The Effect of Length on Key Fingerprint Verification Security & Usability

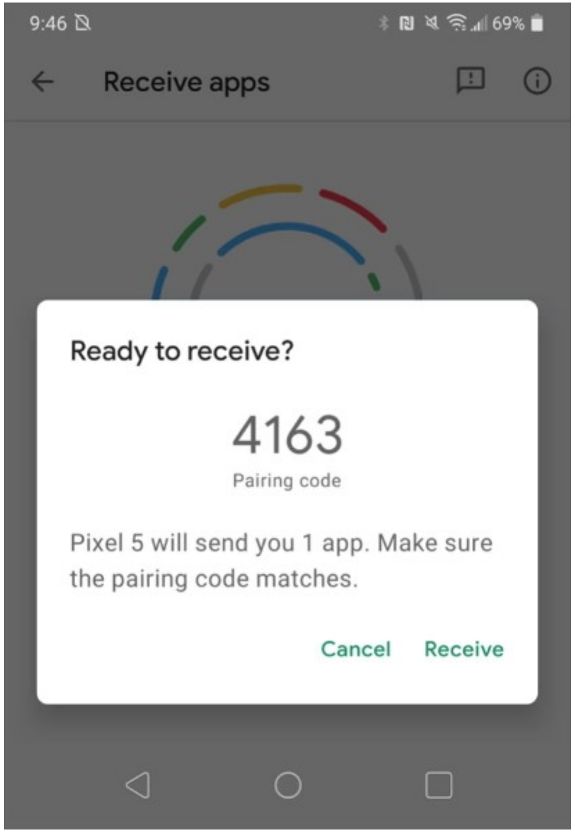
Dan Turner, Siamak F. Shahandashti, Helen Petrie
ePrint: `arxiv.org/abs/2306.04574`

