Application-driven Cipher Design: (2) New Ciphers for MPC and FHE

Christian Rechberger, DTU

Another often quoted myth

“Crypto is dead”, Adi Shamir, RSA 2013

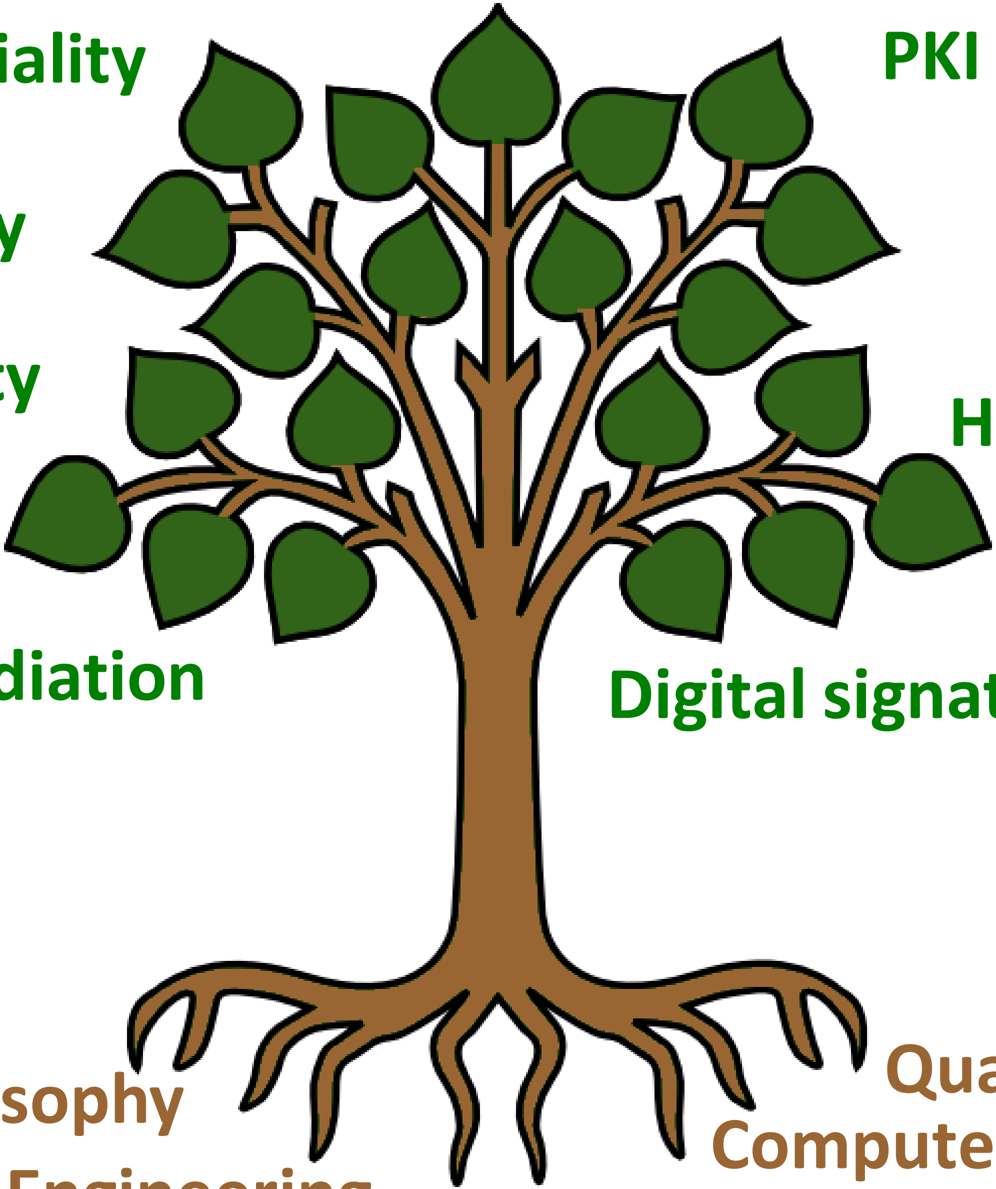
The bad news

Traditional crypto often can not solve real-world problems

Example:“Key theft/loss”, Ron Rivest, Crypto 2011

The good news

- New developments in crypto: MPC, FHE can help to remedy the situation
- Moving from mere theoretical results on to practicability
- Distributed cryptography has first applications, efficiency gains needed to allow for many more
- Symmetric crypto was (so far) outside this development



Confidentiality

Integrity

Authenticity

Non-Repudiation

PKI

Ciphers

Hash functions

Digital signatures

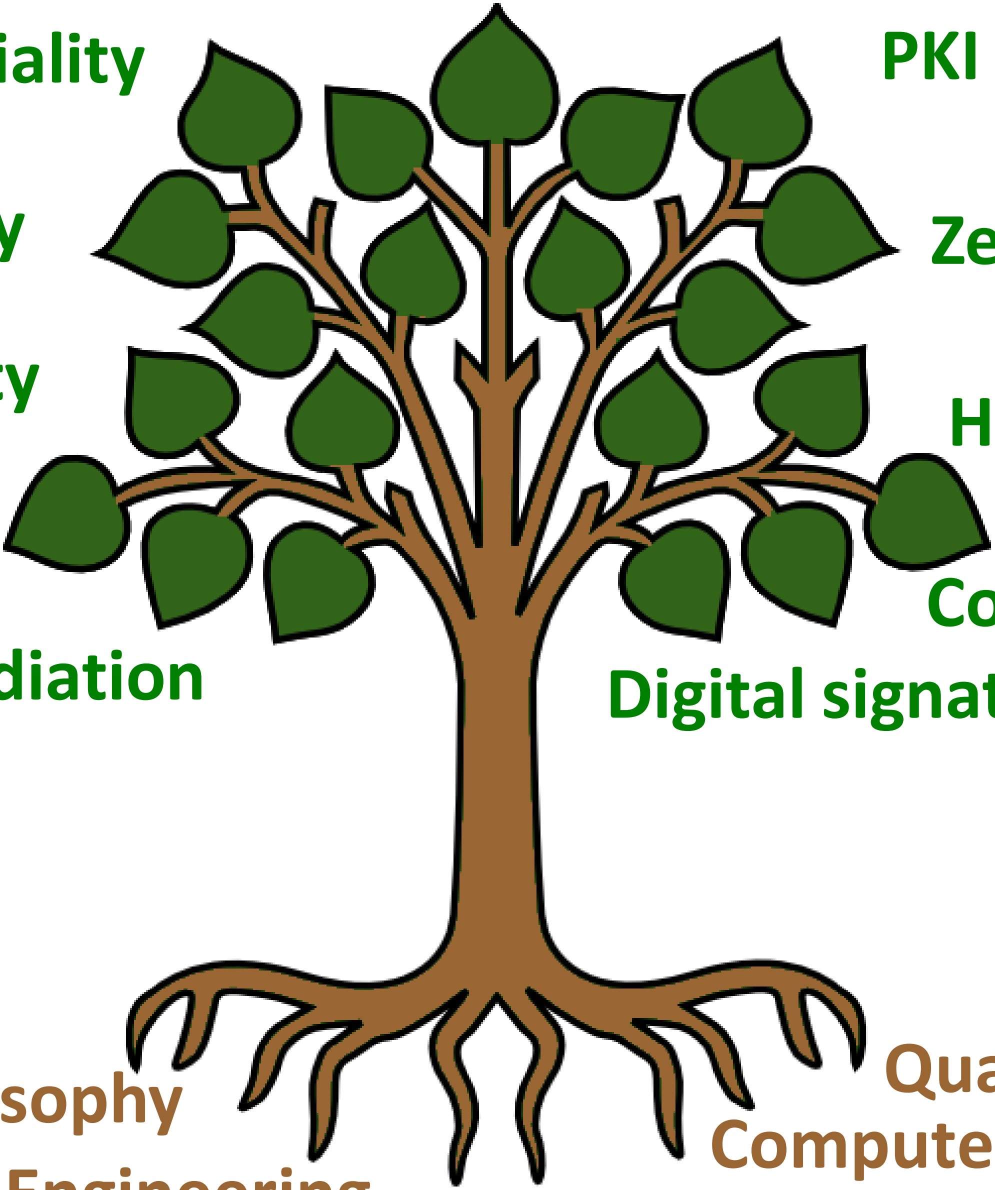
Philosophy

Electrical Engineering

Mathematics

Computer Science

Quantum Physics



Confidentiality

Integrity

Authenticity

FHE

Non-Repudiation

PKI

Ciphers

Zero-Knowledge Proofs

Hash functions

Secure Computation

Digital signatures

Philosophy

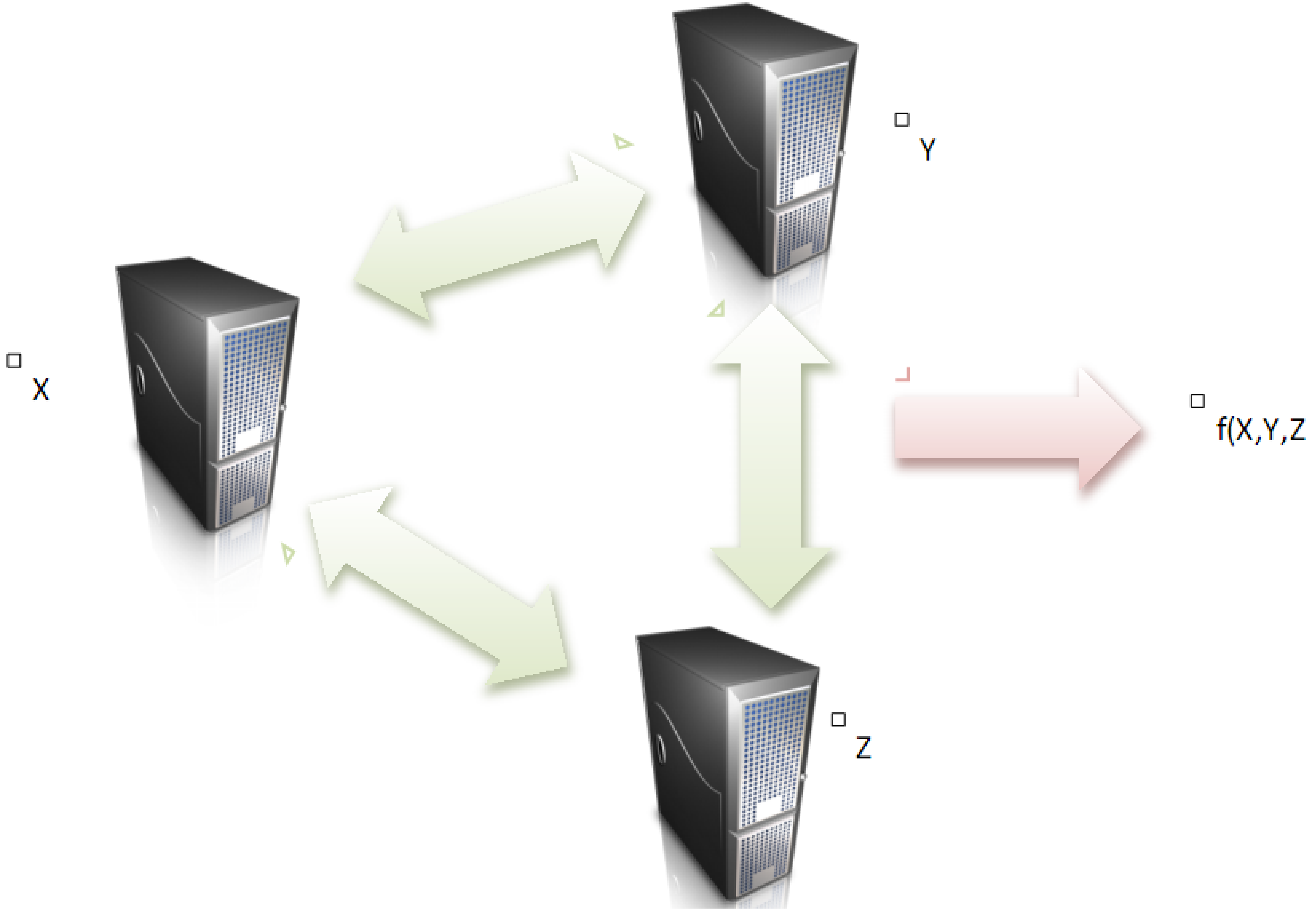
Electrical Engineering

Mathematics

Computer Science

Quantum Physics

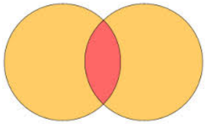
MPC



MPC applications

Block ciphers have various applications in MPC

- **Server-side one-time passwords**, commercialized by Dyadic Security (server-side derivation of one-time passwords via MPC)



- Oblivious Pseudorandom Functions (OPRFs) for **privacy-preserving keyword search**, **private set intersection**, **secure database join**, etc.

