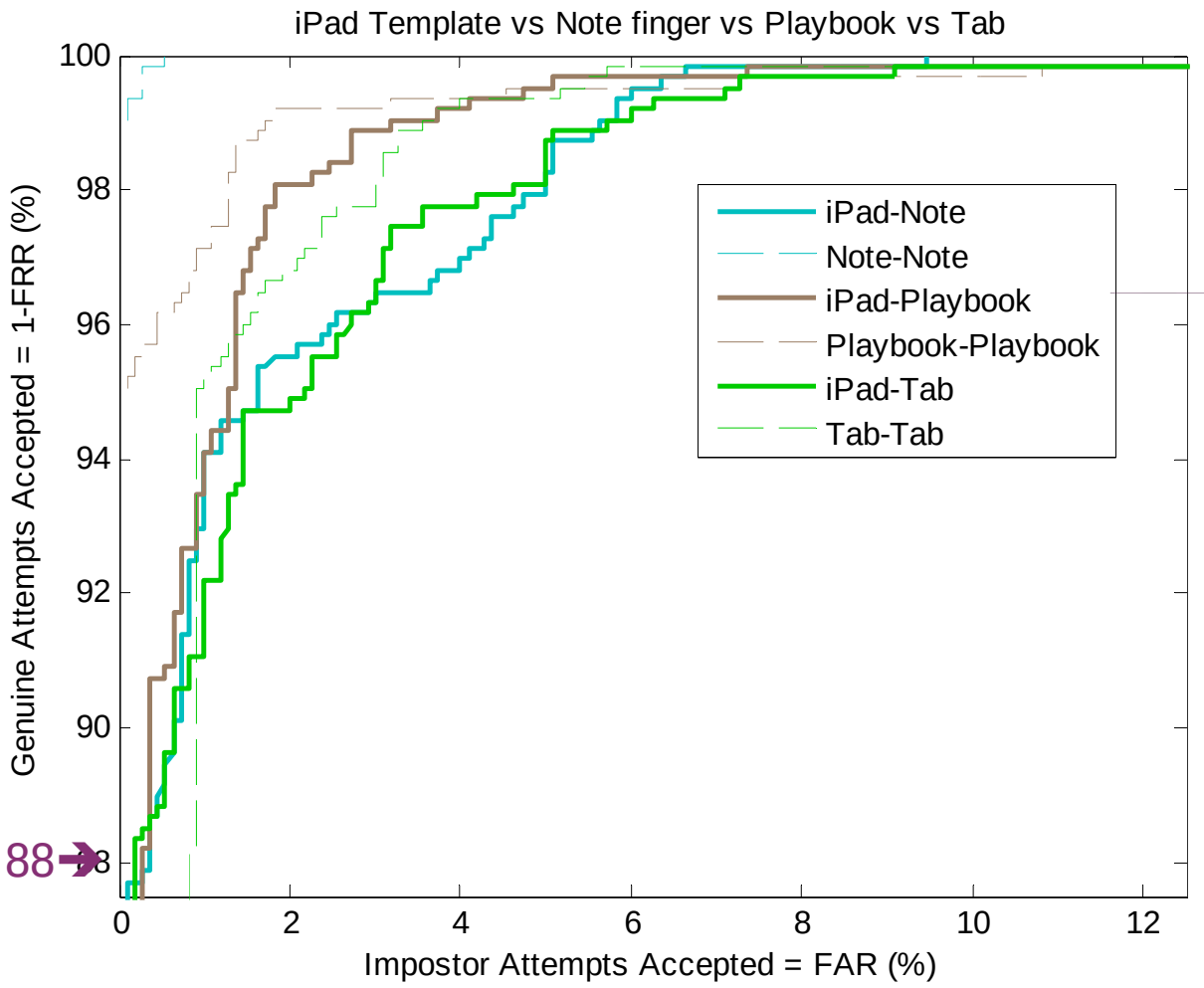
INTEROPERABILITY (FINGER) - 1

- Note-F for reference (best result intra-device):
  - Tab:
    - 2.38 ➔ 1.45 !!
  - iPad:
    - 0.47 ➔ 2.21
  - Playbook:
    - 1.39 ➔ 0.93 !!



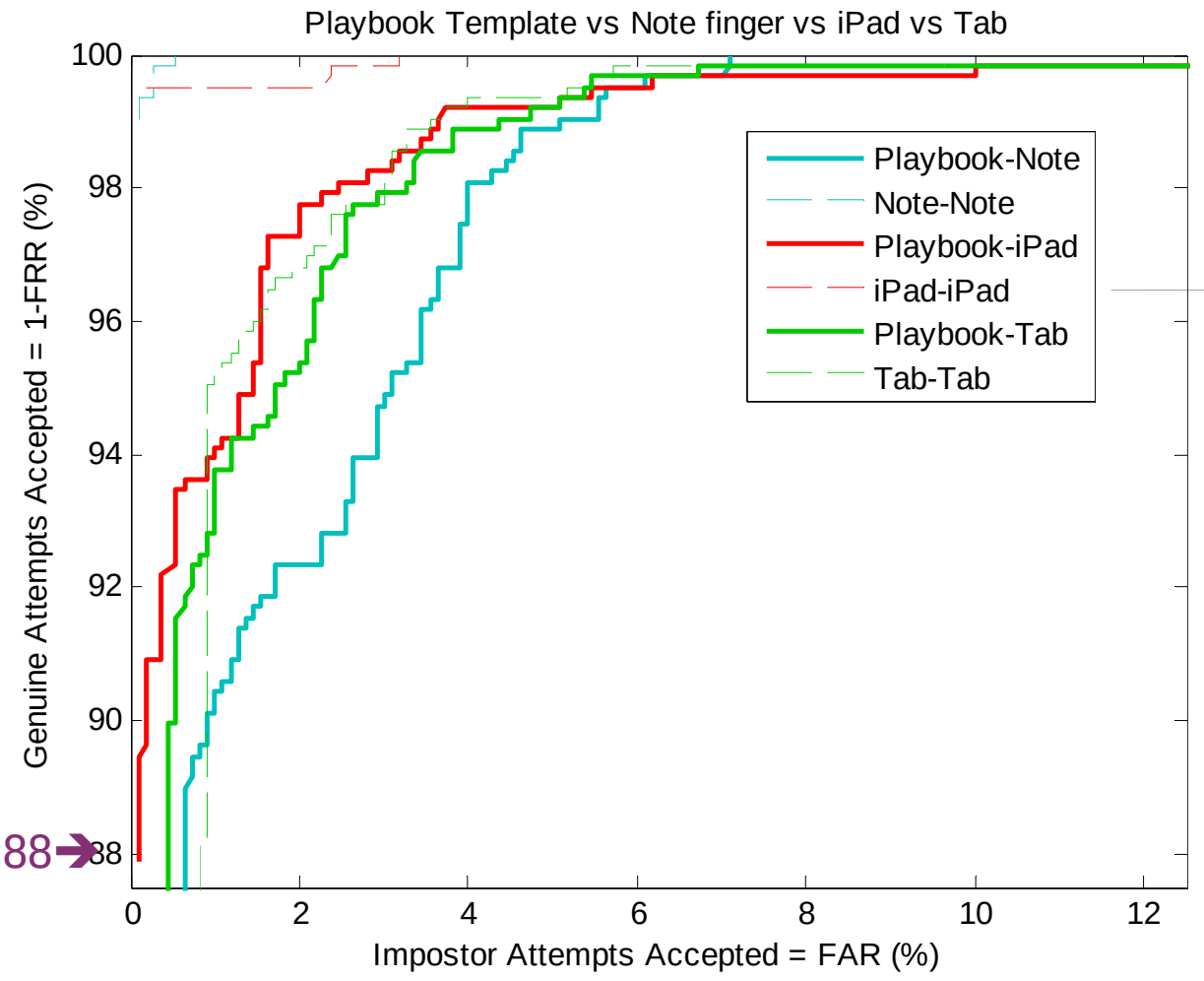
INTEROPERABILITY (FINGER) - 2

- iPad for reference (second best):
  - Tab:
    - 2.38 ➔ 3.06
  - Note-F:
    - 0.29 ➔ 3.52
  - Playbook:
    - 1.39 ➔ 1.91



INTEROPERABILITY (FINGER) - 3

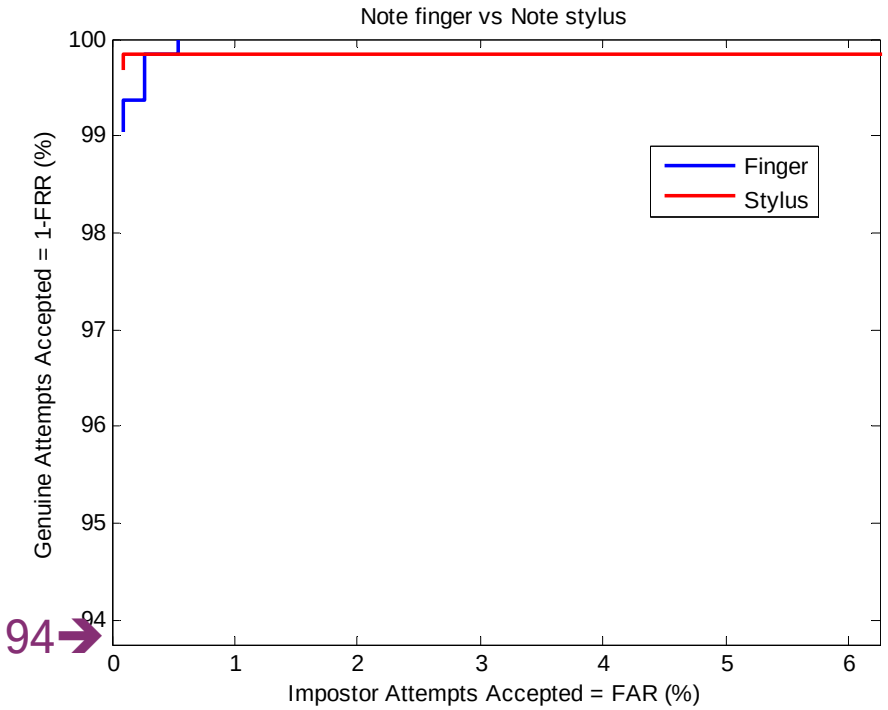
- Playbook for reference:
  - Tab:
    - 2.38 ➔ 2.55
  - Note-F:
    - 0.29 ➔ 3.65
  - iPad:
    - 0.47 ➔ 2.25