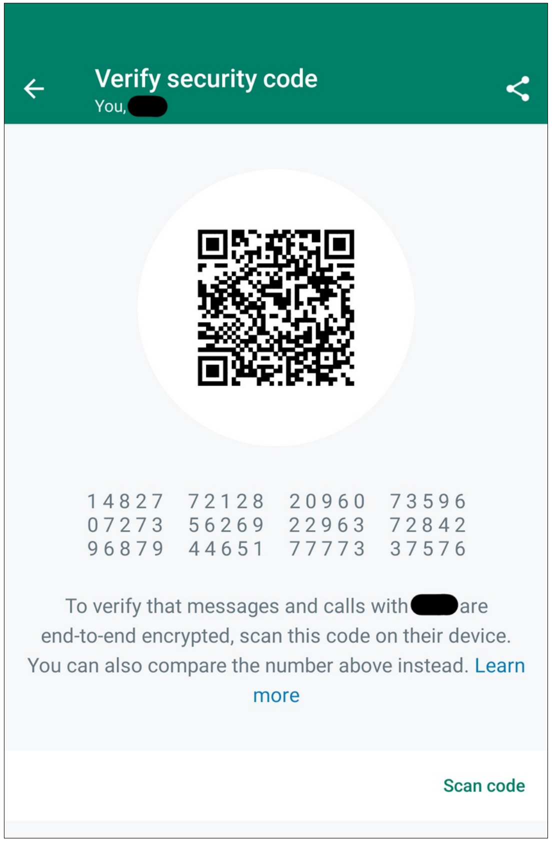
UNIVERSITY of York

Paper ePrint

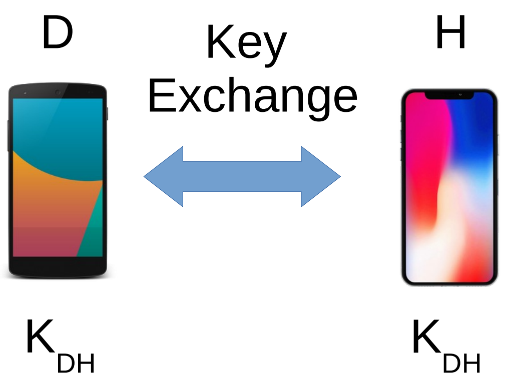
The Context

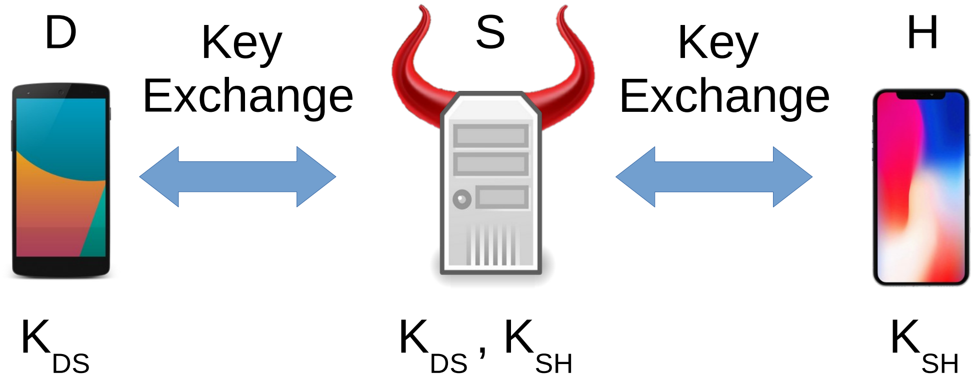
Key Fingerprint Verification



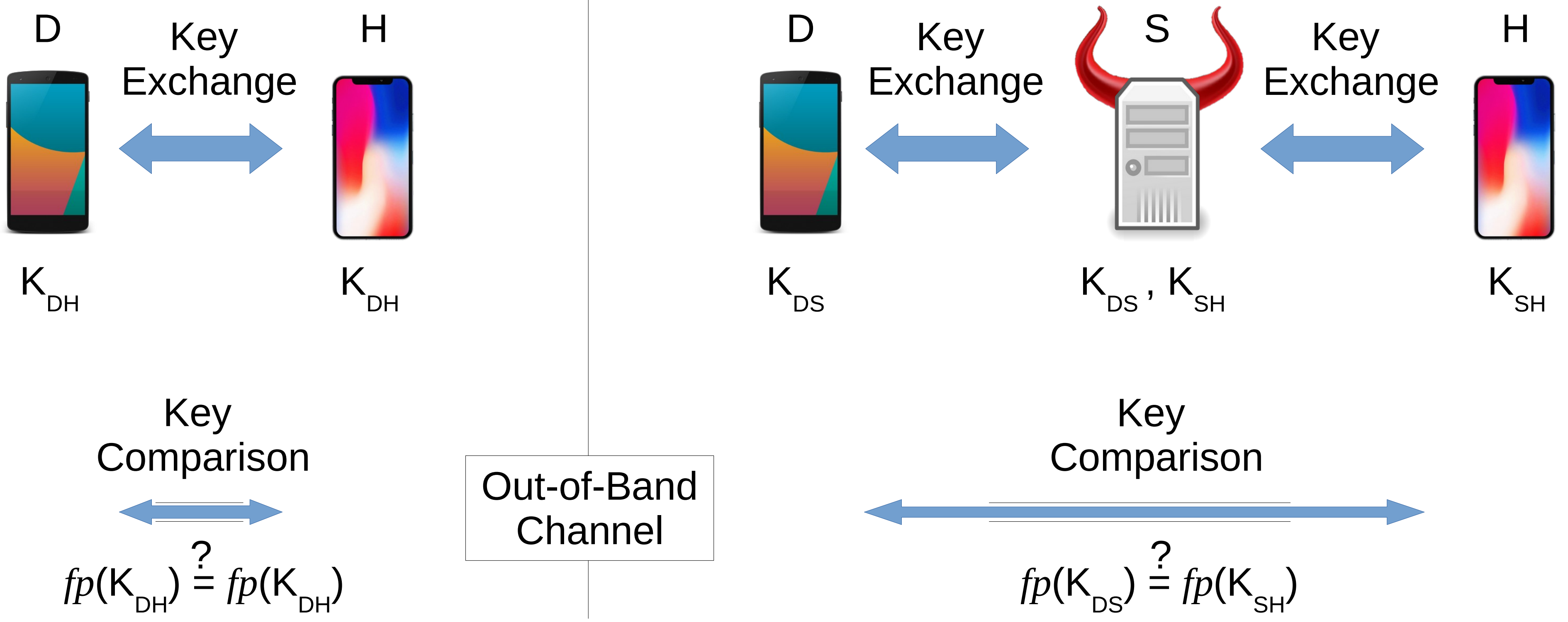


Adversary in the Middle (AitM) Attacks



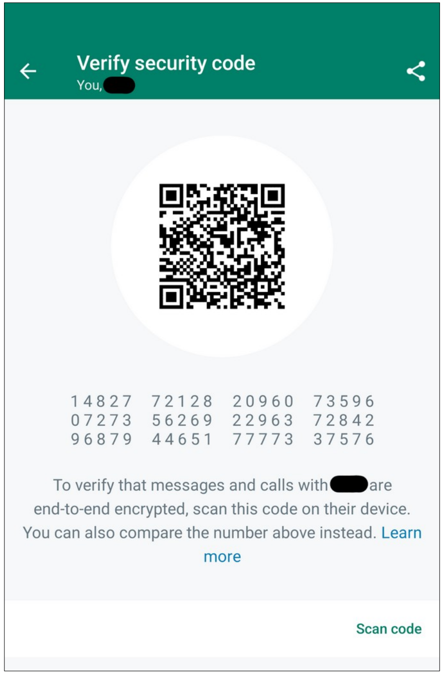


Detection of AitM Attacks



Key Fingerprint Comparison Task

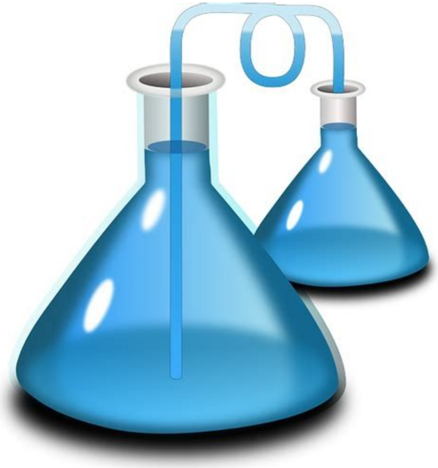
- Ideally needs to be done in an *automated* way
  - e.g. QR code scanning