- **Secure storage**: store symmetrically encrypted intermediate MPC values in untrusted storage

AES circuit is used a lot

- Often protocols need PRF evaluations
- AES is the standard choice for that
- Designed in 1997, standardized in 2001
- Novel security arguments (proofs) against powerful classes of attacks

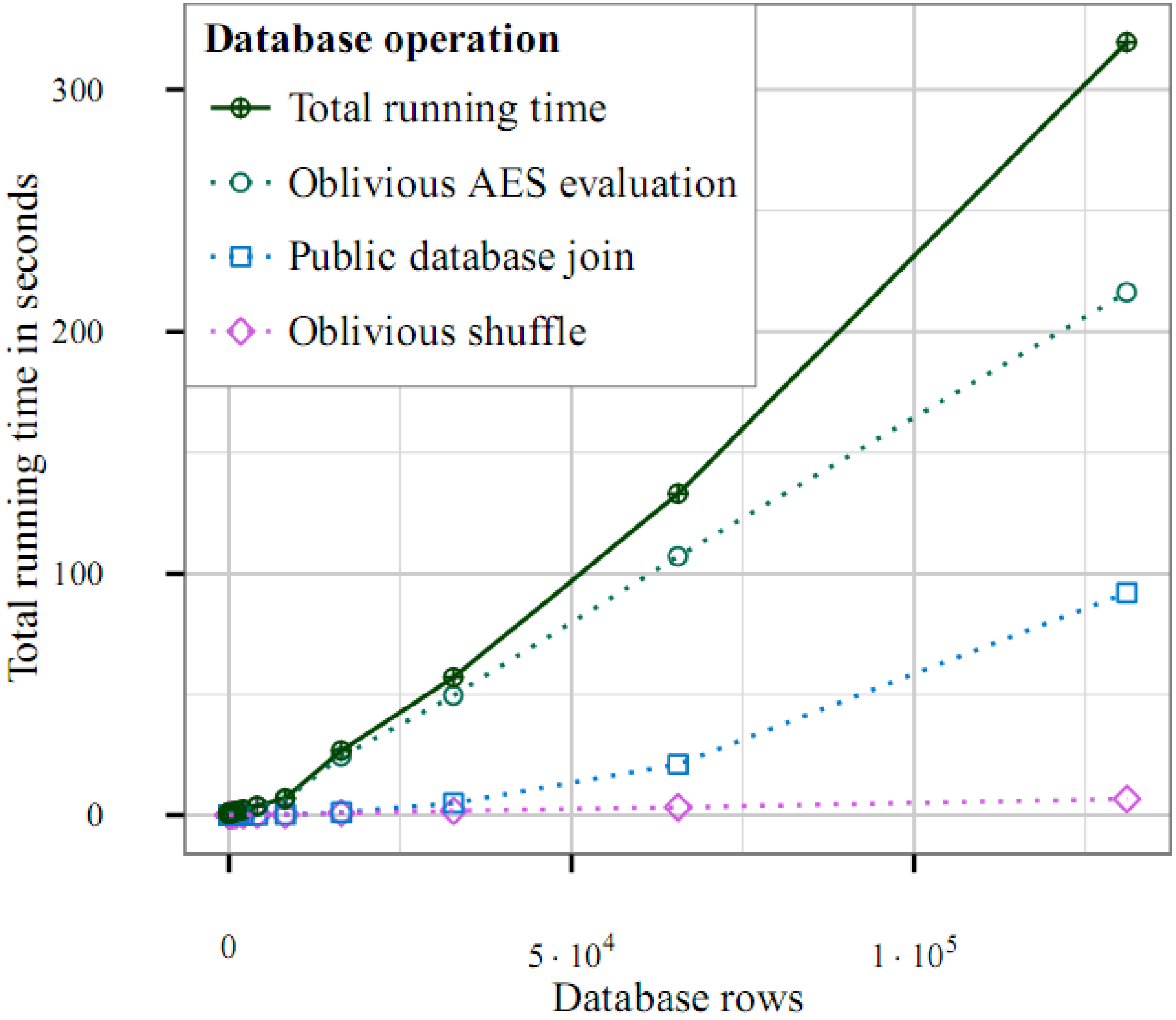
Application: Secure database join, three parties

Way to combine several data sources in privacy preserving manner

Source: Cybernetica

Application:

Merging databases from two different ministries in Estonia, while obeying various data-protection laws.



FHE

Protecting Data via Encryption:

Homomorphic encryption

1. Put your gold in a locked box.
2. Keep the key.
3. Let your jeweler work on it through a glove box.
4. Unlock the box when the jeweler is done!

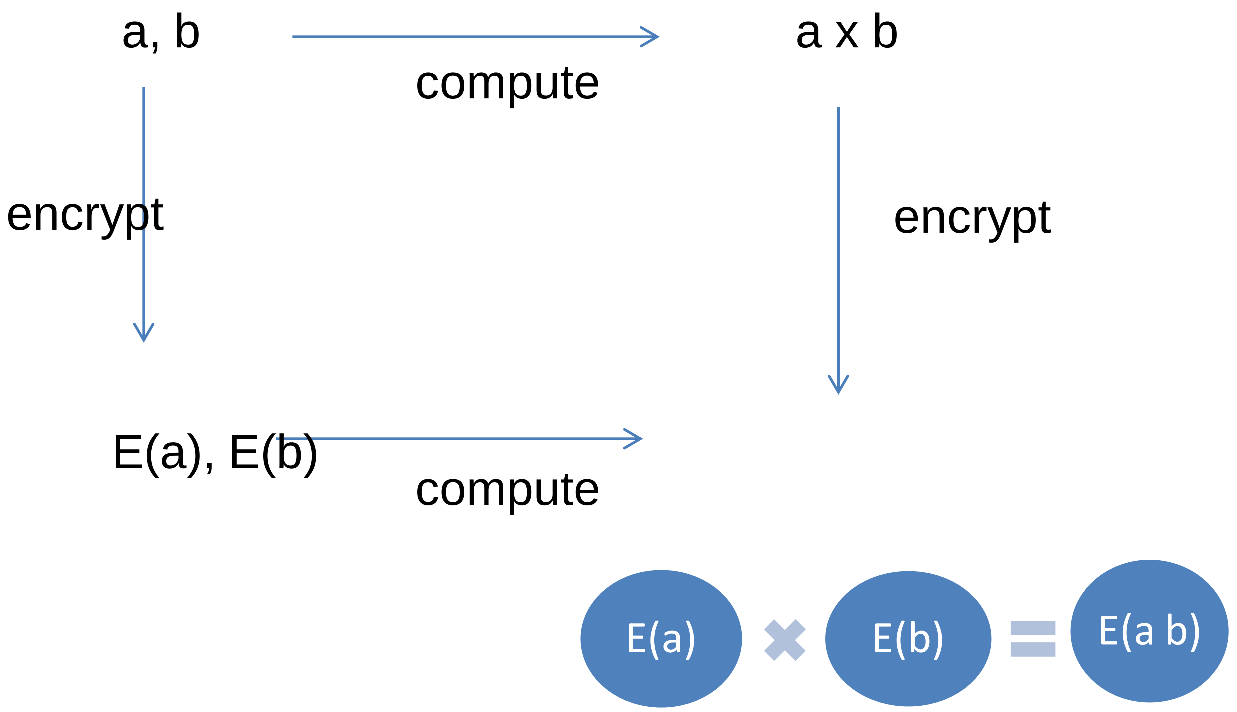
Homomorphic Encryption: addition

$$\begin{tikzcd}
a,\ b \arrow[r, "\text{compute}"'] \arrow[d, "\text{encrypt}"'] & a+b \arrow[d, "\text{encrypt}"] \\
E(a),\ E(b) \arrow[r, "\text{compute}"'] & {}
\end{tikzcd}$$

$E(a) + E(b) = E(a+b)$

Homomorphic Encryption:

multiplication



a, b

compute

a x b

encrypt

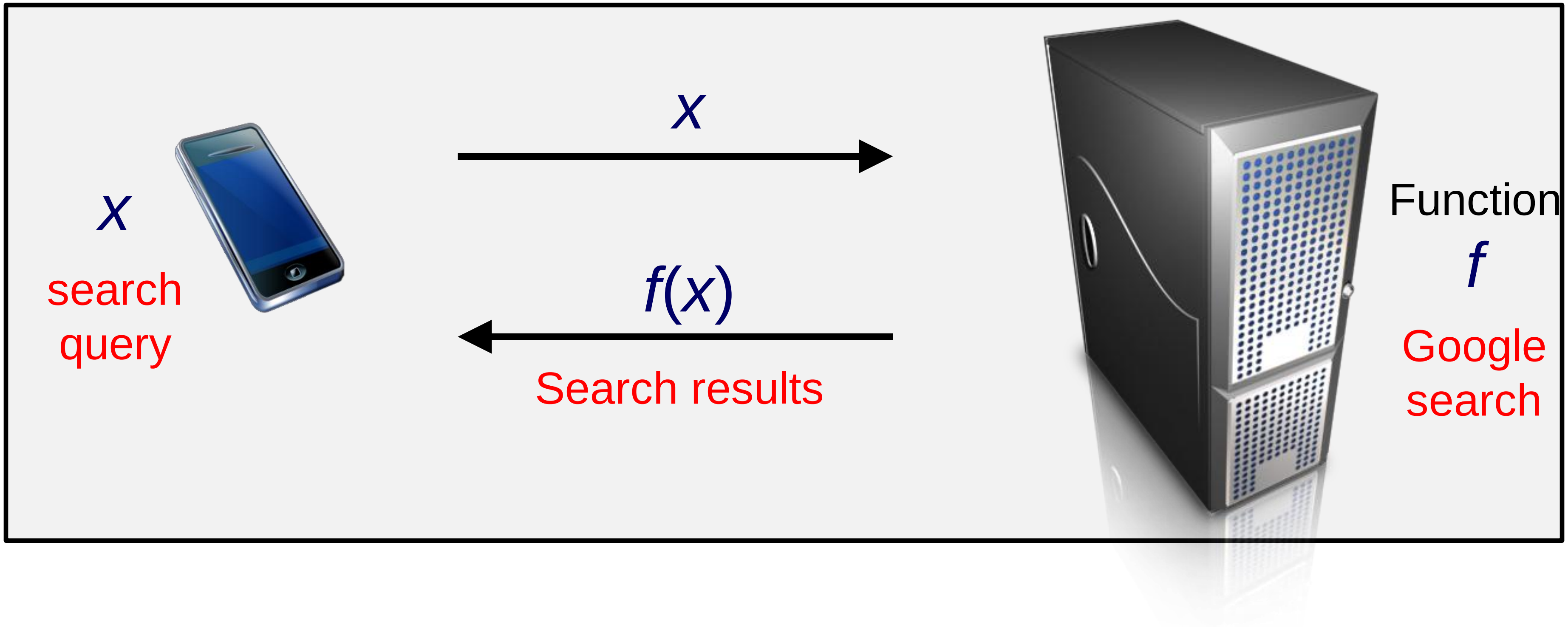
encrypt

E(a), E(b)

compute

E(a) ✖ E(b) = E(a b)

FHE



x

search query

x

$f(x)$

Search results

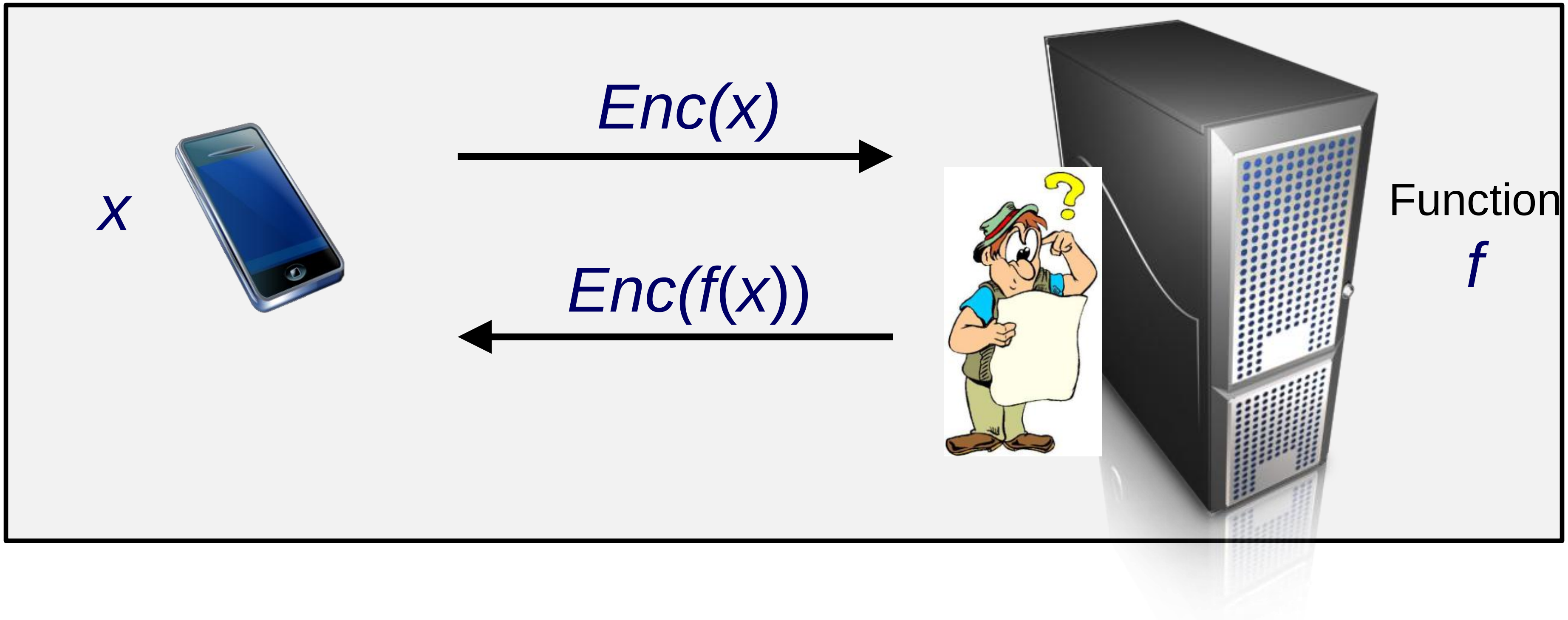
Function

f

Google search

WANT PRIVACY!