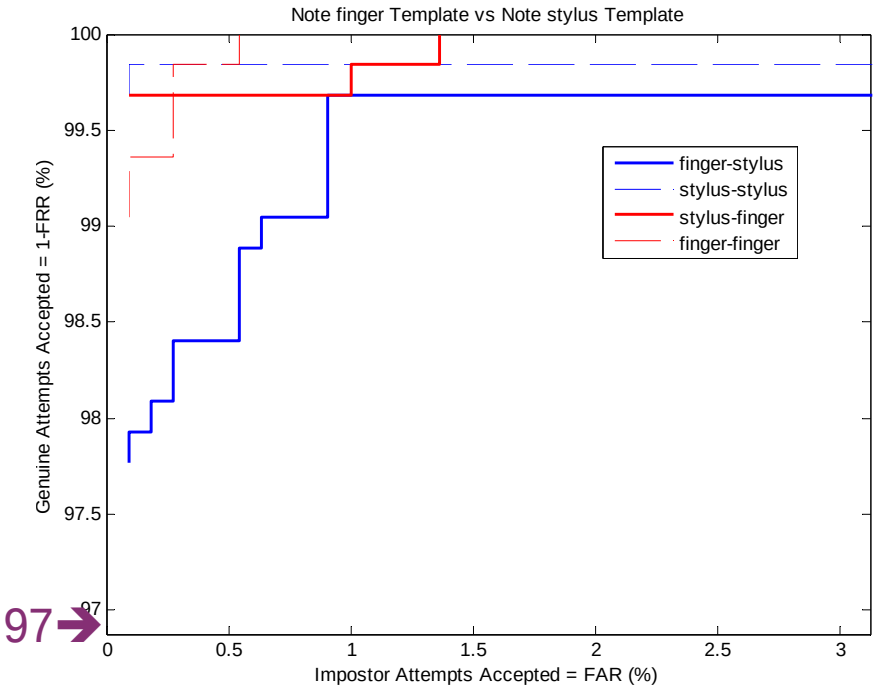


88➔

ONE OR TWO MODALITIES - 1

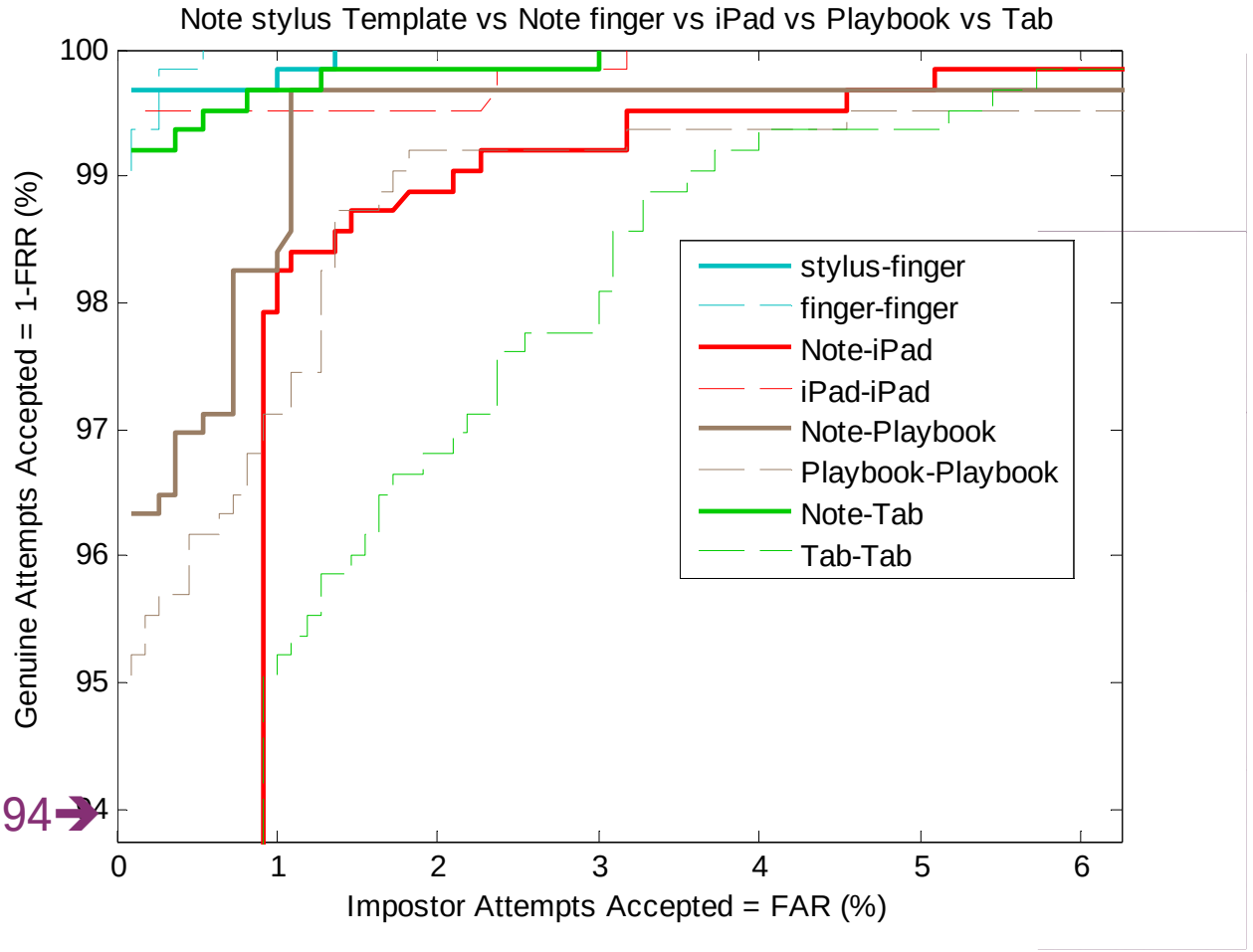
- Comparing within the same device (Note):
  - EER Stylus = 0.17; EER Finger = 0.29
- Comparing "inter-modality"
  - Enrolling with Stylus: 0.34; Enrolling with Finger: 0.93
- Preliminary Result: Not the same, but comparable!





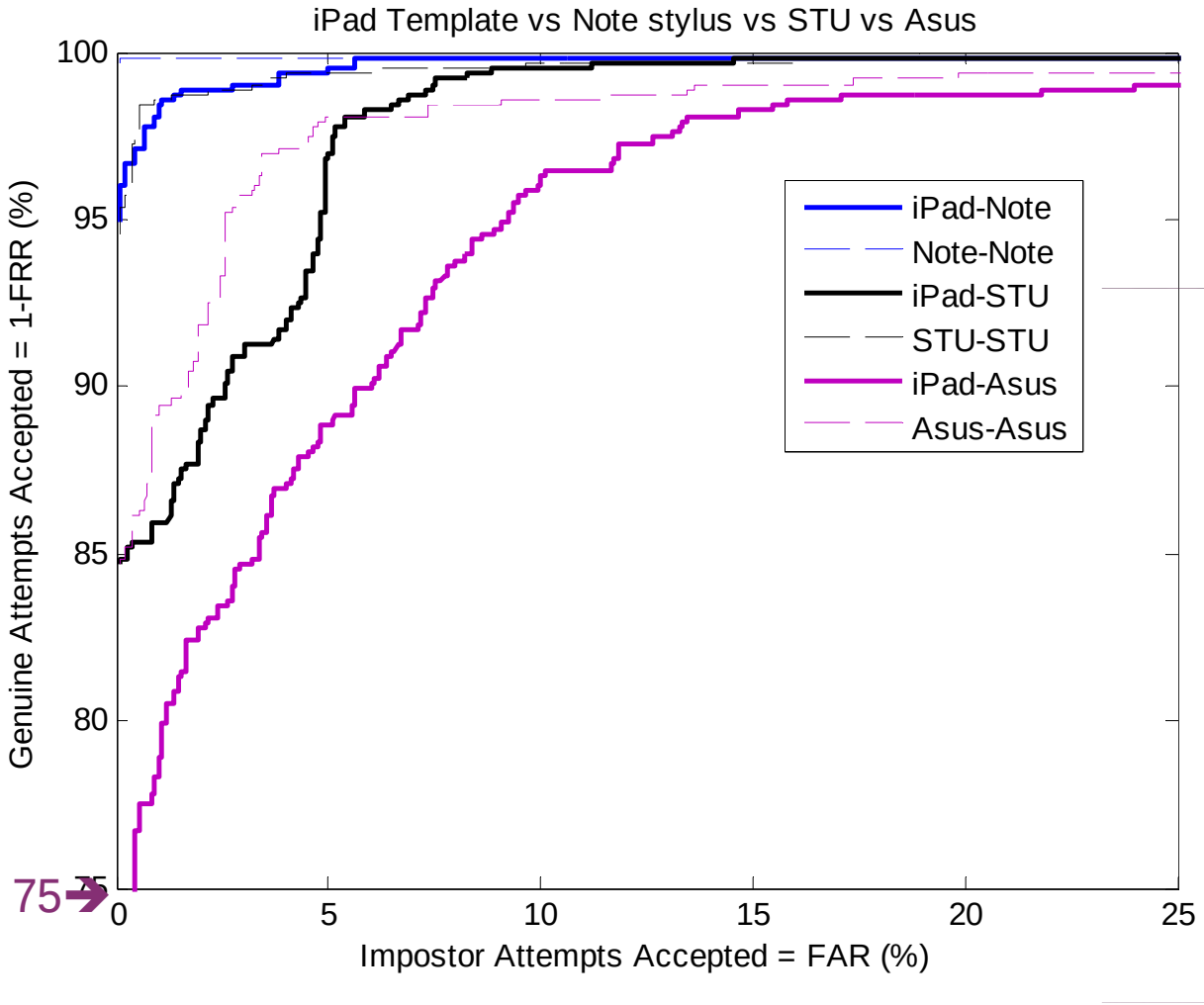
ONE OR TWO MODALITIES - 2

- Note-Stylus for reference:
- Tab:
  - **2.38 → 0.55 !!**
- Note-F:
  - 0.29 → 0.34
- Playbook:
  - **1.39 → 1.10 !!**
- iPad:
  - 0.47 → 1.45