  - Only *(fully) matching* fingerprints will pass
- When not possible, needs to be done *manually*
  - *Nearly matching* fingerprints may pass as well
  - The focus of this work



Key Fingerprint Variations

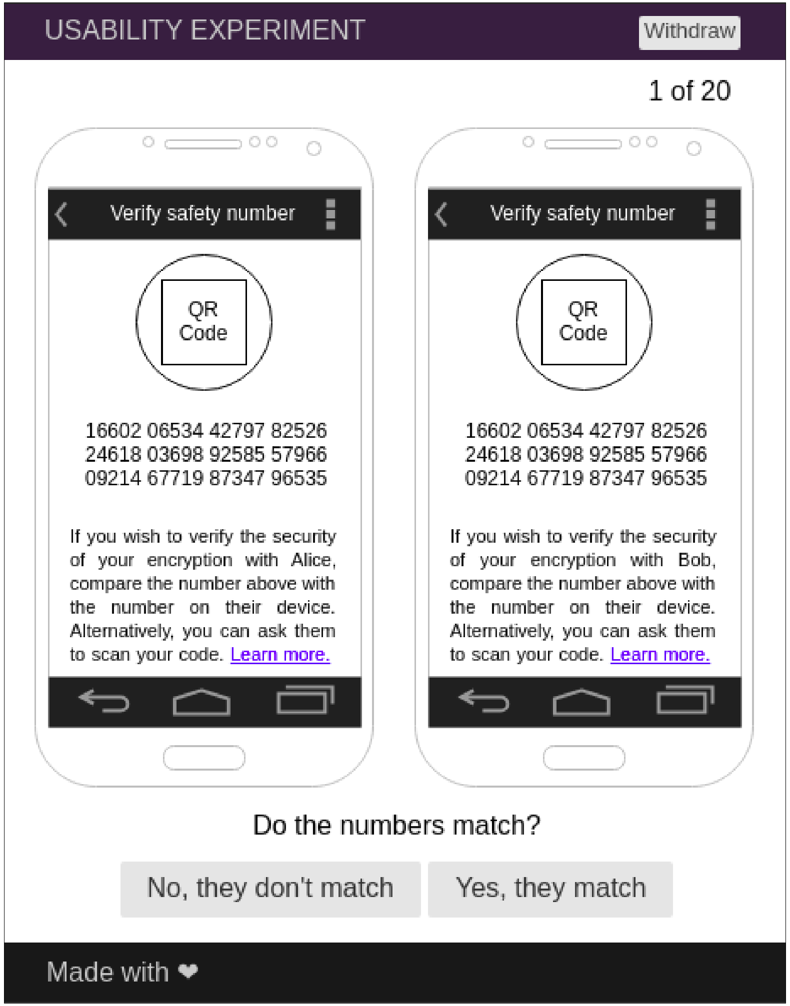
- Format
  - *(Alpha)numeric*, e.g. Signal / WhatsApp, Open PGP, SAS
  - *Words* or *sentences*, e.g. Pretty Easy Privacy
  - *Graphical*, e.g. ASCII art, snowflakes, unicorns
- Comparison mode, e.g. visual or auditory
- Length, e.g. 60 digits for Signal / WhatsApp, 2 words for SAS