Computing on Encrypted Data



x

$Enc(x)$

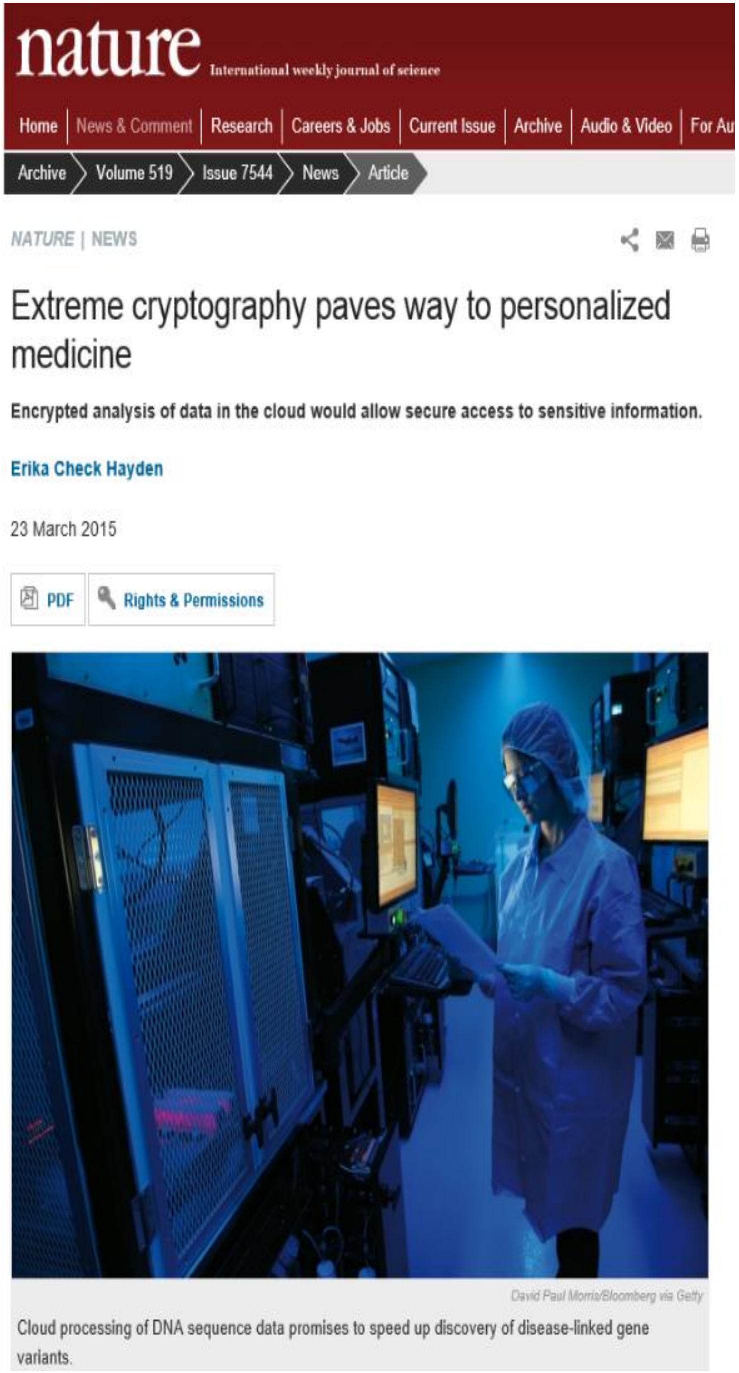
$Enc(f(x))$

Function f

WANT PRIVACY!

Breakthrough

- Gentry, 2009: Mathematical possibility of computing on encrypted data
  - Any combination of linear and non-linear operation
- Problem: **extremely** inefficient
- Sub-discipline became very active, funded e.g. by a 20M$ DARPA grant, ERC, …



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

Archive > Volume 519 > Issue 7544 > News > Article

NATURE | NEWS

Extreme cryptography paves way to personalized medicine

Encrypted analysis of data in the cloud would allow secure access to sensitive information.

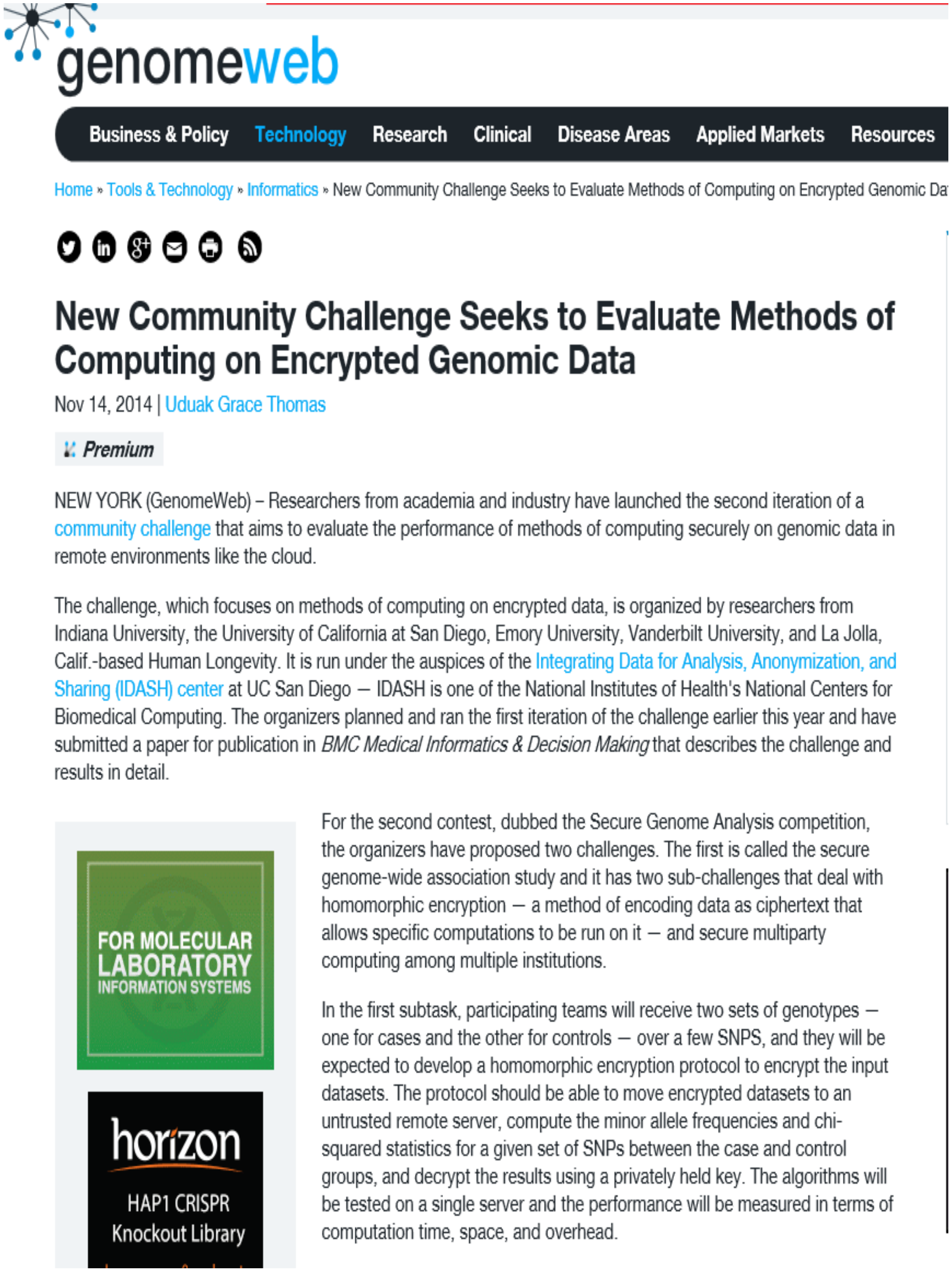
Erika Check Hayden

23 March 2015

PDF | Rights & Permissions

David Paul Morris/Bloomberg via Getty

Cloud processing of DNA sequence data promises to speed up discovery of disease-linked gene variants.



genomeweb

Business & Policy | Technology | Research | Clinical | Disease Areas | Applied Markets | Resources

Home » Tools & Technology » Informatics » New Community Challenge Seeks to Evaluate Methods of Computing on Encrypted Genomic Da

New Community Challenge Seeks to Evaluate Methods of Computing on Encrypted Genomic Data

Nov 14, 2014 | Uduak Grace Thomas

Premium

NEW YORK (GenomeWeb) – Researchers from academia and industry have launched the second iteration of a community challenge that aims to evaluate the performance of methods of computing securely on genomic data in remote environments like the cloud.

The challenge, which focuses on methods of computing on encrypted data, is organized by researchers from Indiana University, the University of California at San Diego, Emory University, Vanderbilt University, and La Jolla, Calif.-based Human Longevity. It is run under the auspices of the Integrating Data for Analysis, Anonymization, and Sharing (IDASH) center at UC San Diego — IDASH is one of the National Institutes of Health's National Centers for Biomedical Computing. The organizers planned and ran the first iteration of the challenge earlier this year and have submitted a paper for publication in *BMC Medical Informatics & Decision Making* that describes the challenge and results in detail.

For the second contest, dubbed the Secure Genome Analysis competition, the organizers have proposed two challenges. The first is called the secure genome-wide association study and it has two sub-challenges that deal with homomorphic encryption — a method of encoding data as ciphertext that allows specific computations to be run on it — and secure multiparty computing among multiple institutions.

In the first subtask, participating teams will receive two sets of genotypes — one for cases and the other for controls — over a few SNPS, and they will be expected to develop a homomorphic encryption protocol to encrypt the input datasets. The protocol should be able to move encrypted datasets to an untrusted remote server, compute the minor allele frequencies and chi-squared statistics for a given set of SNPs between the case and control groups, and decrypt the results using a privately held key. The algorithms will be tested on a single server and the performance will be measured in terms of computation time, space, and overhead.

Why the excitement?

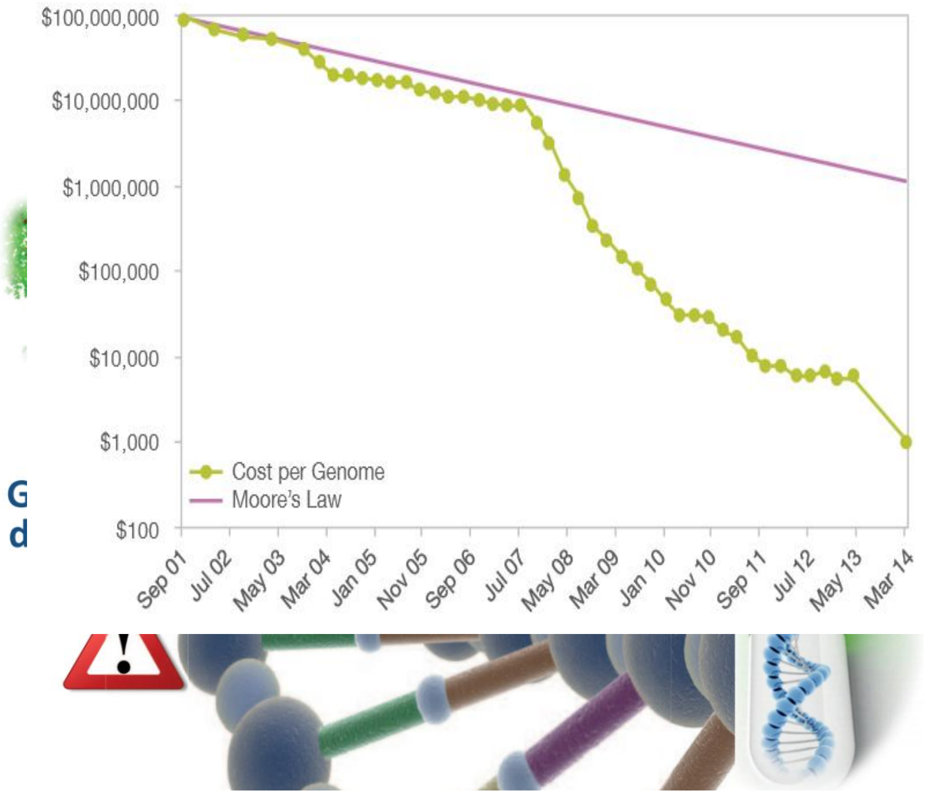
- Fundamental Problem: privacy protection
  - Burgeoning genome sequencing capability
  - Explosion of scientific research possible
  - High risk for personal privacy
- Fundamental Progress through interaction
  - Computer Scientists
  - Mathematicians
  - Bioinformaticians
  - Policy-makers