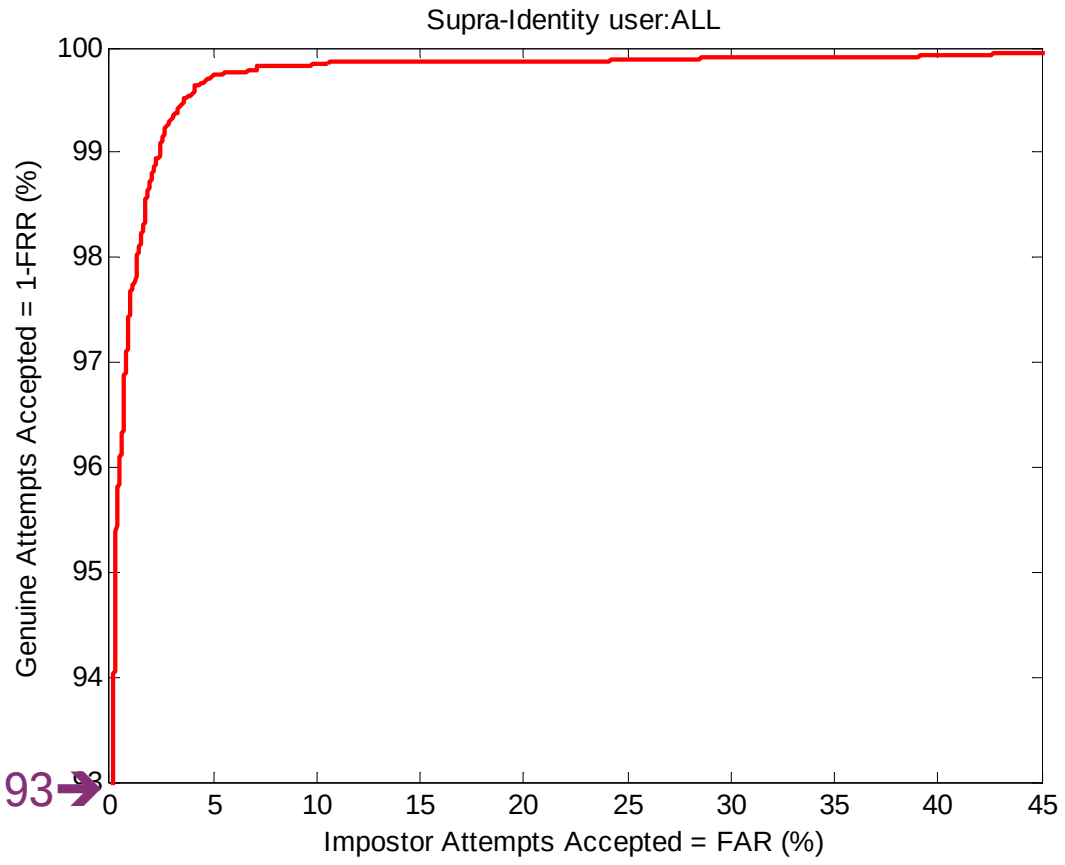
ONE OR TWO MODALITIES - 3

- iPad for reference:
- Note-Stylus:
  - 0.17 ➔ 1.40
- Asus:
  - 3.48 ➔ 7.30
- STU:
  - 1.27 ➔ 4.80



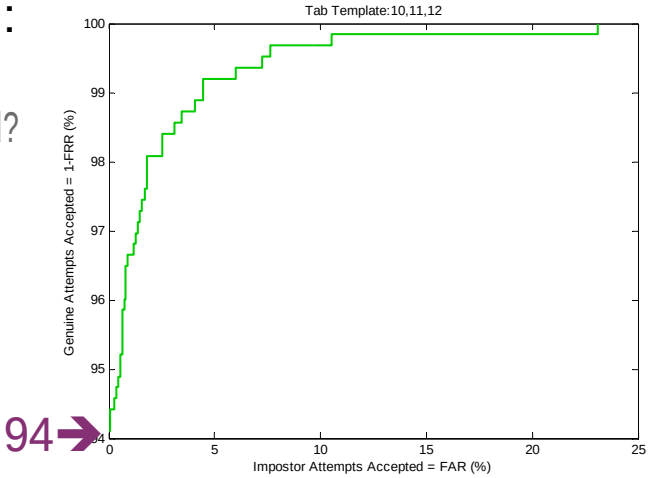
INTEROPERABILITY - II

- Supra-Identity:
  - Each user has 7 different biometric references
  - Best score is used
  - EER = 1.66



CONCLUSIONS (PERFORMANCE) - 1

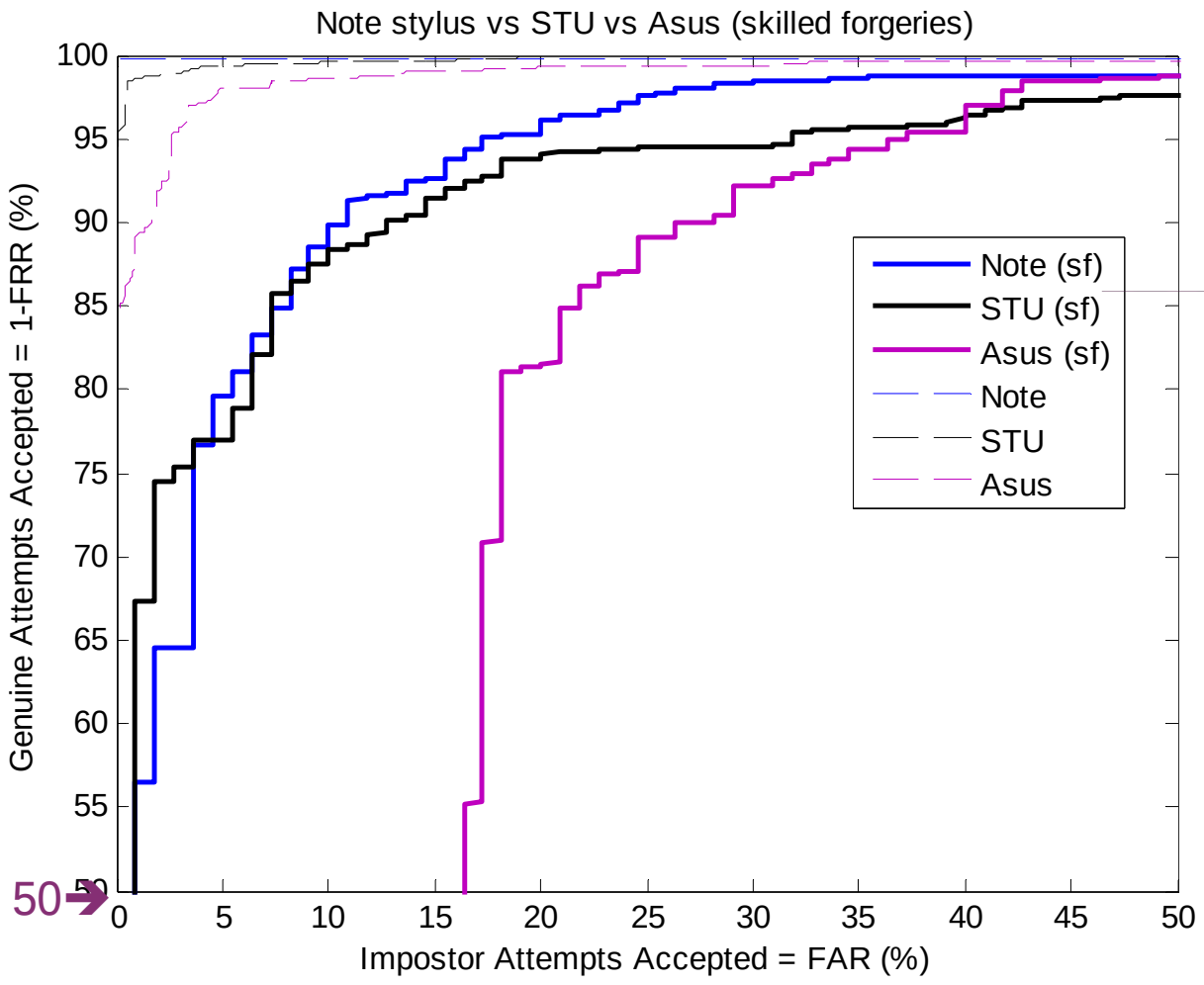
- Visual Feedback is a major parameter for performance!
- **No relationship with technology has been found for different behaviours**
- Within genuine samples, interoperability is affordable:
  - Within stylus devices is really good
  - Within finger-based devices in affordable
    - Improvement of EER in Tab or Playbook, having the template with Note-F, **shows a potential problem in enrolling with Tab or Playbook**.
      - Habituation? Enrolling with samples 10, 11 and 12 improves EER: 2.38 ➔ 1.91 (Figure below)
      - **But the problem still exists!**
- Results initially shows that a single modality can be considered
  - Interoperability is acceptable
    - Enrolling with stylus results may even get better
- Supra-identity shows really good results:
  - But enrolling is tiring for users
  - And what about future devices being used?