The Study



Study Design

- Signal / WhatsApp numeric key fingerprints
- *Conditions*: 1, 2, 3 Line(s) corresponding to 20, 40, 60 digits
  - Between participants: each does 1 length
- *Types*: Safe (matching), Adversarial (nearly matching, 1 chunk diff), Random
  - Within participants: each does 12+4+4 in random order



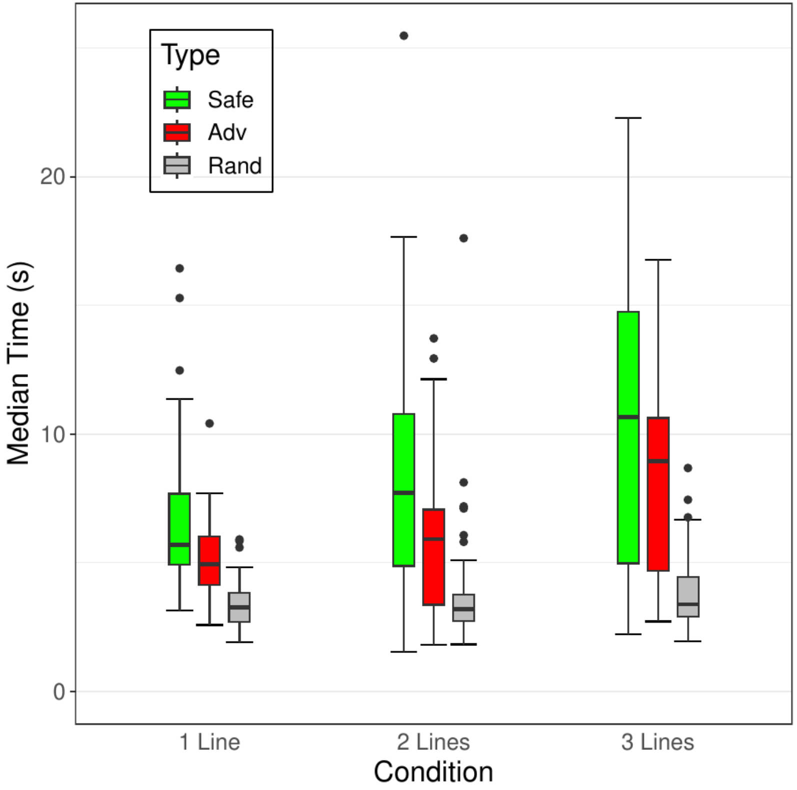
Tested Hypotheses

- **H(t~l)**: longer key → longer comparison time
  - 3 type-specific hypotheses for safe, adv., rand. fingerprints
- **H(t~s)**: higher similarity → longer comparison time
  - 3 length-specific hypotheses for 1L, 2L, 3L fingerprints
- **H(e~l)**: longer key → more errors
  - 2 hypotheses: false acceptance / rejection errors