Genomic Revolution

- Fast drop in the cost of genome-sequencing
  - 2000: $3 billion
  - Mar. 2014: $1,000
  - Genotyping 1M variations: below $200

- Unleashing the potential of the technology
  - Healthcare: e.g., disease risk detection, personalized medicine
  - Biomedical research: e.g., geno-phono association
  - Legal and forensic
  - DTC: e.g., ancestry test, paternity test

  ……



Genome Privacy

- Privacy risks
  - Genetic disease disclosure
  - Collateral damage
  - Genetic discrimination

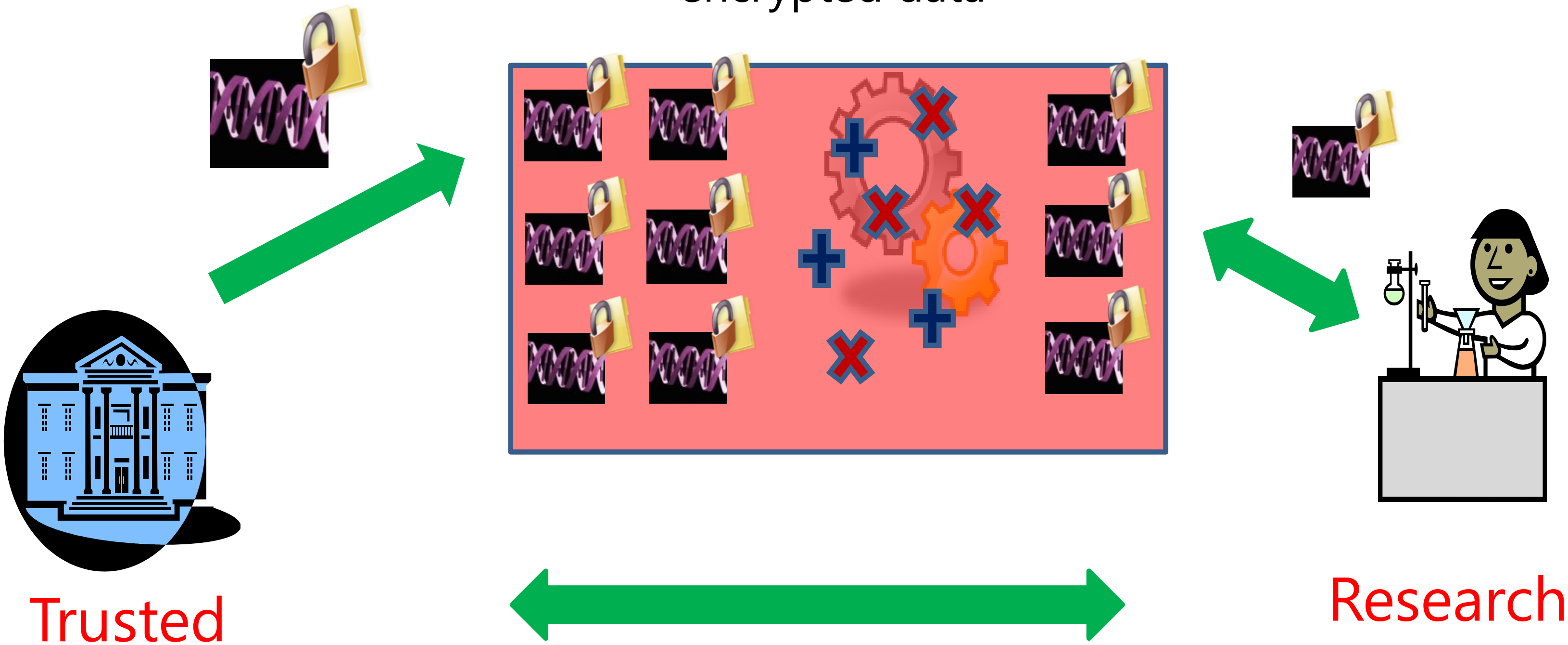
- Grand Challenges:
  - **How to share genomic data or learning in a way that preserves the privacy of the data donors, without undermining the utility of the data or impeding its convenient dissemination?**
  - **How to perform a LARGE-SCALE, PRIVACY-PRESERVING analysis on genomic data, in an untrusted cloud environment or across multiple users?**

Computations on Genome sequence data

- Building predictive models
- Predictive analysis
  - Classification tasks
  - Disease prediction
  - Sequence matching
- Data quality testing
- Basic statistical functions
- Statistical computations on genomic data

Untrusted cloud service

Stores, computes on encrypted data



Trusted party

hosts data and regulates access

Requests for decryption of results (requires a policy)

Researcher:

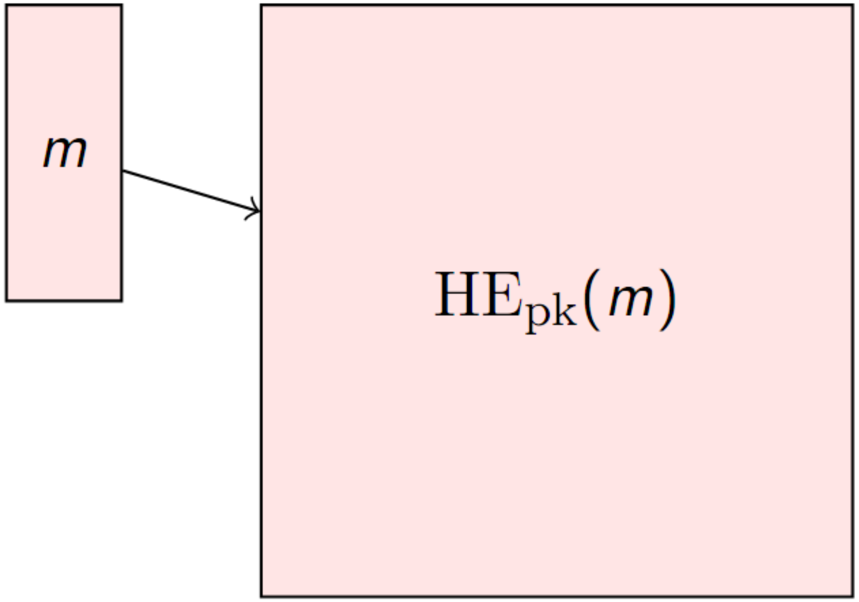
requests encrypted results of specific computations

What are the Costs? Challenges? Obstacles?

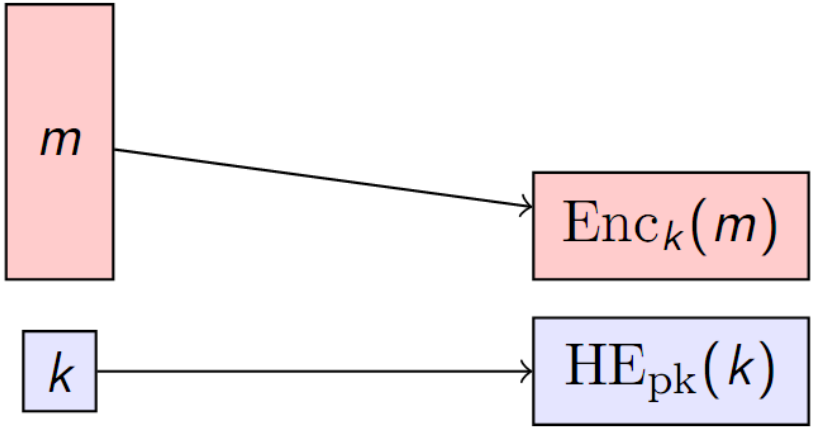
For homomorphic encryption

- Storage costs (large ciphertexts)
- New hard problems (introduced 2010-2015)
- Efficiency at scale (large amounts of data, deep circuits)

FHE Application of Ciphers



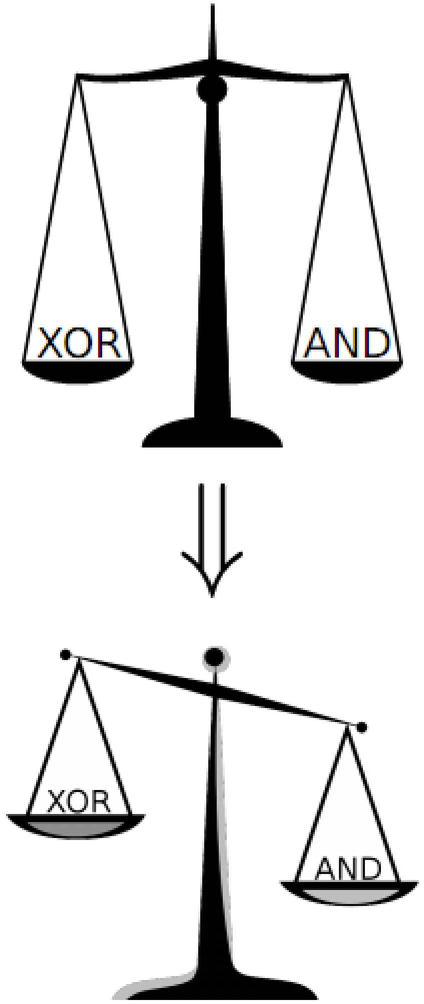
FHE schemes typically come with a ciphertext expansion in the order of 1000s to 1000000s.



Proposed solution: encrypt messages!

Cloud homomorphically decrypts them.

New designs for new computational models



- Since 1970s: balance between linear and non-linear operations
- Idea: Explore ***extreme*** trade-offs

How would an efficient cipher look like if linear operations were for free?

Towards LowMC

- Metrics to optimize:

  AND-depth,

  #AND/bit

  #ANDs

Related work

Ciphers that try to minimizing cost of side-channel attack countermeasures

- Noekeon
- LS-designs (Robin, Fantomas)

Ciphers that try to minimize the latency when implemented in hardware

- Prince

LowMC

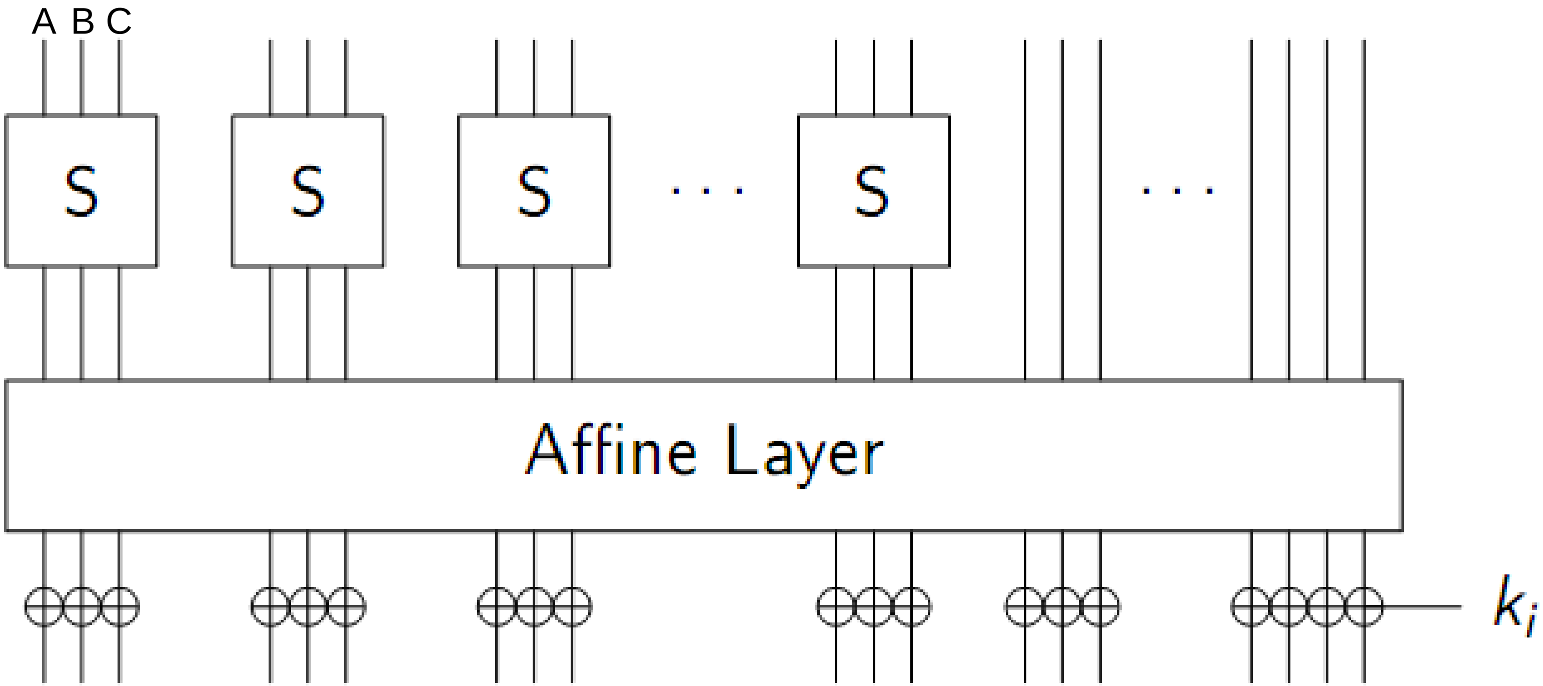
- Joint work with Martin Albrecht (RHUL), Thomas Schneider (TUD), Michael Zohner (TUD) and Tyge Tiessen (DTU)