94➔

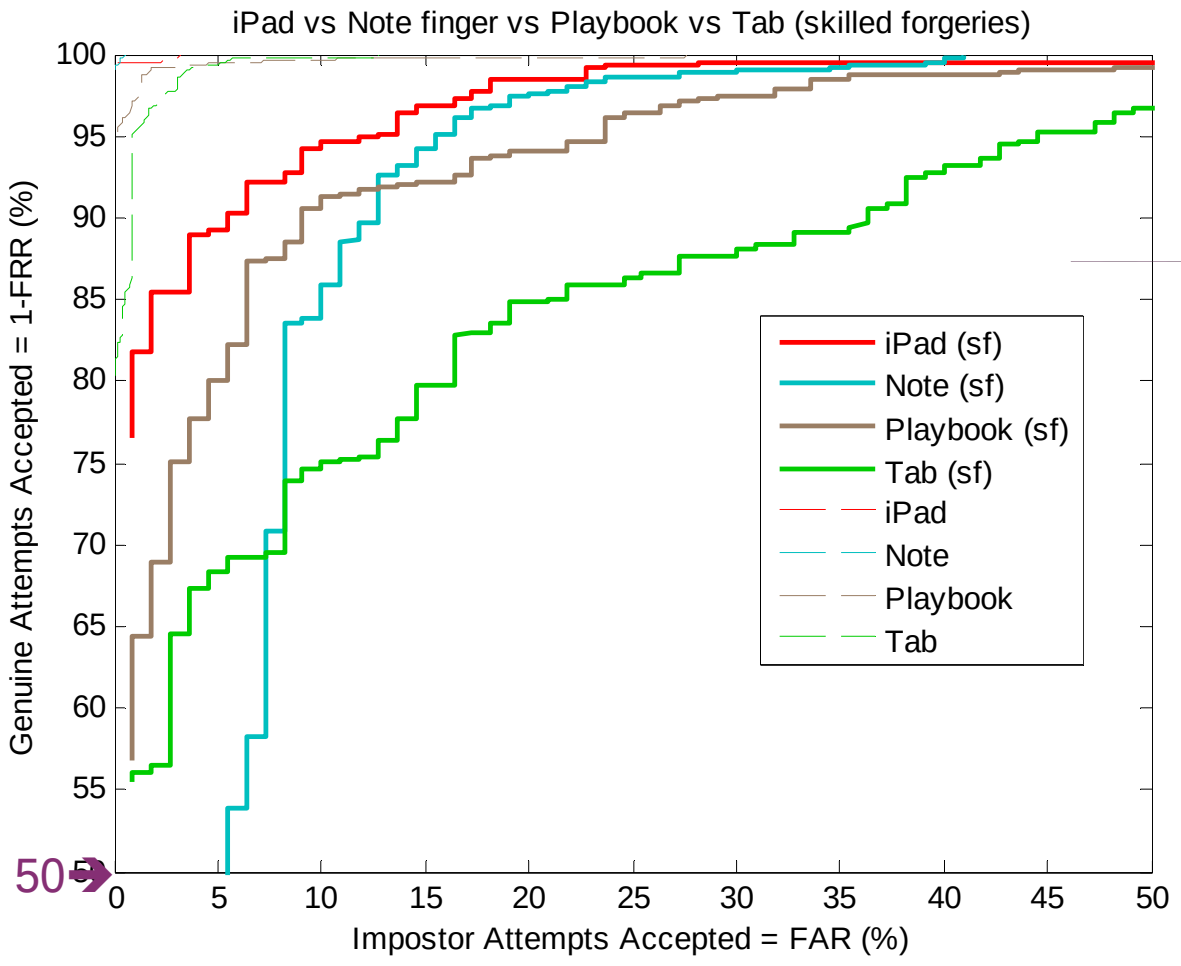
WHAT ABOUT SKILLED FORGERIES? 1

- For each of the stylus-based devices:
  - Note-Stylus:
    - 0.17 ➔ 10.1
  - Asus:
    - 3.48 ➔ 19.0
  - STU:
    - 1.27 ➔ 11.1



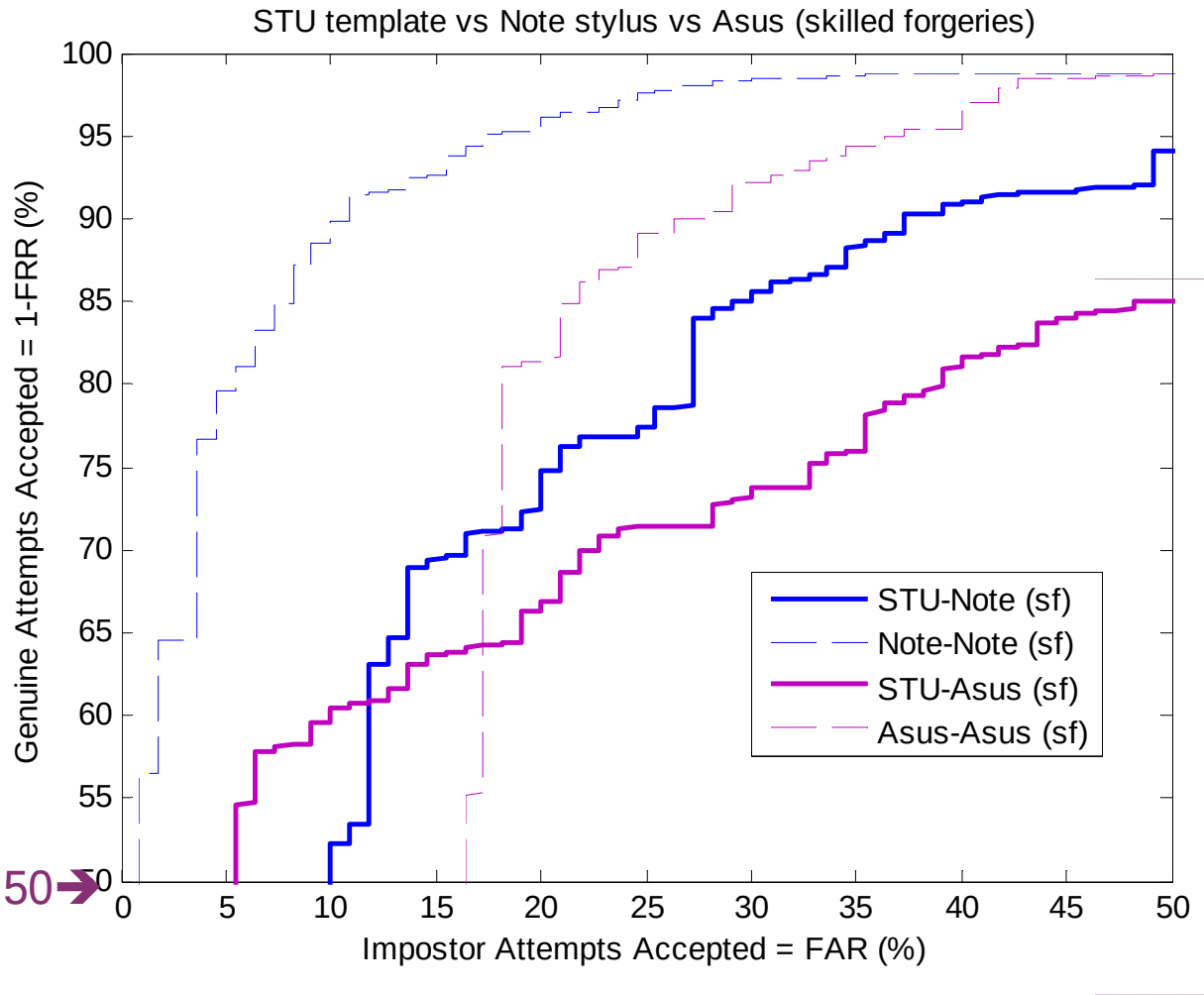
WHAT ABOUT SKILLED FORGERIES? 2

- For each of the finger-based devices:
  - Tab:
    - 2.38 ➔ 17.18
  - Note-F:
    - 0.29 ➔ 10.10
  - Playbook:
    - 1.39 ➔ 9.25
  - iPad:
    - 0.47 ➔ 7.99