The Results

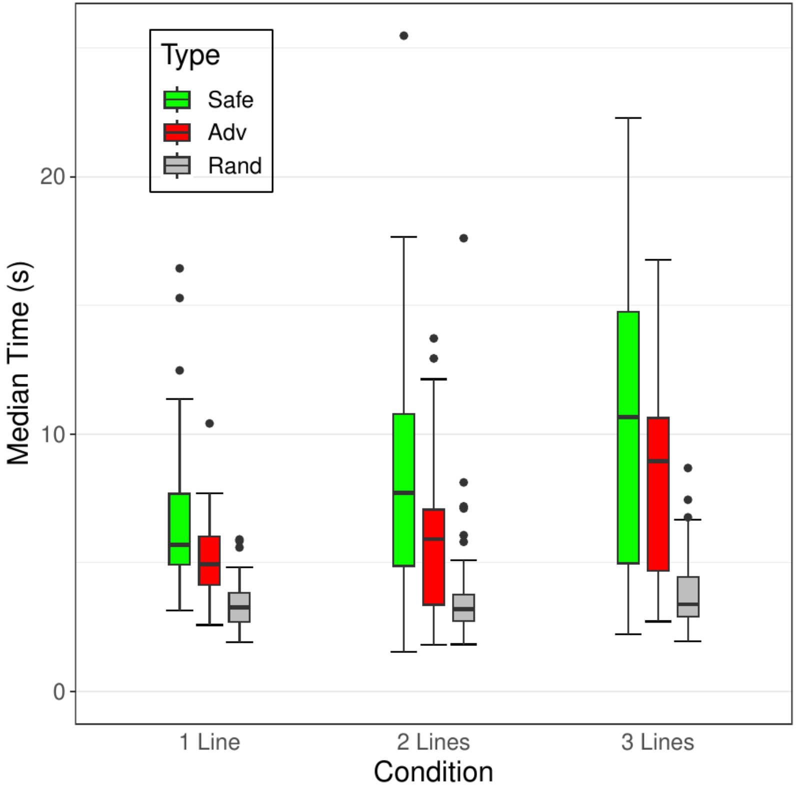
Effect of Length on Comparison Time

- Longer key → longer comparison time: broadly yes, except for Rand
- Kruskal–Wallis + Wilcoxon (Holm)
  - Safe: significant diff 1L–2L–3L
  - Adv: significant diff 1L–3L, 2L–3L
  - Rand: no significant diff



Effect of Type on Comparison Time

- Higher similarity → longer comparison time: emphatic yes
- Friedman + Nemenyi post hoc
  - 1L, 2L, 3L: significant diff safe–adv–rand
- Strong evidence of ‘short-circuit evaluation’



Effect of Length on False Rejection Rate

- Longer key → more errors: Not really for FRE
- Kruskal–Wallis
  - No significant diff b/w lengths

➔ Users are quite efficient & effective in recognising dissimilar fingerprints

#errors	1L	2L	3L
0	92%	85%	80%
1	6%	9%	19%
2–6	0–2%	0–2%	0–2%
7–12	0%	0%	0%

Length	1L	2L	3L
Lower Limit	0.3%	1.6%	1.1%
Mean Rate	0.9%	2.7%	2.0%
Upper Limit	2.0%	4.3%	3.4%

Effect of Length on False Acceptance Rate

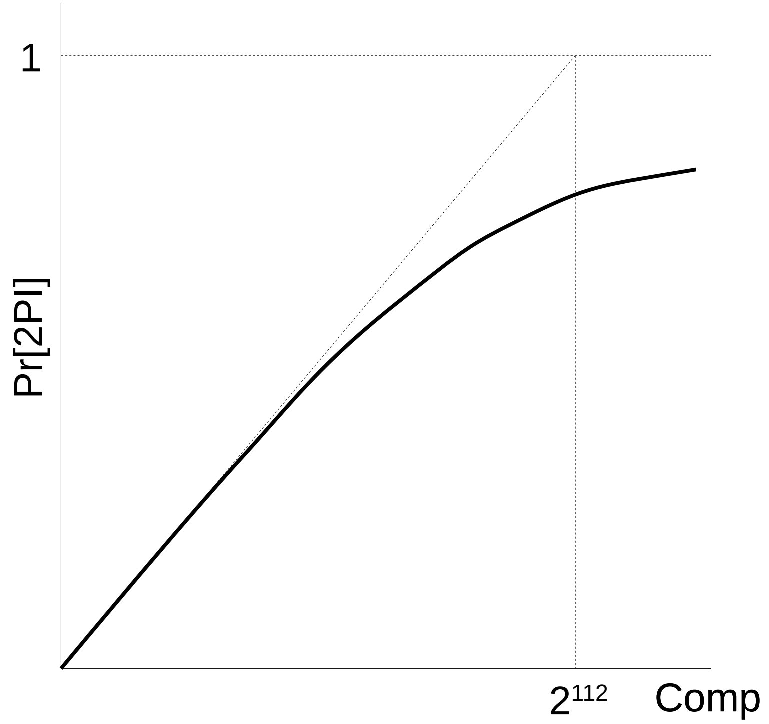
- Longer key → more errors: broadly yes for FAE
- Kruskal–Wallis + Wilcoxon (Holm)
  - Significant diff 1L–3L
- ➔ Users are neither efficient nor effective in comparing highly similar long fingerprints