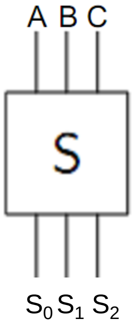
High-level design approach

- Minimise ANDs for confusion
- Maximize diffusion

- Use SPN
- Use small S-box with low multiplicative complexity
- Use partial S-box layer
- Maximise diffusion in affine layer

Round transformation





$$S_0(A, B, C) = A \oplus BC$$

$$S_1(A, B, C) = A \oplus B \oplus AC$$

$$S_2(A, B, C) = A \oplus B \oplus C \oplus AB$$

Affine layer

Let block-size be n

Multiplication of internal state with randomly chosen invertible matrix in GF(2) with n rows/columns

Add randomly chosen n-bit vector

Distinct for every rounds

Key schedule

- Re-use random matrix approach
  - Derive round keys from master key by multiplication with nxk binary matrix
  - Choose matrices uniformly at random from all binary nxk matrices of rank min(n,k)

Design space

Size

- n: Block size
- m: Number of Sboxes

Security

- k: Key size (allowed time complexity)
- d: allowed data complexity

r: Number of rounds that is computed as a function of (n,m,k,d)

How to determine #rounds r

- Cryptanalysis
  - How long is the longest distinguishers
    - Statistical distinguisher (e.g. differential, linear)
    - Combined attacks: special case Boomerang attacks
    - Low-degree attacks
    - …
  - How many rounds could be peeled off?
    - ?

Resistance against differential attacks

- Standard method to determine probability of best differential characteristic:
  - Determine minimal number of active Sboxes.
  - Combine with maximal differential probability of Sbox to determine lower bound on best possible characteristic.
- To determine the minimal number of active Sboxes the branch number would be helpful.
- We do not know the branch number of the randomly chosen matrix

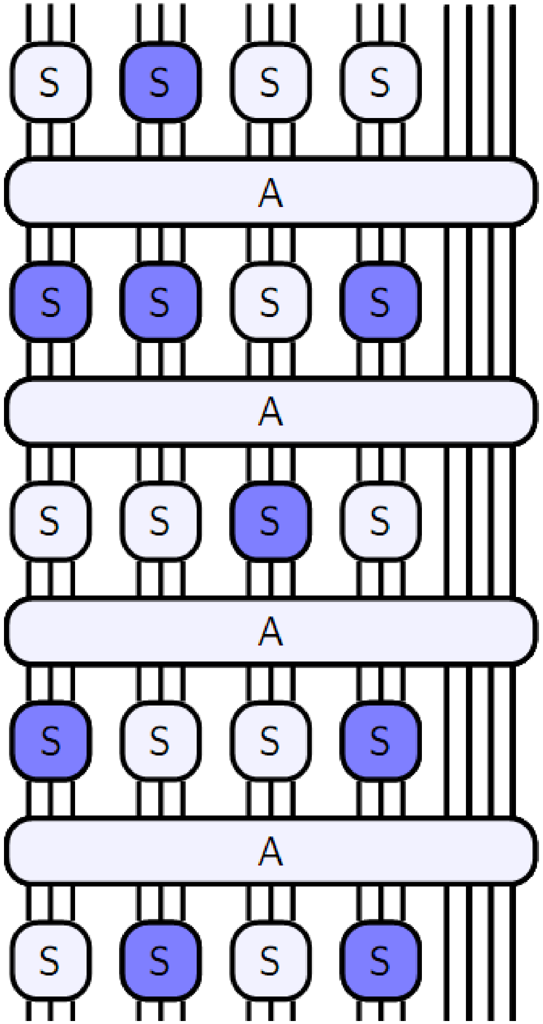
Resistance against differential attacks

- Idea:
  - Calculate for each possible good differential characteristic probability that it is realized in instantiation of LowMC. Sum all these probabilities to get upper bound for probability that at least one is realized.

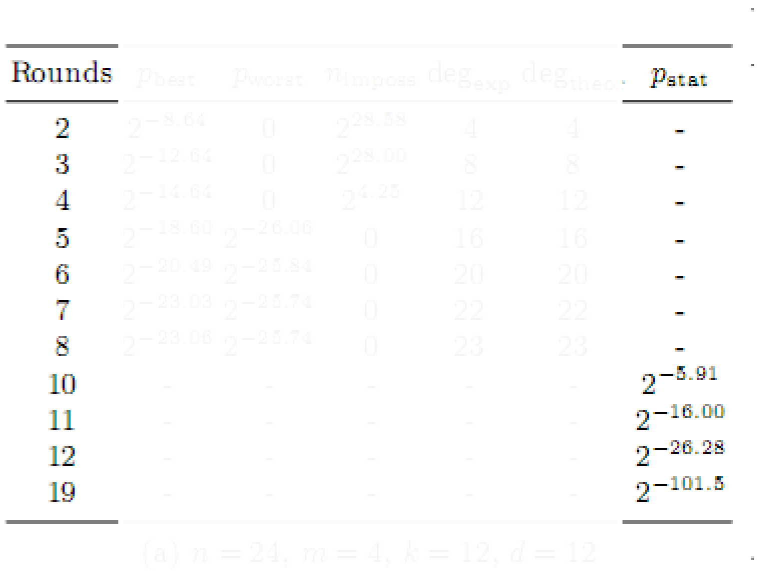
Let C set of possible good characteristics.

Sum over all c∈C:

Pr(c exists in cipher) ≤Pr(good characteristic exists)



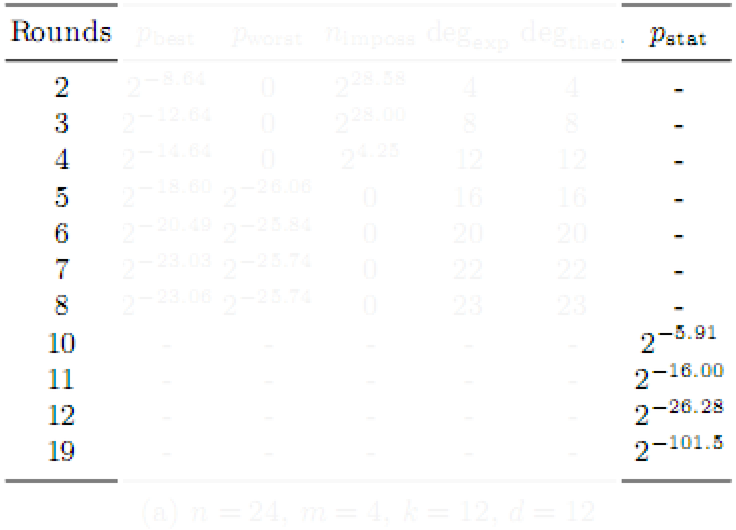
Bounds against differential attacks



Rounds	p_{best}	p_{worst}	n_{imposs}	$\deg_{\text{exp}}$	$\deg_{\text{theo}}$	p_{stat}
2	$2^{-8.64}$	0	$2^{28.58}$	4	4	-
3	$2^{-12.64}$	0	$2^{28.00}$	8	8	-
4	$2^{-14.64}$	0	$2^{4.25}$	12	12	-
5	$2^{-18.60}$	$2^{-26.06}$	0	16	16	-
6	$2^{-20.49}$	$2^{-25.84}$	0	20	20	-
7	$2^{-23.03}$	$2^{-25.74}$	0	22	22	-
8	$2^{-23.06}$	$2^{-25.74}$	0	23	23	-
10	-	-	-	-	-	$2^{-5.91}$
11	-	-	-	-	-	$2^{-16.00}$
12	-	-	-	-	-	$2^{-26.28}$
19	-	-	-	-	-	$2^{-101.5}$

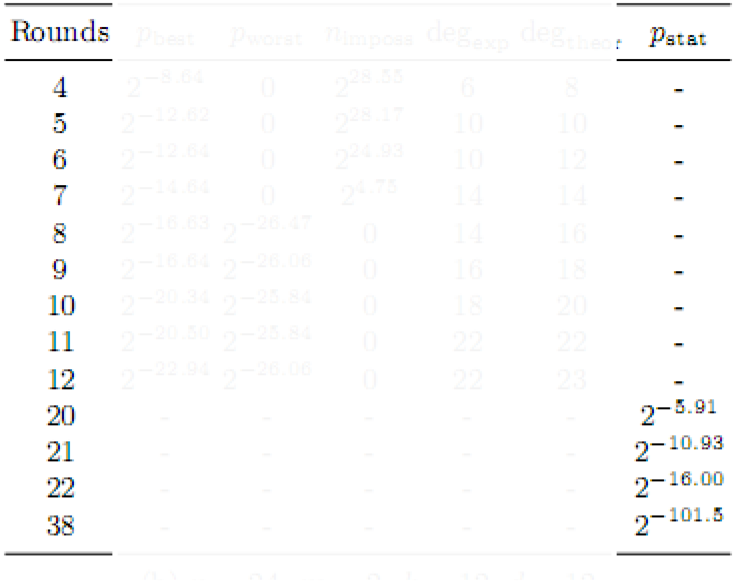
(a) $n = 24$, $m = 4$, $k = 12$, $d = 12$

Bounds against differential attacks



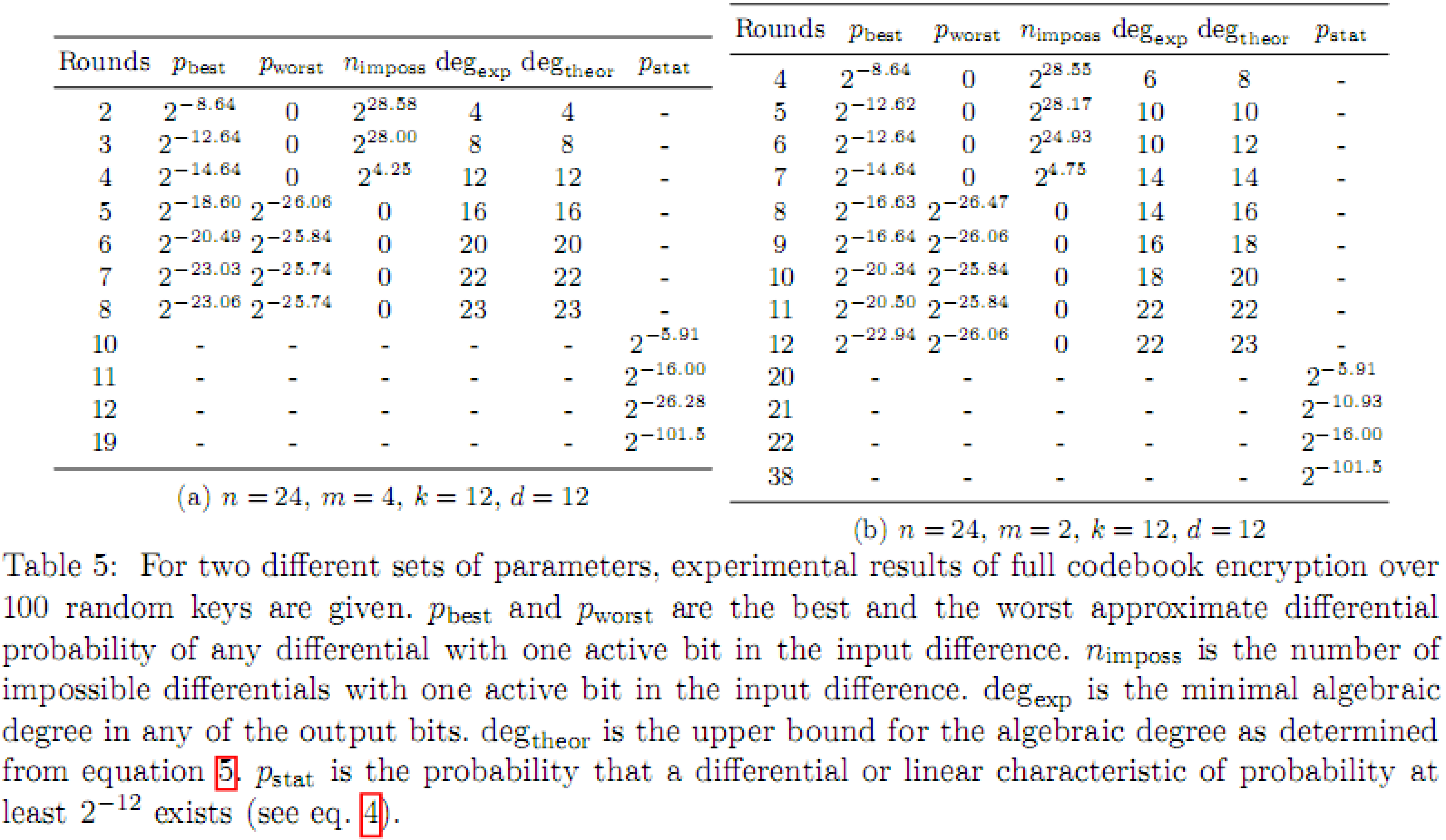
Rounds	p_{best}	p_{worst}	n_{imposs}	$\deg_{\text{exp}}$	$\deg_{\text{theo}}$	p_{stat}
2	$2^{-8.64}$	0	$2^{28.58}$	4	4	-
3	$2^{-12.64}$	0	$2^{28.00}$	8	8	-
4	$2^{-14.64}$	0	$2^{4.25}$	12	12	-
5	$2^{-18.60}$	$2^{-26.06}$	0	16	16	-
6	$2^{-20.49}$	$2^{-25.84}$	0	20	20	-
7	$2^{-23.03}$	$2^{-25.74}$	0	22	22	-
8	$2^{-23.06}$	$2^{-25.74}$	0	23	23	-
10	-	-	-	-	-	$2^{-5.91}$
11	-	-	-	-	-	$2^{-16.00}$
12	-	-	-	-	-	$2^{-26.28}$
19	-	-	-	-	-	$2^{-101.5}$