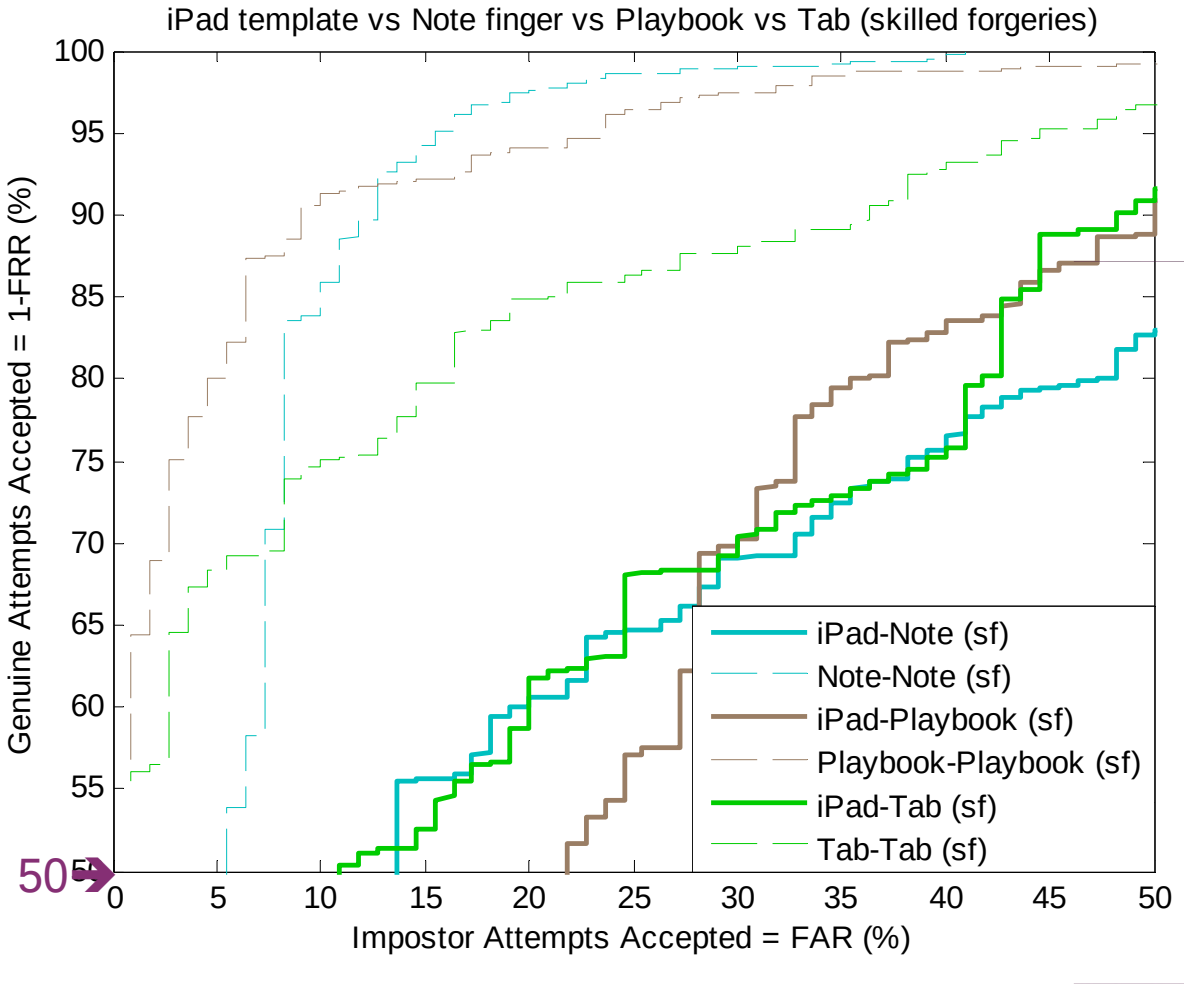
WHAT ABOUT SKILLED FORGERIES? 3

- "Intra-modality" interoperability for stylus-based devices (enrolling with STU):
  - Note-Stylus:
    - 0.98 ➔ 22.9
  - Asus:
    - 6.84 ➔ 28.2



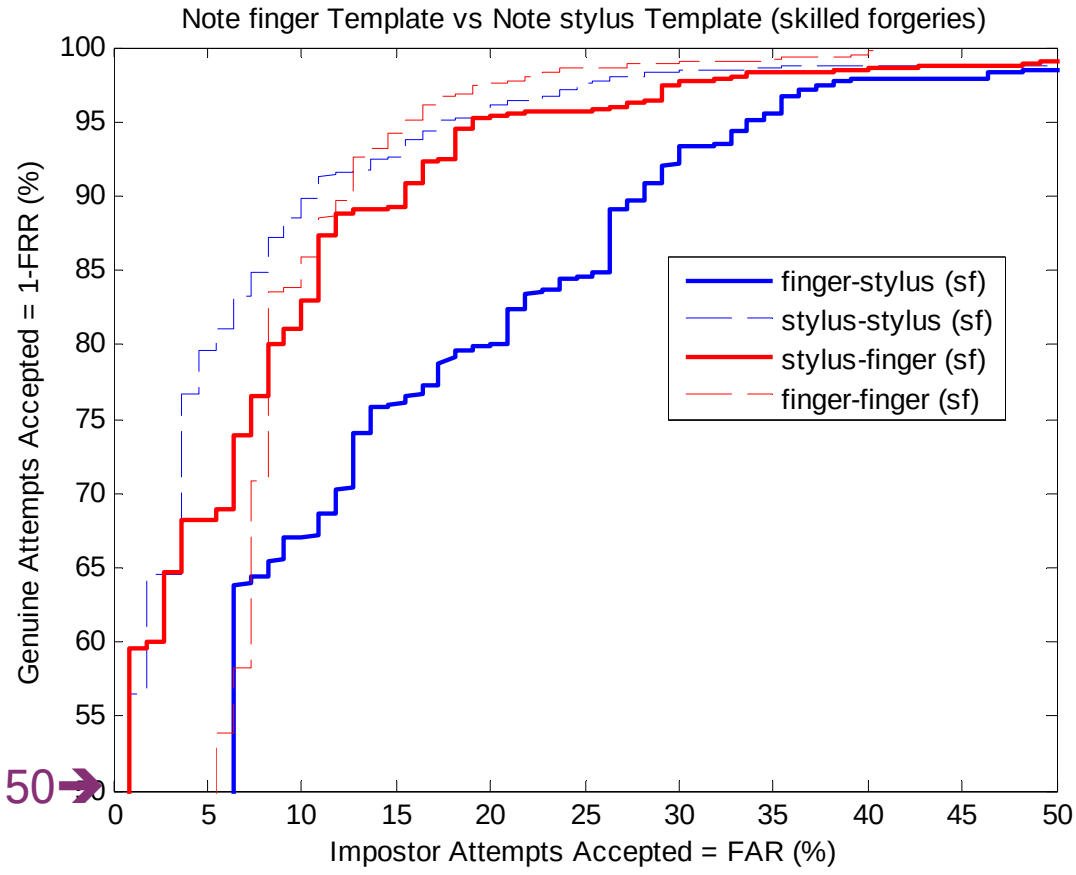
WHAT ABOUT SKILLED FORGERIES? 4

- "Intra-modality" interoperability for finger-based devices (enrolling with iPad):
  - Tab:
    - 3.06 ➔ 29.99
  - Note-F:
    - 3.52 ➔ 30.85
  - Playbook:
    - 1.91 ➔ 30.07



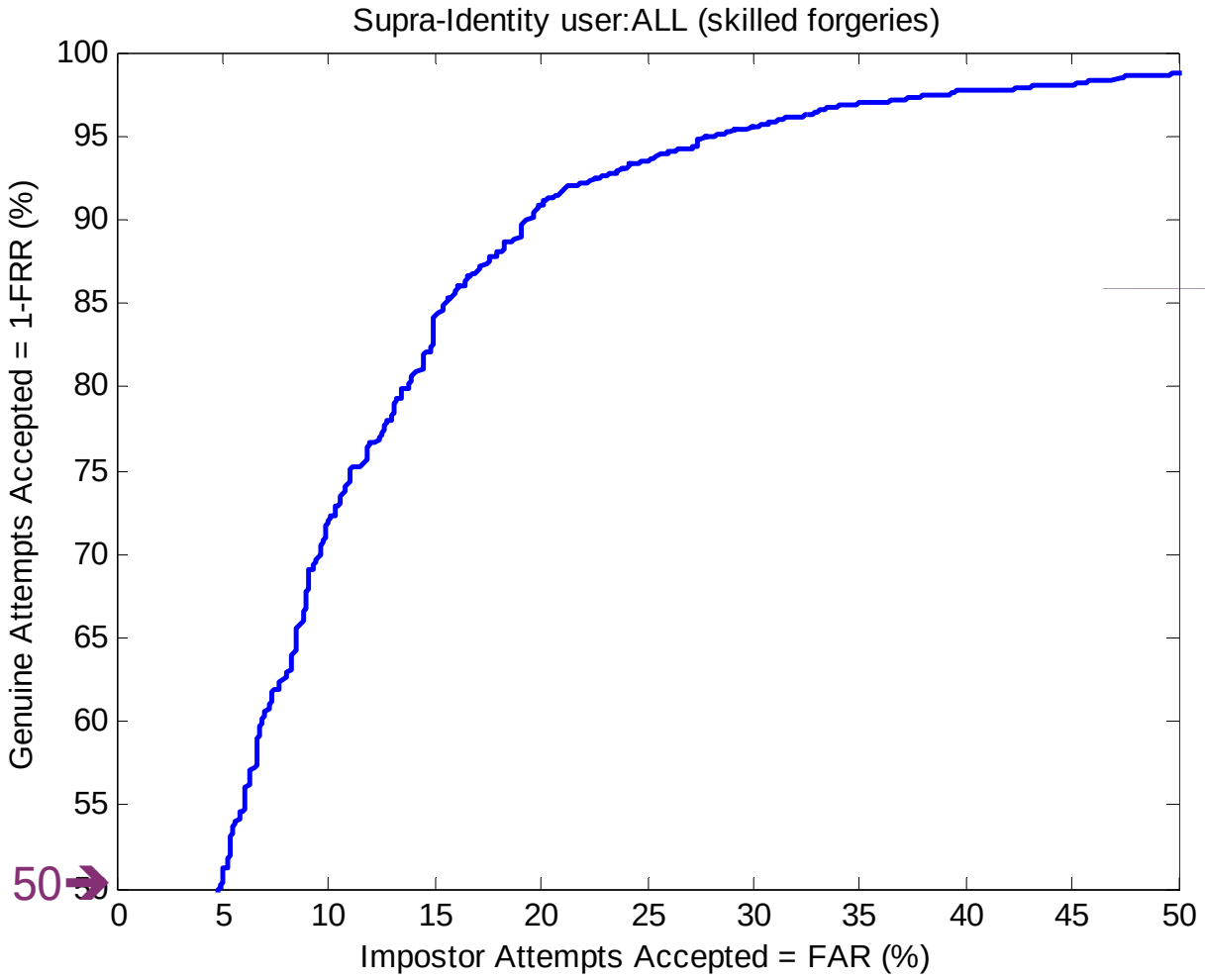
WHAT ABOUT SKILLED FORGERIES? 5

- "Inter-modality" interoperability (Note):
  - Reference Note-F:
    - Note-S: 0.98 ➔ 11.81
  - Reference Note-S:
    - Note-F: 3.52 ➔ 19.97



