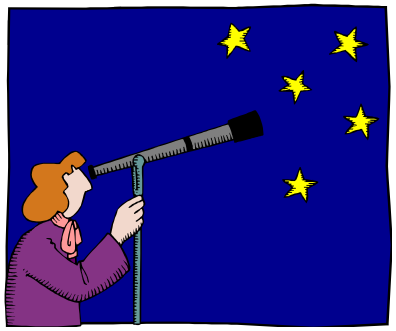
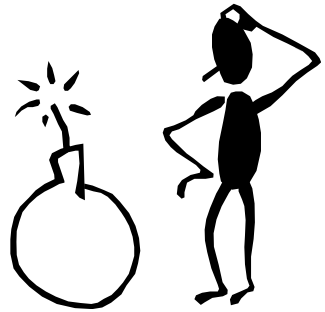
A New Frontier in Symmetric Cryptanalysis



Nicolas T. Courtois



University College of London, UK

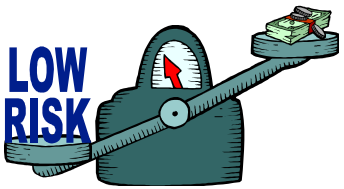
This Talk

- Part 1: a lot of bla bla bla,
  - Is Cryptology as Science? Etc.
- Part 2: HOW to do Algebraic Cryptanalysis [AC]

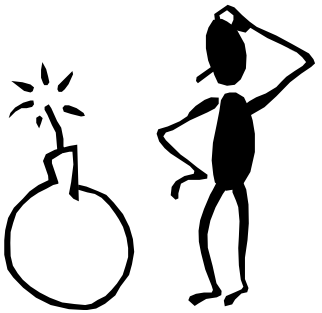
  Remark: AC is as much about

  1. sophisticated algorithms such as MiniSat and F5,
  2. much simpler algorithms and clever ‘tricks’ that allow sudden jumps in complexity.

0. Feel Secure or Paranoid Today?



?



The Curious "Science" of Security

"We need – today again -- to
re-discover the frontiers
of what is secure
that have just moved
yesterday…

Are Cryptologists Always Wrong?

Neal Koblitz: 

"The Uneasy Relationship Between Mathematics and Cryptography", In Notices of the American Mathematical Society, September 2007, see www.ams.org

[…] Once I heard a speaker from NSA complain about university researchers who are cavalier about proposing untested cryptosystems. He pointed out that in the real world if your cryptography fails, you lose a million dollars or your secret agent gets killed.

In academia, if you write about a cryptosystem and then a few months later find a way to break it, you've got two new papers to add to your résumé![…]

Optimistic View

Nothing bad
has ever happened.

Anybody ever broke DES in
practical sense?

Fundamental Research:

Claim:

Some most fundamental questions that pertain to more or less all symmetric cryptosystems were never seriously studied

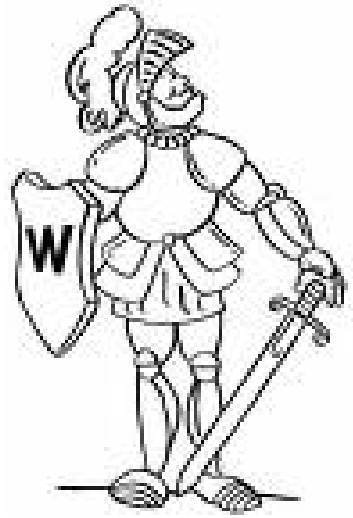
Frontiers

1. Maths vs. Crypto
Science vs. Fiction

Frontiers

Frontiers are natural: people from one place will naturally have trouble understanding other people.

- Some people come from
  Pure Orthodox Mathematics
- Some people are in
  Information Security
  - Cryptology/Computer Science/Law/ Crime Science/Finance and Economics/Marketing/Sociology/…



Very Recent Paper

Neal Koblitz: 

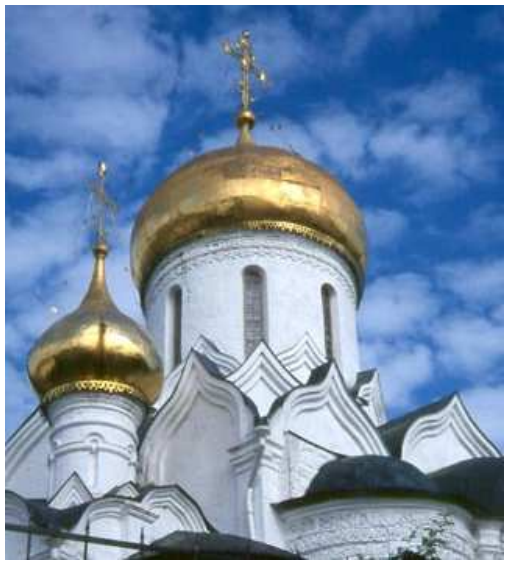
"The Uneasy Relationship Between Mathematics and Cryptography",
In Notices of the American Mathematical Society, September 2007, see www.ams.org