(a) $n = 24$, $m = 4$, $k = 12$, $d = 12$



Rounds	p_{best}	p_{worst}	n_{imposs}	$\deg_{\text{exp}}$	$\deg_{\text{theor}}$	p_{stat}
4	$2^{-8.64}$	0	$2^{28.55}$	6	8	-
5	$2^{-12.62}$	0	$2^{28.17}$	10	10	-
6	$2^{-12.64}$	0	$2^{24.93}$	10	12	-
7	$2^{-14.64}$	0	$2^{4.75}$	14	14	-
8	$2^{-16.63}$	$2^{-26.47}$	0	14	16	-
9	$2^{-16.64}$	$2^{-26.06}$	0	16	18	-
10	$2^{-20.34}$	$2^{-25.84}$	0	18	20	-
11	$2^{-20.50}$	$2^{-25.84}$	0	22	22	-
12	$2^{-22.94}$	$2^{-26.06}$	0	22	23	-
20	-	-	-	-	-	$2^{-5.91}$
21	-	-	-	-	-	$2^{-10.93}$
22	-	-	-	-	-	$2^{-16.00}$
38	-	-	-	-	-	$2^{-101.5}$

Bounds + concrete security against differential attacks



Rounds	p_{best}	p_{worst}	n_{imposs}	$\deg_{\text{exp}}$	$\deg_{\text{theor}}$	p_{stat}
2	$2^{-8.64}$	0	$2^{28.58}$	4	4	-
3	$2^{-12.64}$	0	$2^{28.00}$	8	8	-
4	$2^{-14.64}$	0	$2^{4.25}$	12	12	-
5	$2^{-18.60}$	$2^{-26.06}$	0	16	16	-
6	$2^{-20.49}$	$2^{-25.84}$	0	20	20	-
7	$2^{-23.03}$	$2^{-25.74}$	0	22	22	-
8	$2^{-23.06}$	$2^{-25.74}$	0	23	23	-
10	-	-	-	-	-	$2^{-5.91}$
11	-	-	-	-	-	$2^{-16.00}$
12	-	-	-	-	-	$2^{-26.28}$
19	-	-	-	-	-	$2^{-101.5}$

(a) $n = 24$, $m = 4$, $k = 12$, $d = 12$

Rounds	p_{best}	p_{worst}	n_{imposs}	$\deg_{\text{exp}}$	$\deg_{\text{theor}}$	p_{stat}
4	$2^{-8.64}$	0	$2^{28.55}$	6	8	-
5	$2^{-12.62}$	0	$2^{28.17}$	10	10	-
6	$2^{-12.64}$	0	$2^{24.93}$	10	12	-
7	$2^{-14.64}$	0	$2^{4.75}$	14	14	-
8	$2^{-16.63}$	$2^{-26.47}$	0	14	16	-
9	$2^{-16.64}$	$2^{-26.06}$	0	16	18	-
10	$2^{-20.34}$	$2^{-25.84}$	0	18	20	-
11	$2^{-20.50}$	$2^{-25.84}$	0	22	22	-
12	$2^{-22.94}$	$2^{-26.06}$	0	22	23	-
20	-	-	-	-	-	$2^{-5.91}$
21	-	-	-	-	-	$2^{-10.93}$
22	-	-	-	-	-	$2^{-16.00}$
38	-	-	-	-	-	$2^{-101.5}$

(b) $n = 24$, $m = 2$, $k = 12$, $d = 12$

Table 5: For two different sets of parameters, experimental results of full codebook encryption over 100 random keys are given. p_{best} and p_{worst} are the best and the worst approximate differential probability of any differential with one active bit in the input difference. n_{imposs} is the number of impossible differentials with one active bit in the input difference. $\deg_{\text{exp}}$ is the minimal algebraic degree in any of the output bits. $\deg_{\text{theor}}$ is the upper bound for the algebraic degree as determined from equation 5. p_{stat} is the probability that a differential or linear characteristic of probability at least 2^{-12} exists (see eq. 4).

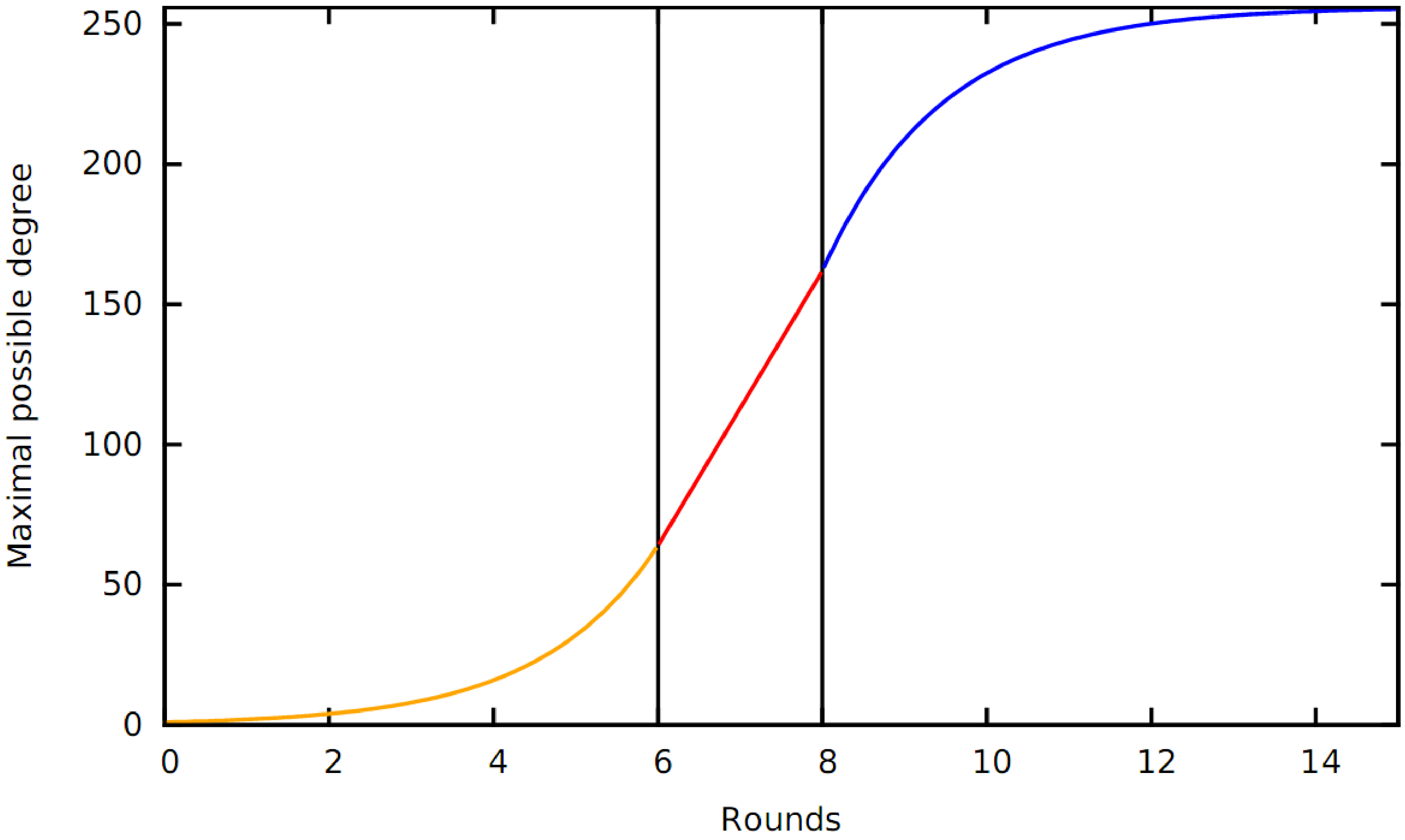
Resistance against combined attacks

- Example: Boomerang attacks, which use good differentials that meet half-way from both sides
- Partial non-linear layer allows probability 1 differential for a few rounds
- Solution:
  - Re-use approach from before for the heightened requirements
  - Double the length

Resistance against higher order attacks

- Question: What is the minimal number of rounds needed to reach a given algebraic degree?
- Lemma: If algebraic degree is dr after r rounds, max. degree in round r+1 is min (**2dr**, **m+dr**, **n/2+dr/2**)

Growth of degree



Round formular

r ≥ max(rstat, rdeg, rcmbnd) + router

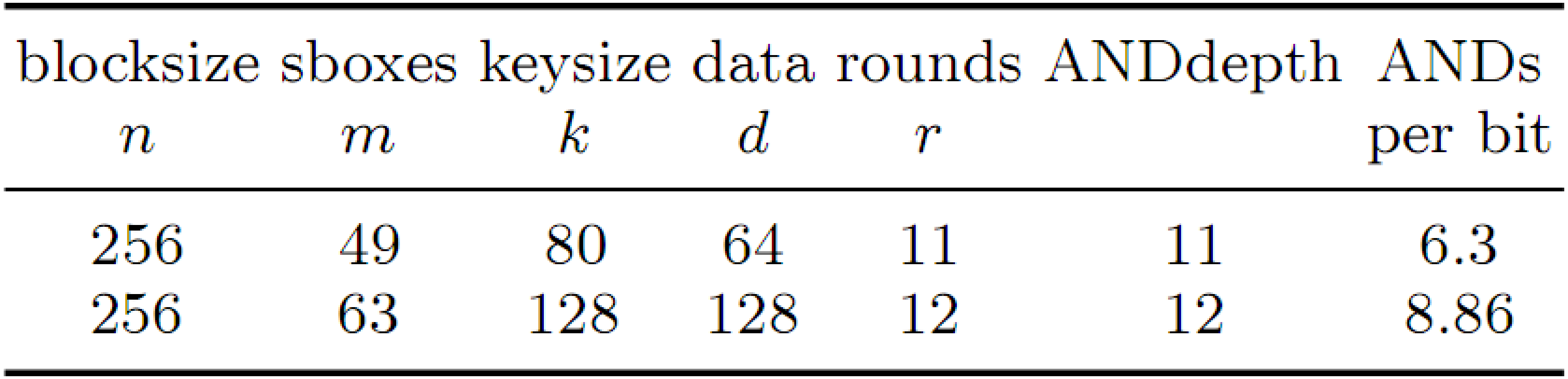
rstat: bound for differential and linear distinguishers