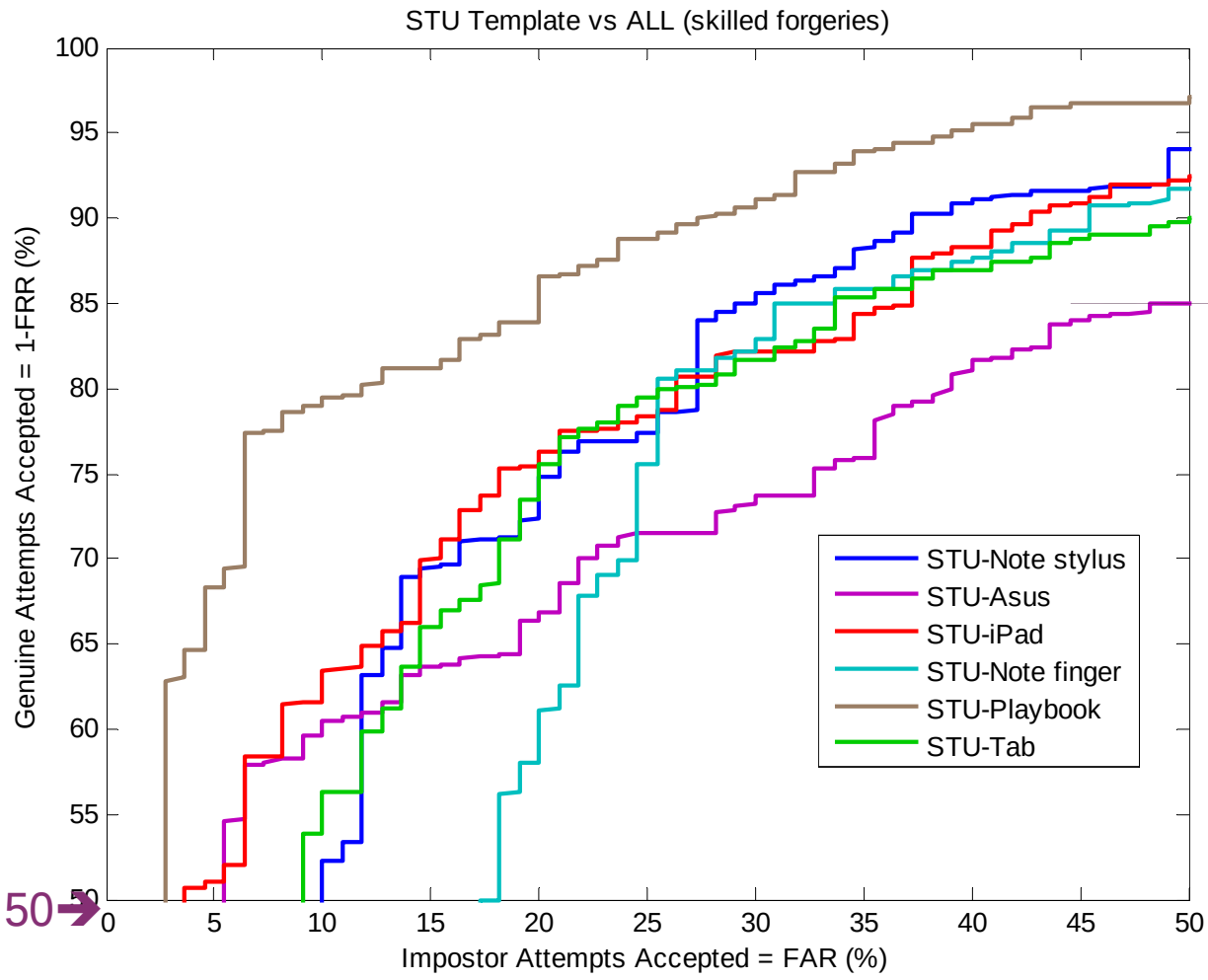
WHAT ABOUT SKILLED FORGERIES? 6

- Supra-identity:
  - 1.66 ➔ 15.33



WHAT ABOUT SKILLED FORGERIES? 7

- "Office enrolment / Multiple device verification" scenario (enrolment with STU):
  - Note-Stylus:
    - 22.93
  - Asus:
    - 28.21
  - iPad:
    - 22.61
  - Note-Finger:
    - 24.55
  - Playbook:
    - 17.17
  - Tab:
    - 22.53



CONCLUSIONS (PERFORMANCE) - 2

- Visual Feedback is a major parameter for performance!
- **No relationship with technology has been found for different behaviours**
- Within genuine samples, interoperability is affordable:
  - Within stylus devices is really good
  - Within finger-based devices in affordable
    - **Potential problem in some enrolments with fingers**
- Results initially shows that a single modality can be considered
  - Interoperability is acceptable
    - Enrolling with stylus results may even get better
- Supra-identity shows really good results:
  - But enrolling is tiring for users
  - And what about future devices being used?
- Skilled Forgeries kill modality performance
  - Worst that State-of-the-Art
    - **May pressure provide the difference?**
  - Intra-device (10% approx.); Intra-modality (>20% for stylus, 30% for finger); Supra-identity (15.33%); Typical scenario (about 20%)
  - The skilled forgeries were obtained with high knowledge on the user's signature
    - **Improvement on algorithm? Or real problem with advanced forgery?**

USABILITY EVALUATION

SPECIFICATIONS

- Devices:
  - ASUS (Tablet PC)
  - STU-500 (Peripheral – Reference Device)
  - iPad (Tablet)
  - HTC Desire (Smartphone)
- Crew: 15 people
  - Age: 16 – 60 years old
  - Other data: Broad-spectrum in familiarity with technology, habituation to sign, etc.
- Sessions: 1
- Signatures/session: 12
- All genuine samples (NO skilled forgeries)
- All users' real signatures
- 5 Scenarios

- Target: Obtain a preliminary idea on the influence of the position for signing
  - First step towards a full usability test (based on HBSI)
- Same algorithm used

SCENARIOS

01
Reference Scenario
(All devices)

02
(HTC, iPad)

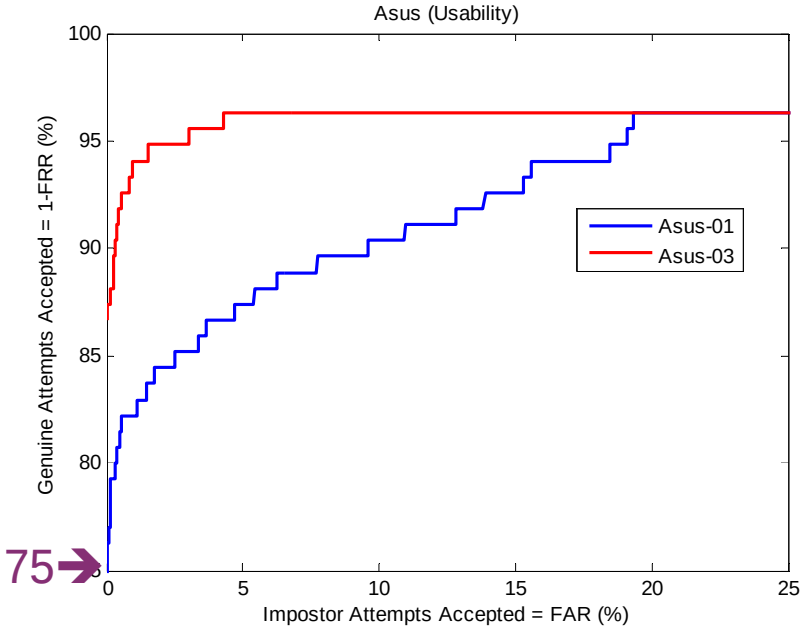
03
(Asus, HTC, STU)

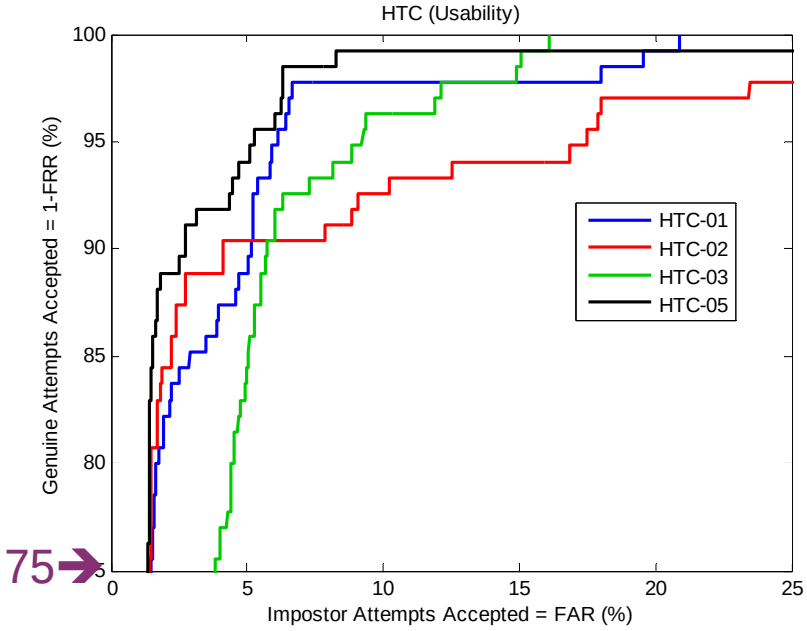
04
(iPad, STU)

05
(HTC, iPad)