Cryptographic community:

- "The "spy vs. spy mentality"
- "constant competition and rivalry"
- "excessive - and even childish at times"

Mathematics [overheard]

- Mathematics: direct relationship with God.
- This cryptology is
  a profane and stupid
  engineering science…
- Cryptologists
  =def=
  people that have
  not grown big enough
  to do maths.



Cryptology

Ignorance Trap:

- We do **NOT WANT TO KNOW** about attacks unless:
  - They are faster than other known attacks on the same cipher (why so? major fallacy)
  - Their importance is already widely recognised (conservatism)

Also unless:

- It breaks their cipher, not ours…
- You pay us consultancy fees for that…

Mathematics

Intelligence Trap:

- Applied maths is bad maths.
- We do not want to consider facts.
  - We want to study ONLY what is provable [+with our favourite tools].
    - Control freak?
  - Zero risk: Do not dare formulate a conjecture that is not true.
    - Cryptology: 40 % risk for experts, 99 % for beginners.
- We have a proof, we don't need to experiment to verify if it's true.
    - Many proofs are actually wrong, subtleties.
- We need to study attacks that are complex and clever.
  - Simple attacks are not interesting?

Mathematics vs. Cryptology

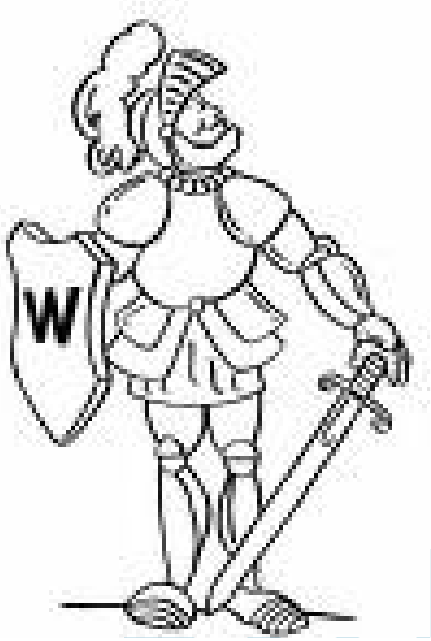
- Some mathematicians are maybe studying the empty set.
  - There are specific examples:
    Inaccessible cardinals, Ramsey cardinals, etc…

- In cryptology we do it ALL the time.

Conjectured assumptions collapse
on a daily basis.

Cryptology:

- Cryptology is almost a separate “science” that defines its own object of study (formal security definitions).
- We need to add axioms to mathematics.
  - Not everything is provable, statements that we love to make are all like: ∀ algorithm… Very few such statements were ever proven and very few will ever be…



- We have a direct relationship with God that specifically made the world an encrypted message to decode…

Courtois, Indocrypt 2008

***Remark:

"The discourse regarding the role of complexity in cryptography has degenerated to a point where it may take some time to recover."

[Kevin McCurley, in a post about Koblitz's criticism of crypto, 14 Sept. 2007]

Cryptologic Community:

Not much is proven…

A group of people with
shared beliefs

- Some deeply rooted
  in a certain reality of hardness
  resulting from precisely this endless confrontation
  of clever designers and clever attackers…
- Some are spectacularly naïve and are to collapse next, as usual in cryptology.
  - Like a religion in which the Gospels are rewritten each year.

AES

Advanced Encryption Standard:

- In 2000 NIST selected Rijndael as the AES.

Science vs. Fiction

Laws of Prediction [Arthur C. Clarke]:

When a distinguished elder scientist tells you something is not possible => he is wrong…

Algebraic Attacks on AES/Serpent/Etc:

"Provably" Secure [2000]
=> Speculative Fiction [2001]
=> Science Fiction [2002]
=> Science [2004-7]
=> Reality ???