#errors	1L	2L	3L
0	72%	55%	39%
1	15%	13%	15%
2	8%	9%	11%
3	0%	2%	4%
4	6%	22%	31%

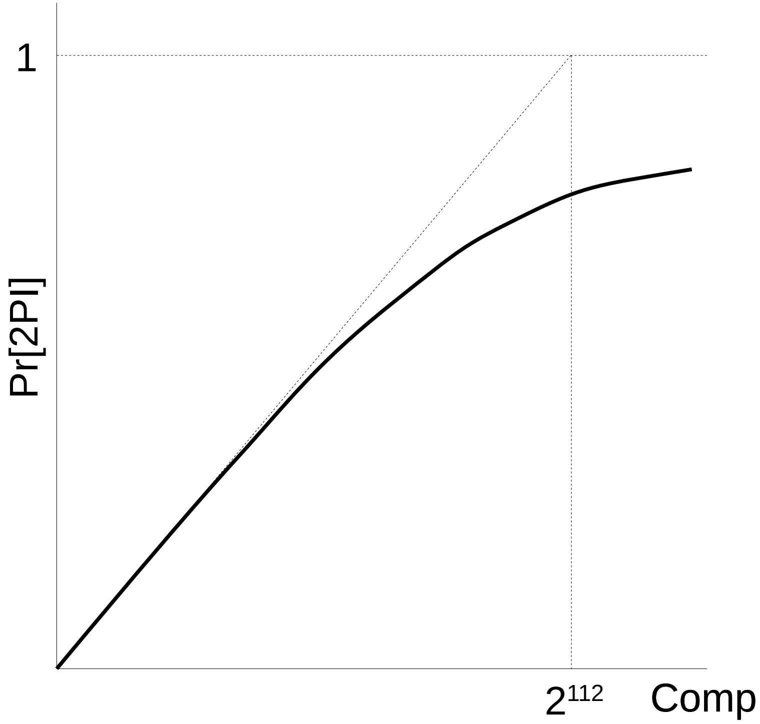
Length	1L	2L	3L
Lower Limit	9%	25%	37%
Mean Rate	13%	31%	44%
Upper Limit	19%	38%	50%