SAME DEVICE & SCENARIO - 1

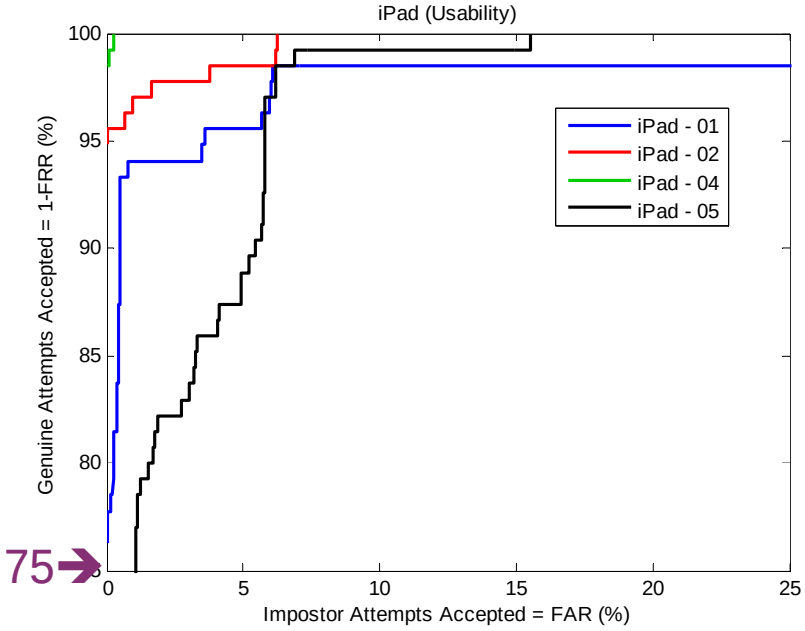
- Comparing enrolling with its own scenario (same device):
  - ASUS:
    - Scenario 01: EER = 9.62 (🡐 3.48)
    - Scenario 03: EER = 4.37
  - HTC:
    - Scenario 01: EER = 5.92 (🡐 2.38 for Tab)
    - Scenario 02: EER = 8.87
    - Scenario 03: EER = 7.35
    - Scenario 05: EER = 5.19

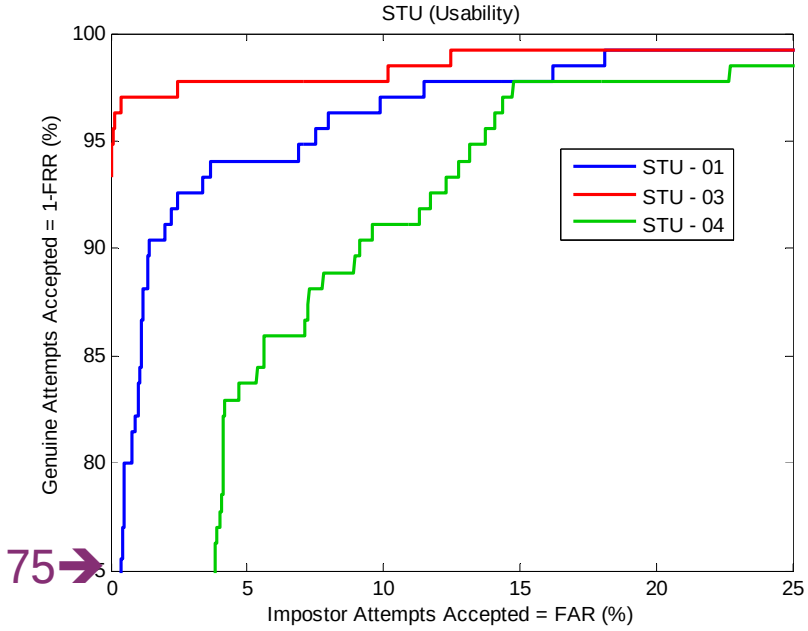




SAME DEVICE & SCENARIO - 2

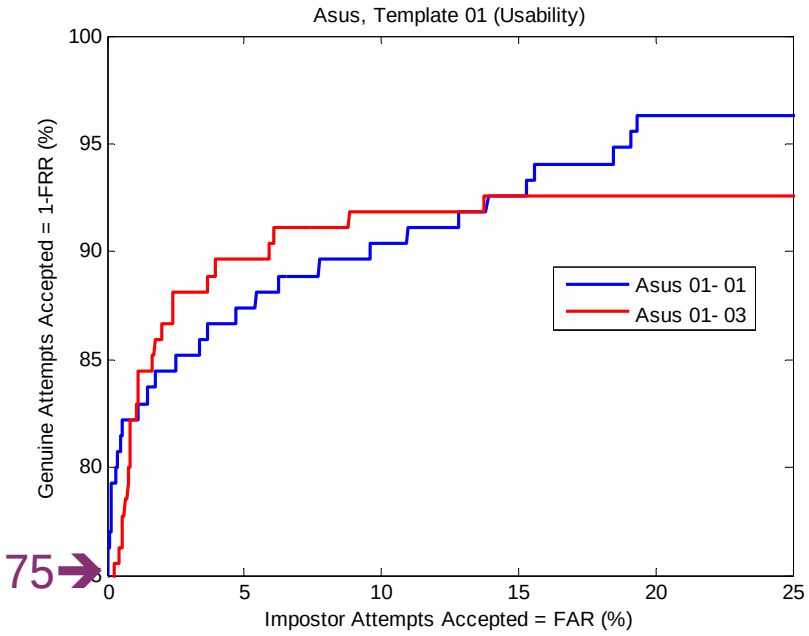
- Comparing enrolling with its own scenario (same device):
  - iPad:
    - Scenario 01: EER = 4.44 (🡐 0.47)
    - Scenario 02: EER = 2.23
    - **Scenario 04: EER = 0.14 !!**
    - Scenario 05: EER = 5.84
  - STU:
    - Scenario 01: EER = 5.92 (🡐 1.27)
    - Scenario 03: EER = 2.35
    - Scenario 04: EER = 9.60

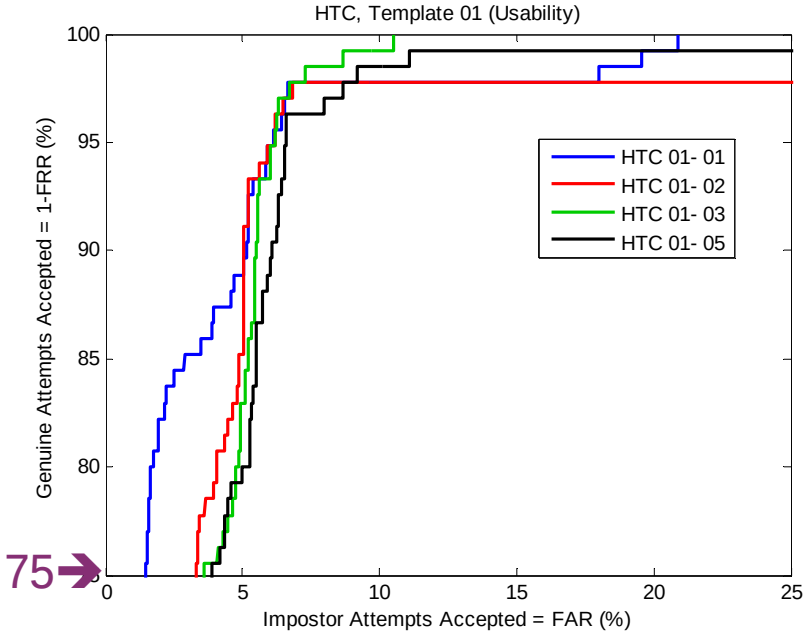




SCENARIO INTEROPERABILITY - 1

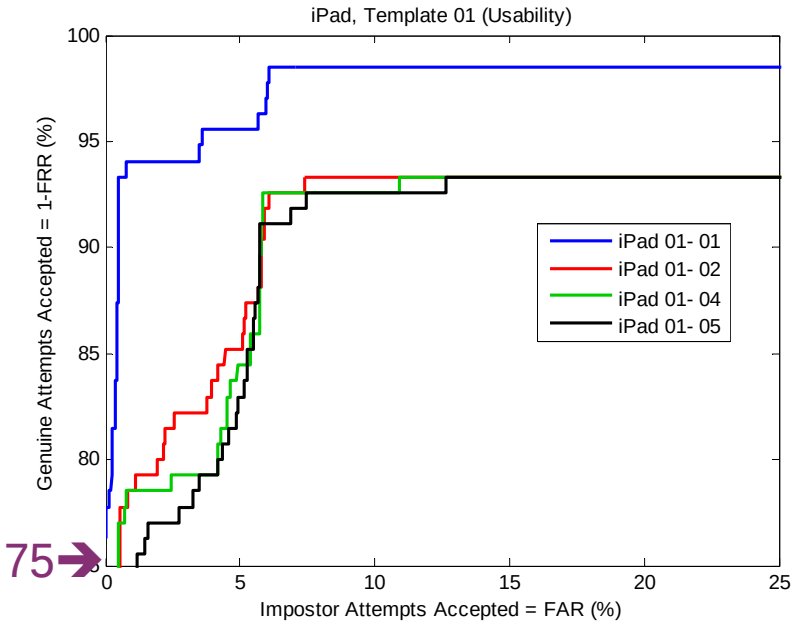
- Comparing enrolling with scenario 01 (same device):
  - ASUS:
    - Scenario 01: EER = 9.62
    - Scenario 03: EER = 4.37 ➔ 8.85
  - HTC:
    - Scenario 01: EER = 5.92
    - **Scenario 02: EER = 8.87 ➔ 5.92 !!**
    - **Scenario 03: EER = 7.35 ➔ 5.96 !!**
    - Scenario 05: EER = 5.19 ➔ 6.60