But in Late 2001

A new kind of “terrorist” appears
and strikes some basic certitudes
about encryption

AL – GEB – RA الجبر

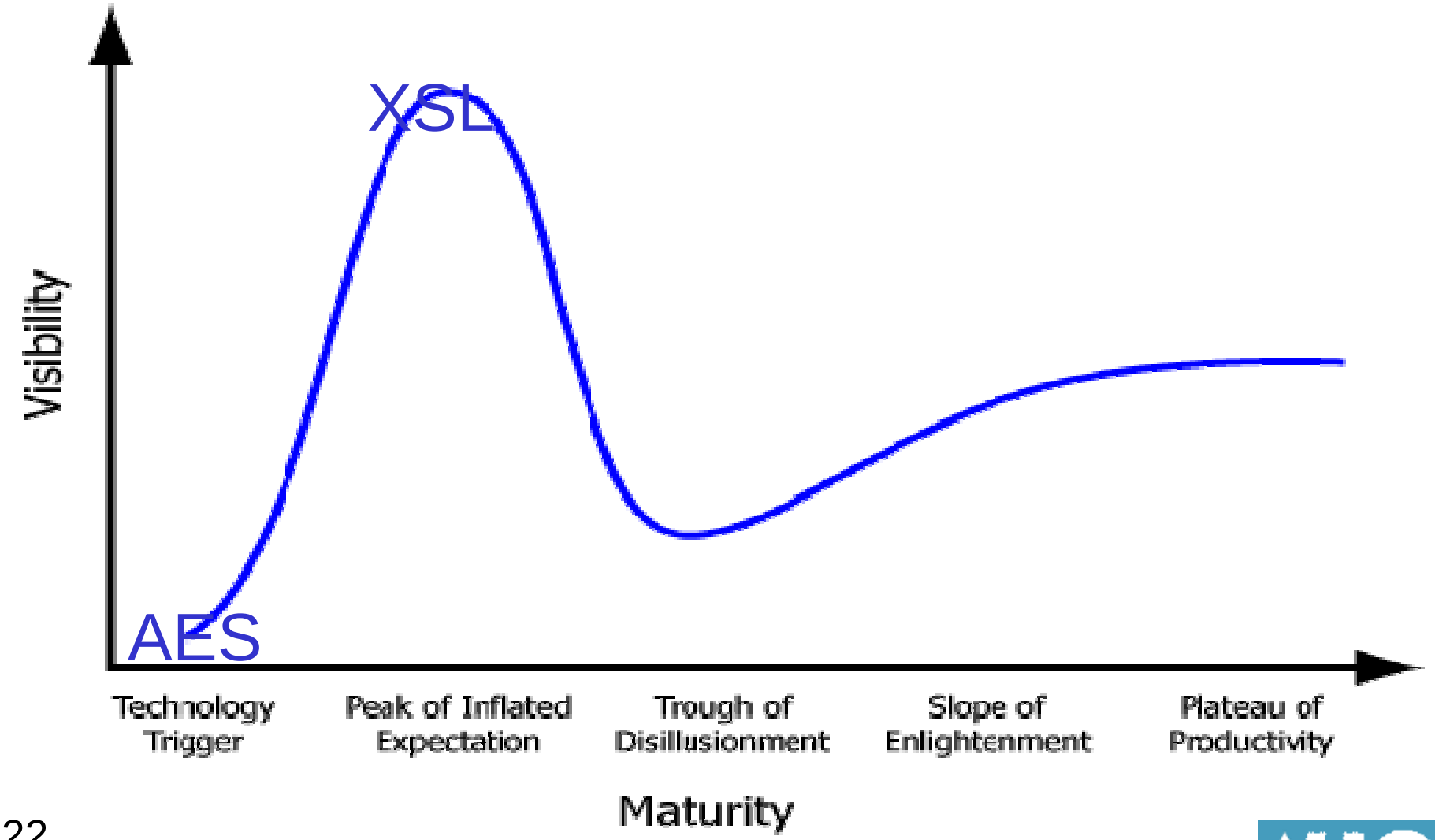
Reportedly the terrorist is not that dangerous...

However: he speaks a foreign language, and nobody really understands what he is up to.
[what really can be done with Gröbner bases and SAT solvers? Etc.]. So he might strike again from his secret basement ☺

Frontier-ology:

Frontiers are opportunities
for discovery and exploration.