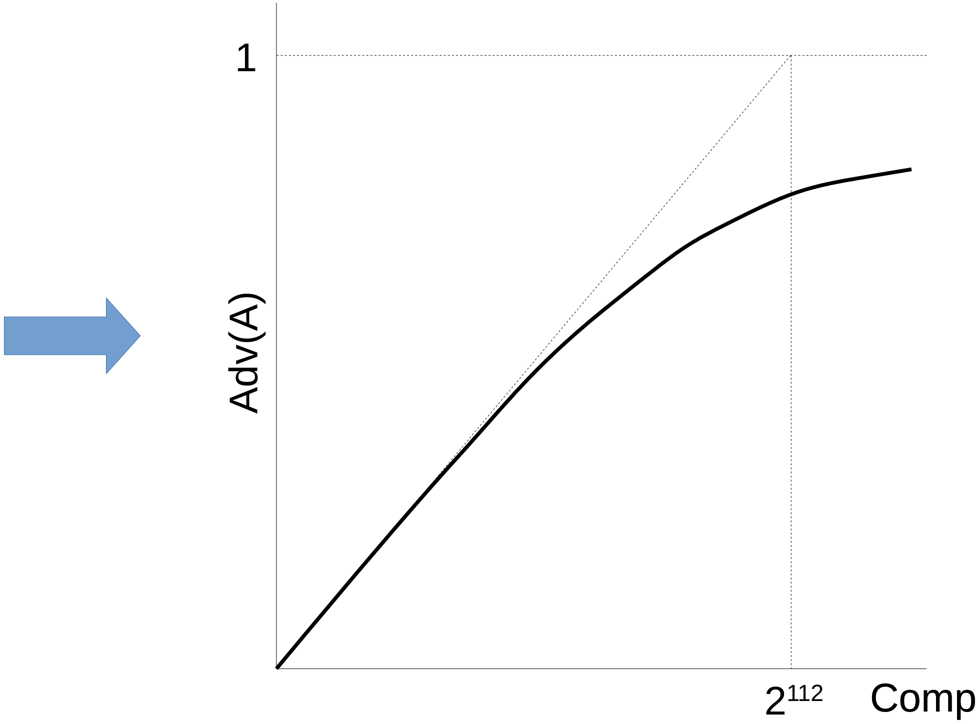
The Security Implications

(Full) 2nd Preimage Attack: Finding 2PI

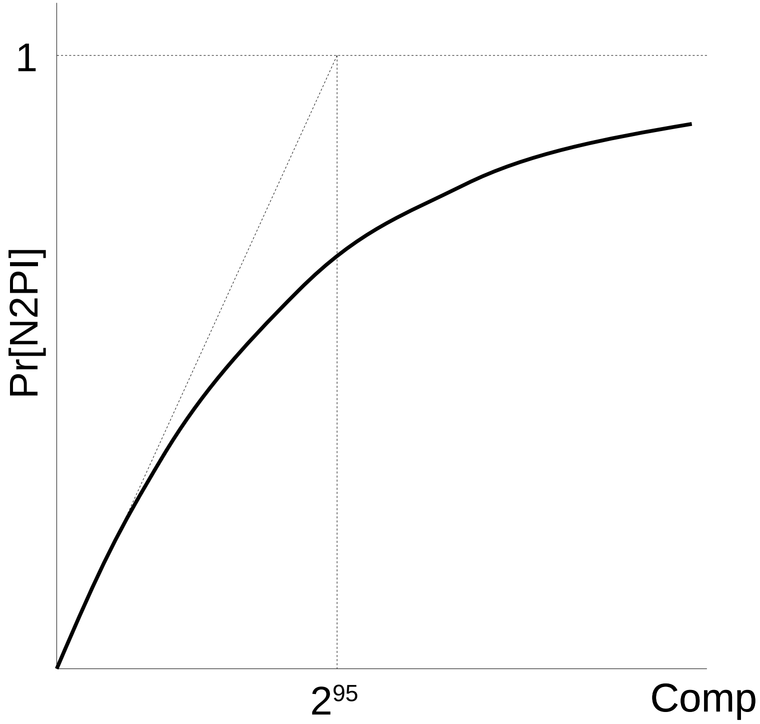


(Full) 2nd Preimage Attack: Overall Success

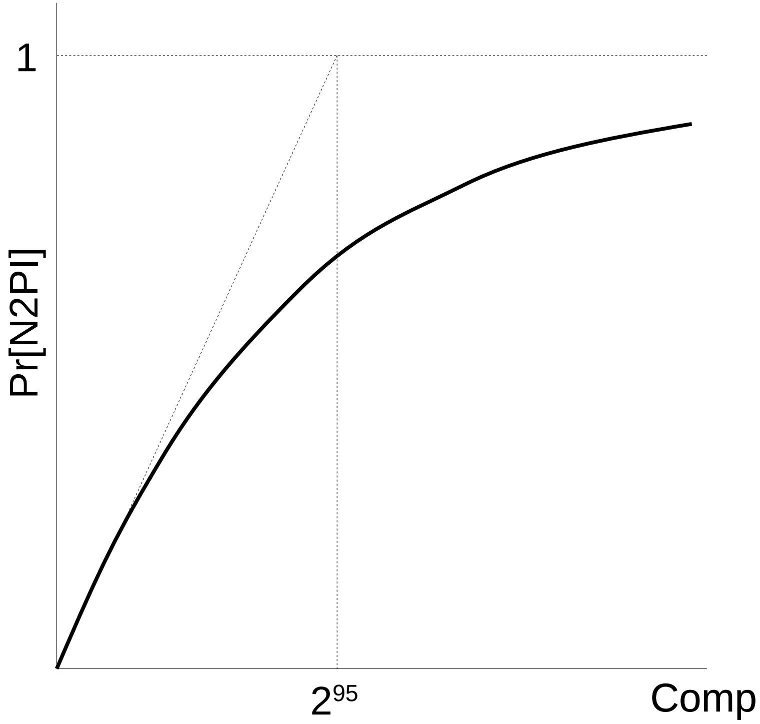


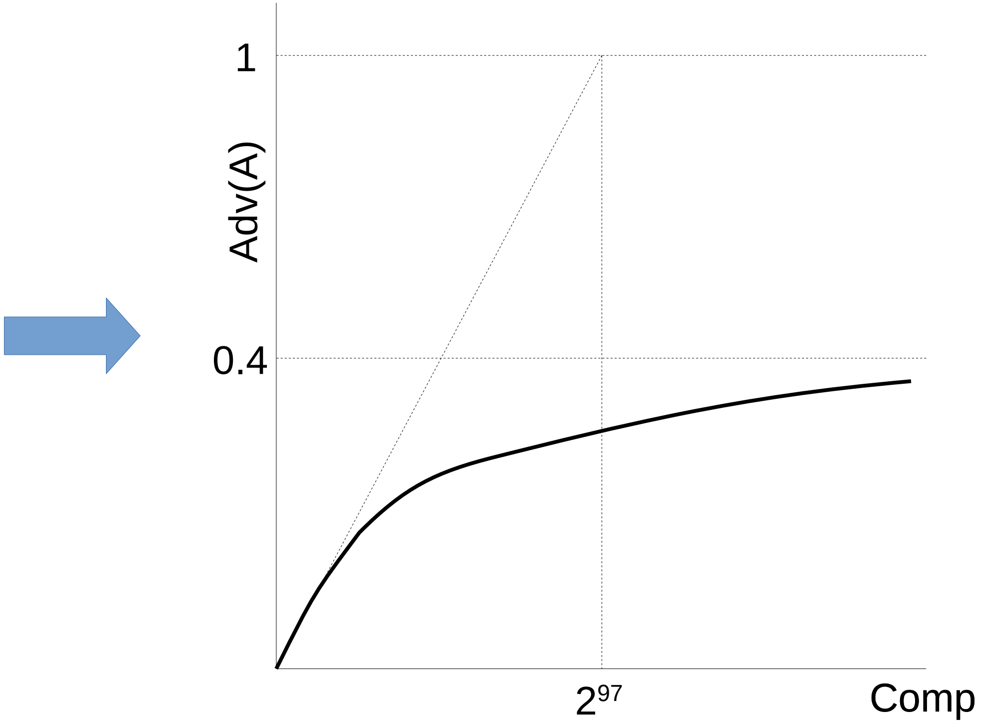


Near 2nd Preimage Attack: Finding N2PI