rdeg: bound for sufficient degree

rcmbnd: bound for combined distinguishers

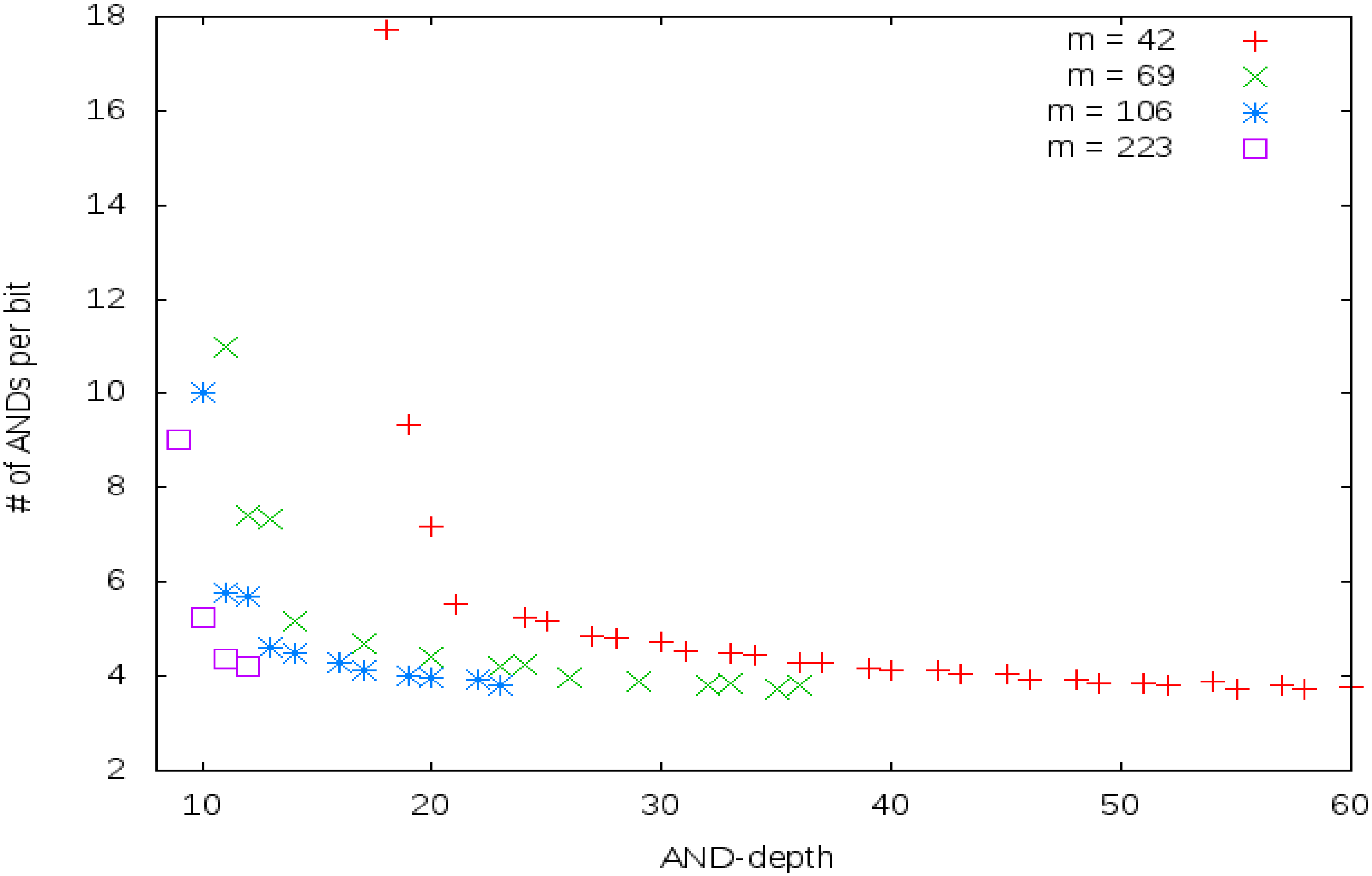
router: bound for rounds that can be peeled off (we choose router=rstat)

Concrete instances

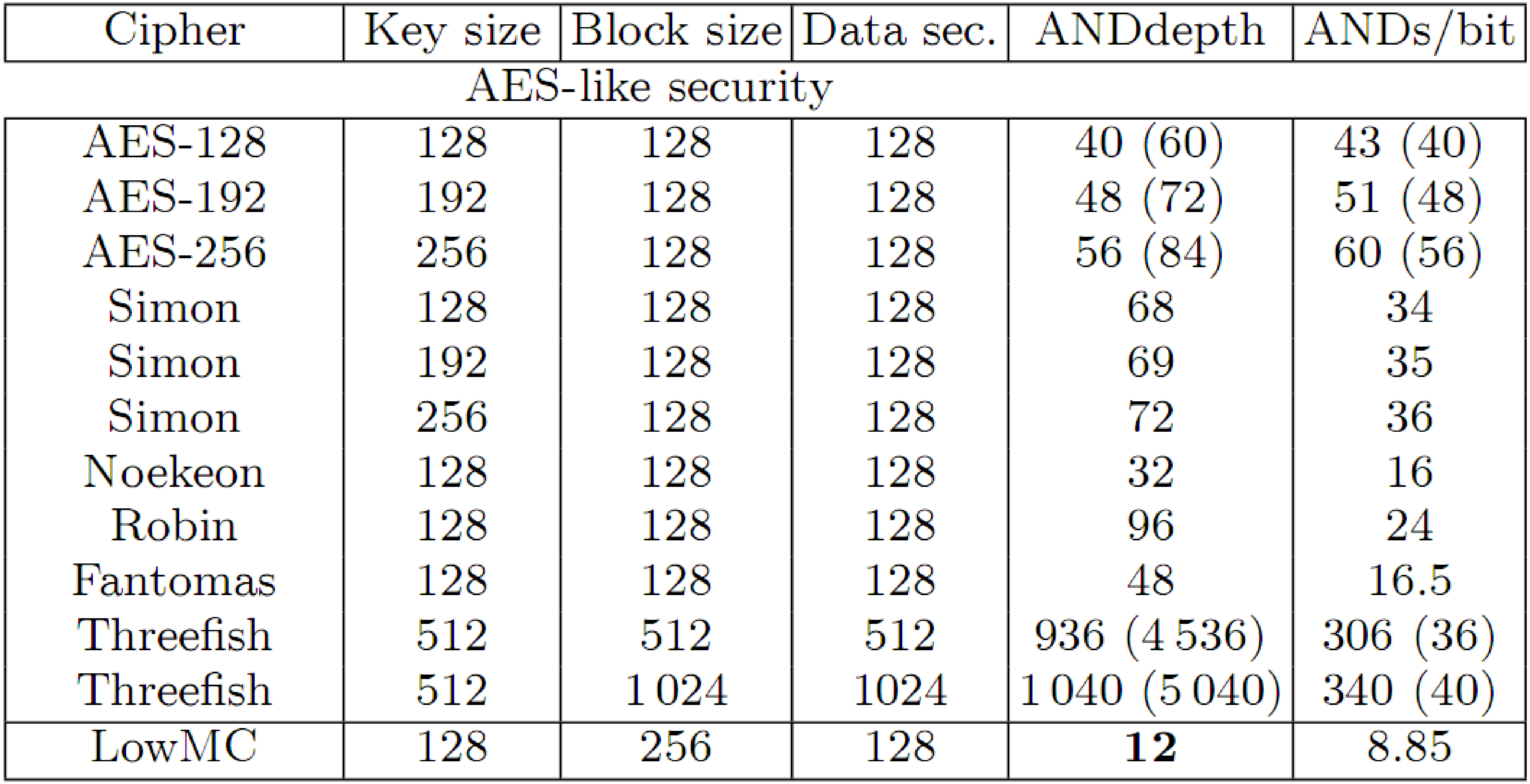


blocksize	sboxes	keysize	data	rounds	ANDdepth	ANDs
n	m	k	d	r		per bit
256	49	80	64	11	11	6.3
256	63	128	128	12	12	8.86

Visualizing the design space

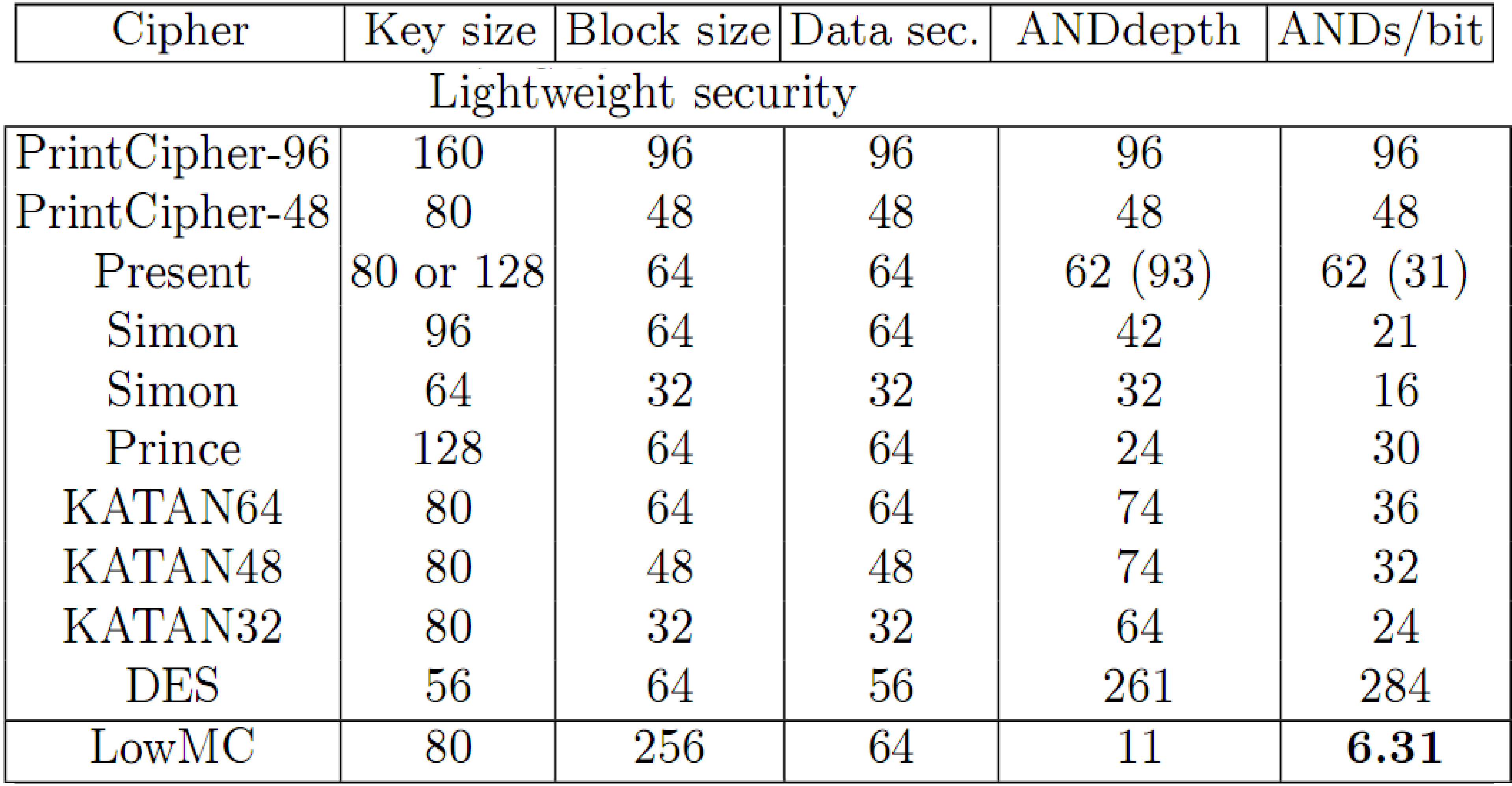


Comparison with other designs AES-like security



Cipher	Key size	Block size	Data sec.	ANDdepth	ANDs/bit
AES-like security					
AES-128	128	128	128	40 (60)	43 (40)
AES-192	192	128	128	48 (72)	51 (48)
AES-256	256	128	128	56 (84)	60 (56)
Simon	128	128	128	68	34
Simon	192	128	128	69	35
Simon	256	128	128	72	36
Noekeon	128	128	128	32	16
Robin	128	128	128	96	24
Fantomas	128	128	128	48	16.5
Threefish	512	512	512	936 (4 536)	306 (36)
Threefish	512	1 024	1024	1 040 (5 040)	340 (40)
LowMC	128	256	128	**12**	8.85