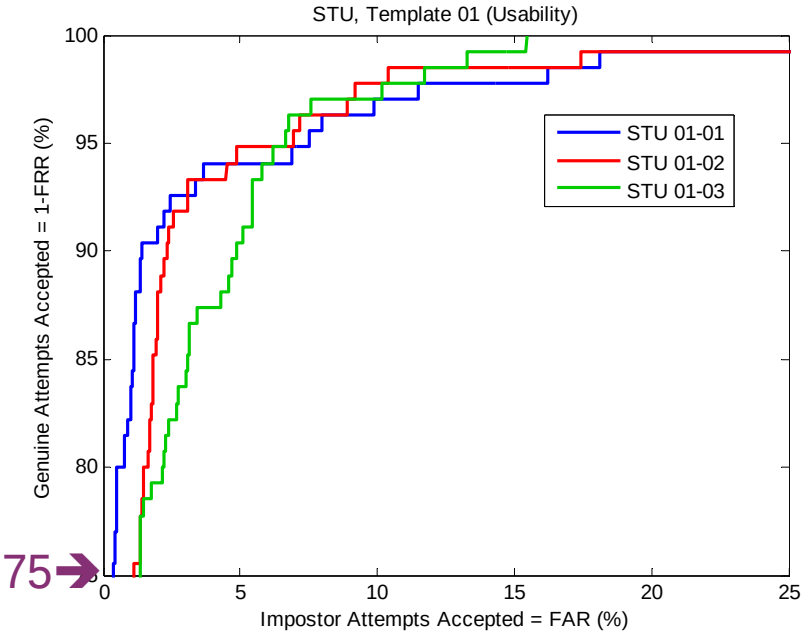




SCENARIO INTEROPERABILITY - 2

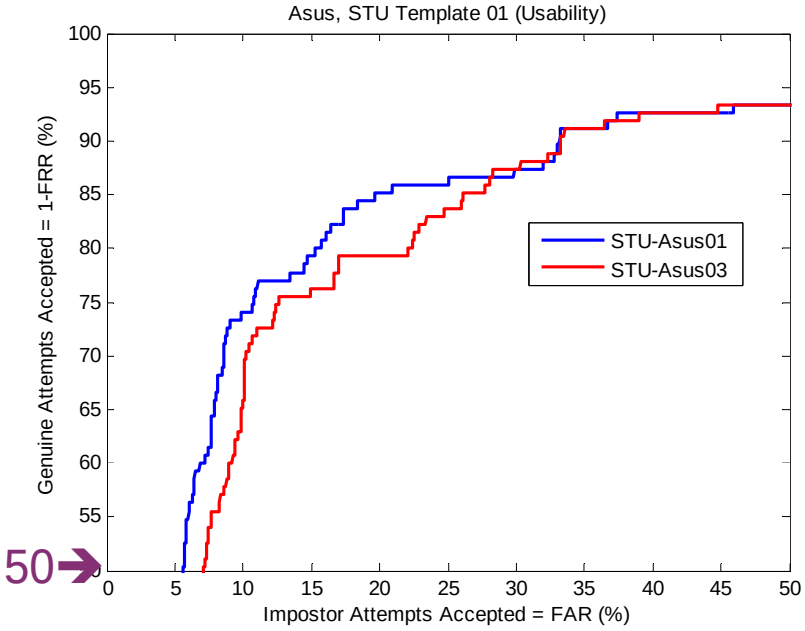
- Comparing enrolling with scenario 01 (same device):
  - iPad:
    - Scenario 01: EER = 4.44
    - Scenario 02: EER = 2.23 ➔ 7.42
    - Scenario 04: EER = 0.14 ➔ 7.42
    - Scenario 05: EER = 5.84 ➔ 7.44
  - STU:
    - Scenario 01: EER = 5.92
    - Scenario 03: EER = 2.35 ➔ 5.19
    - **Scenario 04: EER = 9.60 ➔ 5.92 !!**

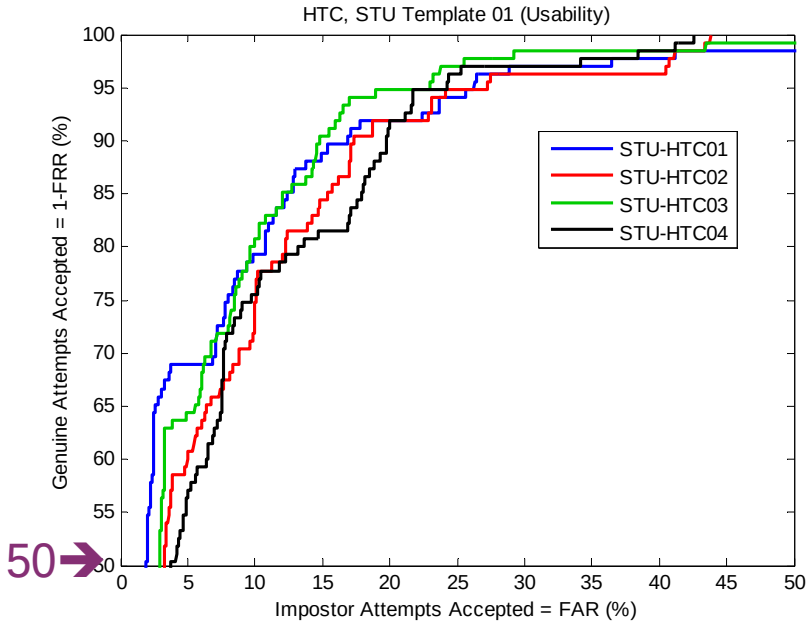




EXPECTED IOP DEPLOYMENT - 1

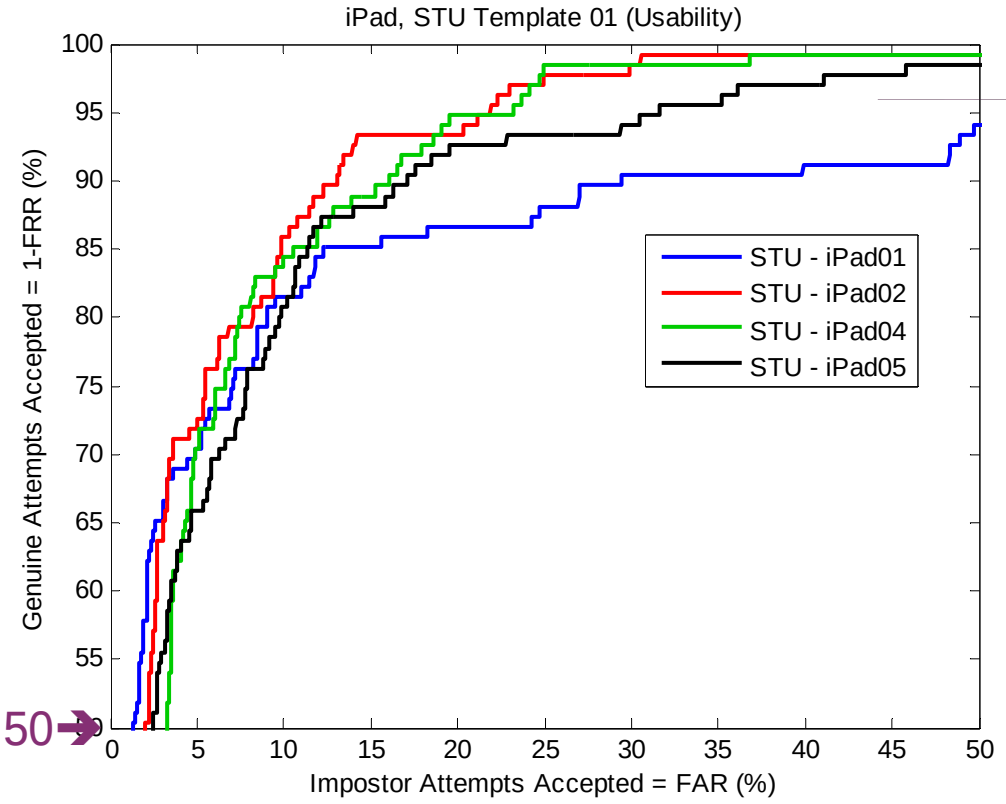
- Comparing enrolling with STU (scenario 01):
  - ASUS:
    - Scenario 01: EER = 9.62 ➔ 17.19
    - Scenario 03: EER = 4.37 ➔ 20.75
  - HTC:
    - Scenario 01: EER = 5.92 ➔ 13.14
    - Scenario 02: EER = 8.87 ➔ 15.49
    - Scenario 03: EER = 7.35 ➔ 13.94
    - Scenario 05: EER = 5.19 ➔ 17.04





EXPECTED IOP DEPLOYMENT - 2

- Comparing enrolling with STU (scenario 01):
  - iPad:
    - Scenario 01: EER = 4.44 ➔ 14.81
    - Scenario 02: EER = 2.23 ➔ 11.76
    - Scenario 04: EER = 0.14 ➔ 12.60
    - Scenario 05: EER = 5.84 ➔ 12.58



CONCLUSIONS (USABILITY)

- A change in the test crew have a major impact in performance
- Smartphones seem to present a lower dependency on the scenario
- Surprisingly the best results are achieved standing and device on a platform
  - Better rates than the reference scenario
  - **May be due to habituation!**
- Results on the 2nd session may show an improvement on the error rates
  - **Enrolment seems to present not stable signatures**
  - **Could be improvement in training a solution?**
- Interoperability (both intra-device and inter-device) show homogeneity on the error rates
- Definitely bad error rates
  - But, … compared against what?
  - **Same kind of evaluation with other modalities?**

FUTURE WORKS!

- A lot!
  - At least all the highlighted text on each of the conclusions
- Increasing size of test crews is a must!
  - For both, Performance and Usability
- Carrying out a real Usability evaluation
  - Following HBSI Methodology
- This kind of studies will lead to the need of further improvement in Scenario and Operational Testing
  - May be new standards to arise in the following years

THANKS! QUESTIONS?

R. Sanchez-Reillo, R. Blanco-Gonzalo, O. Miguel-Hurtado, A. Mendaza-Ormaza

University Group for Identification Technologies (GUTI)

Carlos III University of Madrid

http://guti.uc3m.es

http://idtestinglab.uc3m.es

rsreillo@ing.uc3m.es