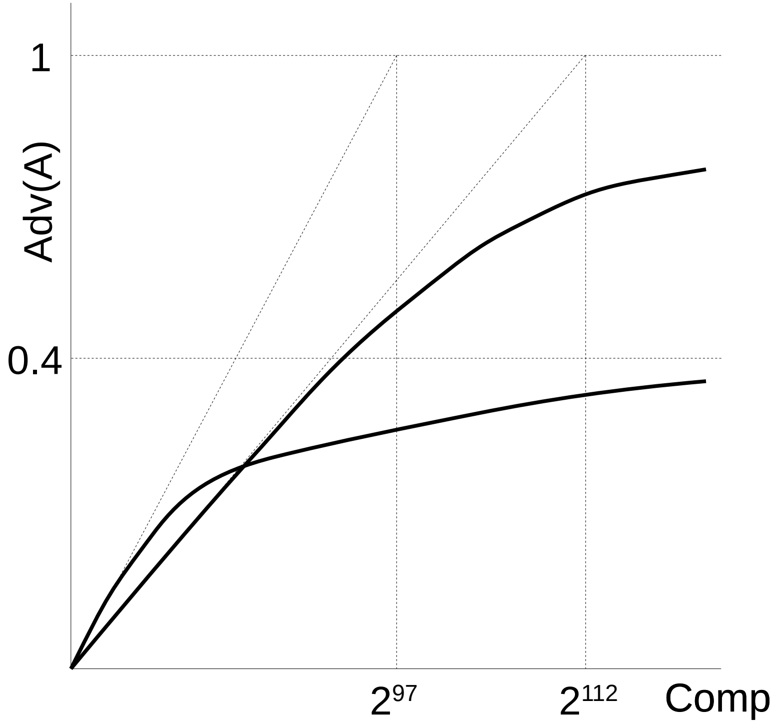
Near 2nd Preimage Attack: Overall Success





Implication of Results on Security

- For adversaries with lower computational budget, manual key fingerprint verification provides a lower security level than usually assumed



Thank you.

Siamak F. Shahandashti
cs.york.ac.uk/~siamak

@SiamakFS

UNIVERSITY of York