Comparison with other designs „lightweight“ security



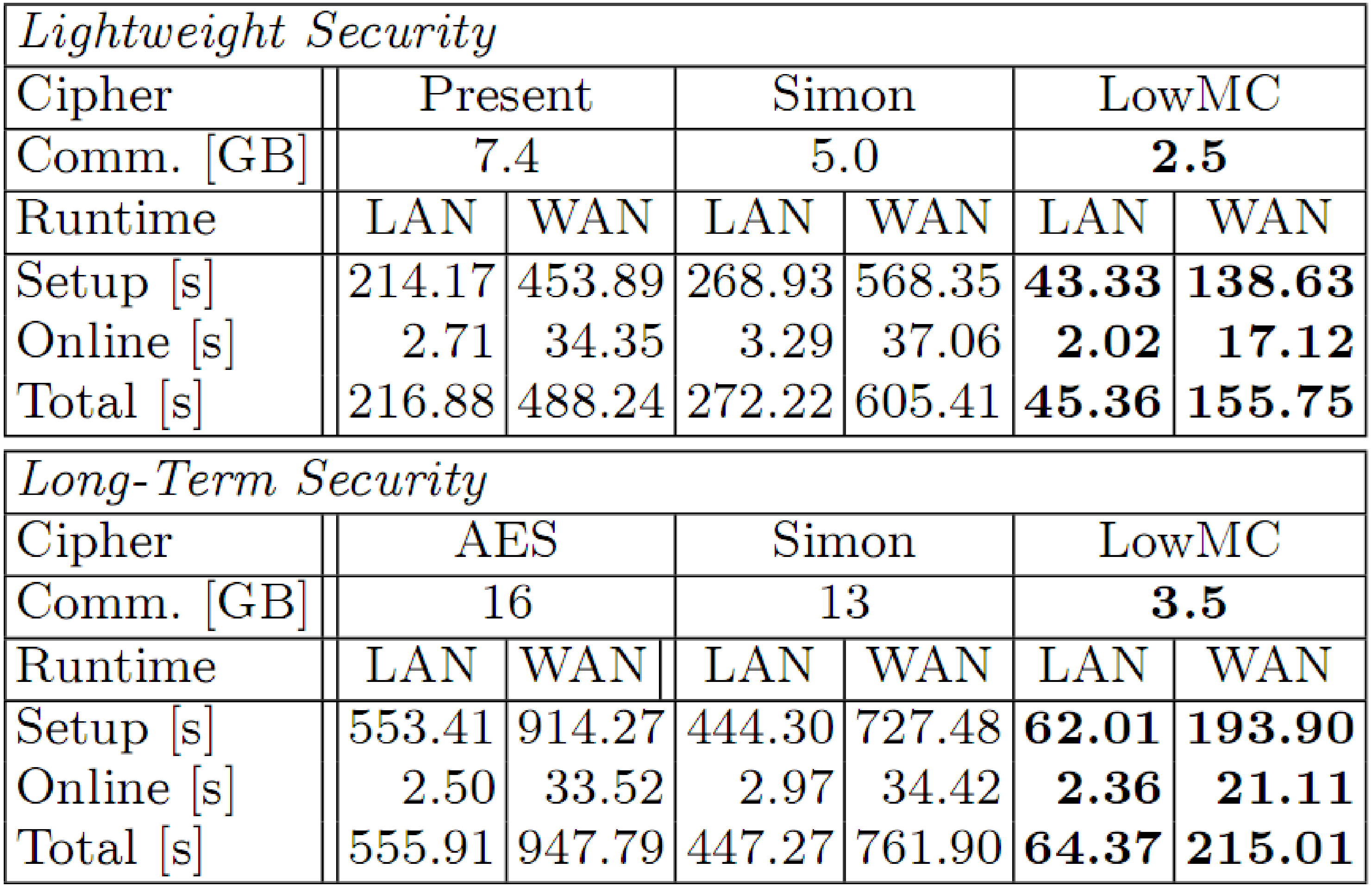
Lightweight security

Cipher	Key size	Block size	Data sec.	ANDdepth	ANDs/bit
PrintCipher-96	160	96	96	96	96
PrintCipher-48	80	48	48	48	48
Present	80 or 128	64	64	62 (93)	62 (31)
Simon	96	64	64	42	21
Simon	64	32	32	32	16
Prince	128	64	64	24	30
KATAN64	80	64	64	74	36
KATAN48	80	48	48	74	32
KATAN32	80	32	32	64	24
DES	56	64	56	261	284
LowMC	80	256	64	11	**6.31**

Properties and Advantages

- Low ANDDepth and ANDs/encrypted bit
- Block size and security(data-complexity) de-coupled
- Differential and linear attacks will *provably* not work, except for extremely unlucky choices of linear layers

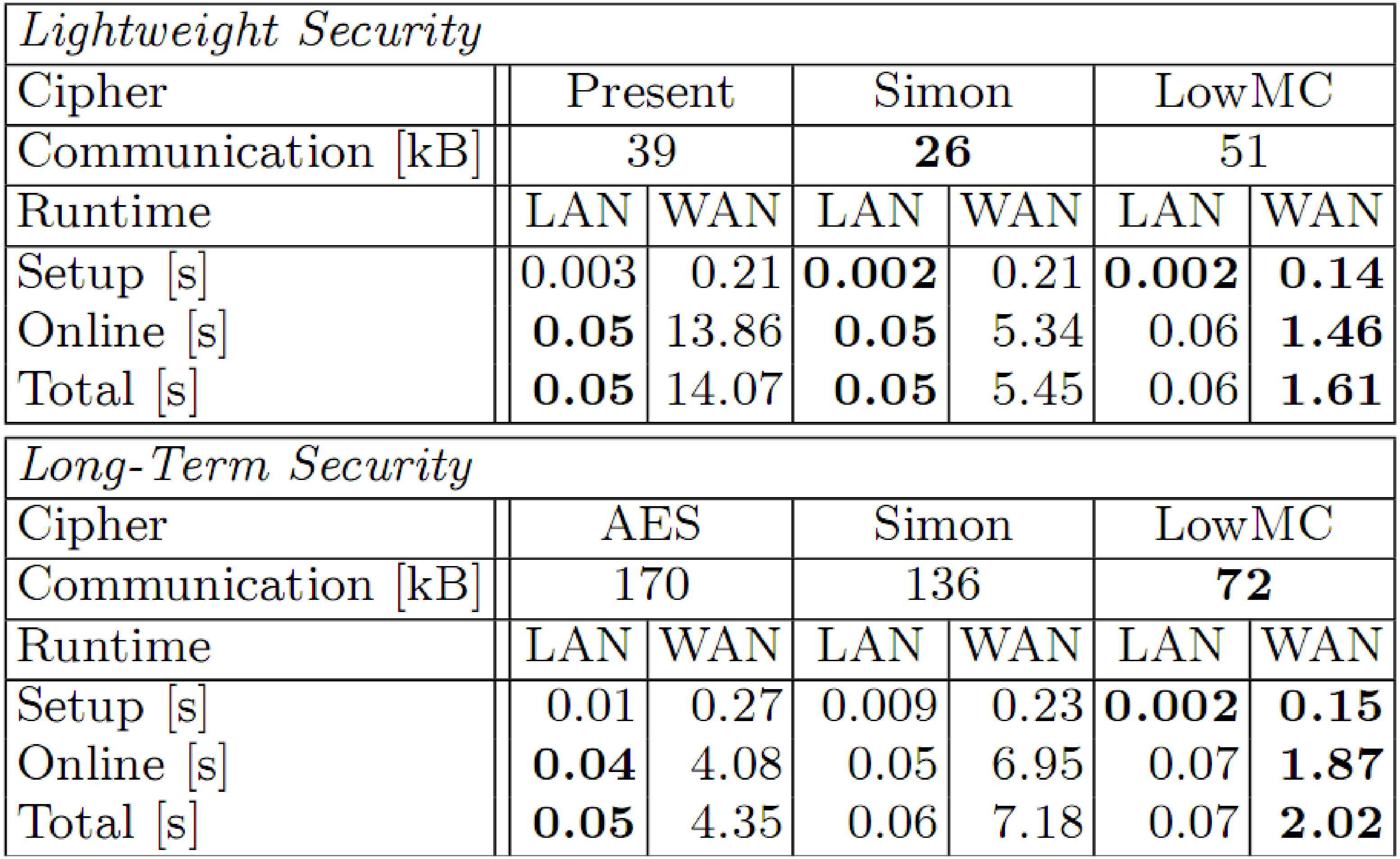
GMW benchmarks – long message



Lightweight Security						
Cipher	Present		Simon		LowMC	
Comm. [GB]	7.4		5.0		**2.5**	
Runtime	LAN	WAN	LAN	WAN	LAN	WAN
Setup [s]	214.17	453.89	268.93	568.35	**43.33**	**138.63**
Online [s]	2.71	34.35	3.29	37.06	**2.02**	**17.12**
Total [s]	216.88	488.24	272.22	605.41	**45.36**	**155.75**

Long-Term Security						
Cipher	AES		Simon		LowMC	
Comm. [GB]	16		13		**3.5**	
Runtime	LAN	WAN	LAN	WAN	LAN	WAN
Setup [s]	553.41	914.27	444.30	727.48	**62.01**	**193.90**
Online [s]	2.50	33.52	2.97	34.42	**2.36**	**21.11**
Total [s]	555.91	947.79	447.27	761.90	**64.37**	**215.01**

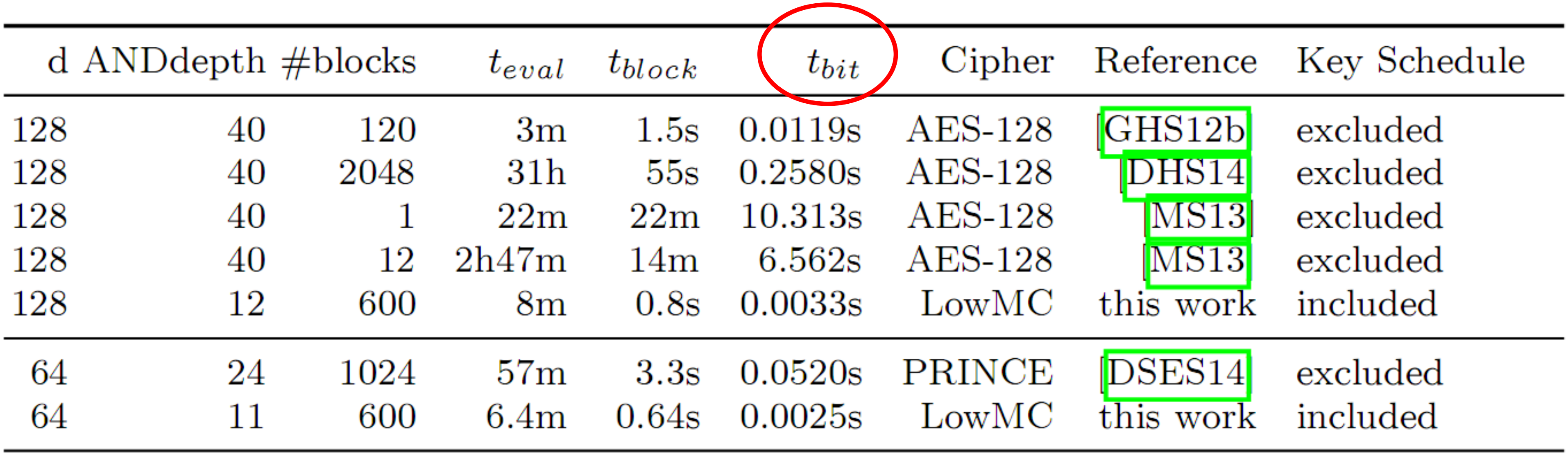
GMW benchmarks – single block



Lightweight Security						
Cipher	Present		Simon		LowMC	
Communication [kB]	39		**26**		51	
Runtime	LAN	WAN	LAN	WAN	LAN	WAN
Setup [s]	0.003	0.21	**0.002**	0.21	**0.002**	**0.14**
Online [s]	**0.05**	13.86	**0.05**	5.34	0.06	**1.46**
Total [s]	**0.05**	14.07	**0.05**	5.45	0.06	**1.61**

Long-Term Security						
Cipher	AES		Simon		LowMC	
Communication [kB]	170		136		**72**	
Runtime	LAN	WAN	LAN	WAN	LAN	WAN
Setup [s]	0.01	0.27	0.009	0.23	**0.002**	**0.15**
Online [s]	**0.04**	4.08	0.05	6.95	0.07	**1.87**
Total [s]	**0.05**	4.35	0.06	7.18	0.07	**2.02**

FHE implementation benchmarks



d	ANDdepth	#blocks	t_{eval}	t_{block}	t_{bit}	Cipher	Reference	Key Schedule
128	40	120	3m	1.5s	0.0119s	AES-128	[GHS12b]	excluded
128	40	2048	31h	55s	0.2580s	AES-128	[DHS14]	excluded
128	40	1	22m	22m	10.313s	AES-128	[MS13]	excluded
128	40	12	2h47m	14m	6.562s	AES-128	[MS13]	excluded
128	12	600	8m	0.8s	0.0033s	LowMC	this work	included
64	24	1024	57m	3.3s	0.0520s	PRINCE	[DSES14]	excluded
64	11	600	6.4m	0.64s	0.0025s	LowMC	this work	included

Caveat: implementations/underlying techniques improve over time

Conclusions

- Explored extreme corner of cipher design space, motivated by new set of applications
- PRF with ANDdepth 11/12 with 128-bit security, balanced with low number of ANDs/bit

- One order of magnitude speed-gain
- Is this the limit?

Open Problems

- Cryptanalysis
- Design
- Implementation

Open Problems: Cryptanalysis

- Analysis of concrete LowMC instances against other attack vectors
  - Algebraic attacks
    - extremely simple structure
    - more information available per PT/CT pair
  - ?
- (Asymptotic) behavior of attacks vectors when blocksize increases
  - Largely solved for differential/linear attacks
  - MITM/Imposs. Differential/Integral/… attacks?

Open Problems: Design

- Application for even more extreme concrete parameterizations for LowMC?
- Larger S-Boxes with low ANDdepth?
- Hash functions using the same design strategy
- Something that is fast, both in the classical as well as in the new MPC/FHE world.
- LowMC design mainly optimizes for ANDdepth and GF(2) multiplication. What about other settings?

Open Problems: Implementations

Improved implementations of LowMC in

- GMW
- Yao
- SPDZ
- …

Other protocols / applications

- Interested in MPC protocols that are slower but have some desirable property
  - More advantages of choosing LowMC over AES
  - Example: Large scale MPC
  - Others?
- Applications in other areas
  - SNARKS
  - Obfuscation

New Ciphers for MPC and FHE

Q&A

Christian Rechberger, DTU

Joint work with Martin Albrecht (RHUL), Thomas Schneider (TUD), Michael Zohner (TUD) and Tyge Tiessen (DTU)