Gartner's Technology Hype Cycle

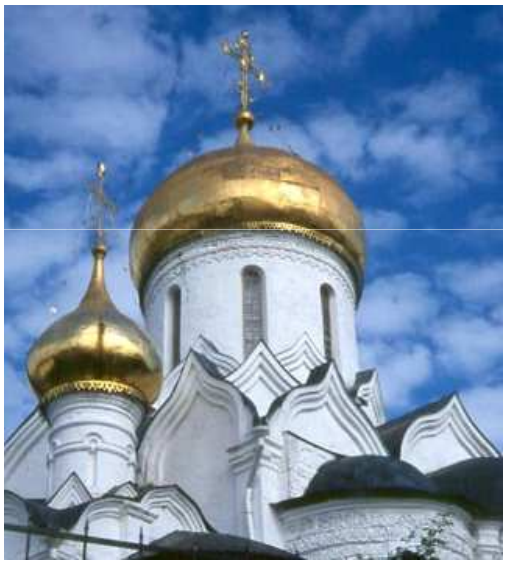


Courtois, Indocrypt 2008

Two Religions [Maths and Crypto]

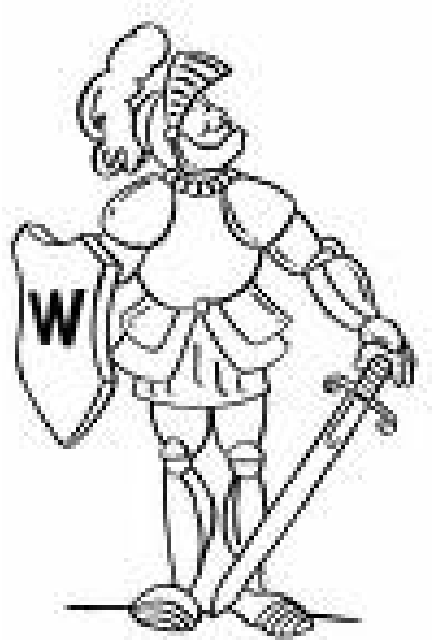
We will not agree on some questions any time soon…

Goal: learn each other's language.



tools

motivation



Mathematics

Cryptography

Frontiers

2. Algebraization, New Frontier in Symmetric Cryptography?

MQ Problem

Find a solution to a system of
m quadratic equations with
n variables
over a field/ring.

Cryptography and MQ

Claim: 95 % of all applied cryptography depends on the hardness of MQ.

1. RSA is based on MQ with m=1 and n=1: factoring N ⇔ solving $x^2=C \bmod N$.

Universality/completeness:
any polynomial system
can be written as quadratics
with added variables…

MQ Problem

Multivariate Version

[n variables]

Jean Dieudonné

[French Mathematician]

Book “Calcul infinitésimal”, Hermann, 1980

[..] Everybody in mathematics knows that going from one to several variables is an important jump that is accompanied by great difficulties and calls for completely new methods. […]

MQ Problem over GF(2)

Find a solution (at least one),

i.e. find $(x_0, \ldots, x_{n-1})$ such that:

$$\begin{cases} 1 = x_1 + x_0x_1 + x_0x_2 + \ldots \\ 0 = x_1x_2 + x_0x_3 + x_7 + \ldots \\ \vdots \end{cases}$$

More Applications of MQ

1. Public key schemes based on MQ directly, e.g. HFE [broken by Courtois, Joux and Faugère] and Sflash [broken by Stern, Shamir et al.]
2. If sparse MQ is easy, any block cipher including AES should be easy to break…
3. Dense MQ is VERY hard. In 2006 Patarin et al. Propose QUAD, a provably secure stream cipher based on MQ directly.
   - Open problem: propose a provably secure block cipher

Courtois, Indocrypt 2008

Schneier [Applied Cryptography book]

[…] Any algorithm that gets its security from the composition of polynomials over a finite field should be looked upon with scepticism, if not outright suspicion. […]

Written before AES ever existed…

Actually any cipher can be seen in this way…

Algebraization:

Theorem:

Every function over finite fields is a polynomial function.

[can be proven as a corollary of Lagrange's interpolation formula] $$P(X) = \sum_{i=1}^{t} Y_i \cdot \prod_{1 \le j \le t, j \neq i} \frac{X - X_j}{X_i - X_j}$$

False over rings!

What Can Said About Frontiers

Frontiers move:

The process can be called CONQUEST.

- Not always pejorative.

Maths – "Algebraization"

Mathematics:

Since, say the second half of XIX-th century, algebra is "conquering" other areas of mathematics. E.g.

- Algebraic Topology
- Algebraic Geometry
- Etc..

"Algebraization" of Cryptology

Since the 70s mathematics started conquering cryptology.

Before cryptography meant "bad mathematics" [at least according to Koblitz].

In April 2006 the NSA have officially decided that people "must/should" use Elliptic Curves [suite B].

The private sector failed to make the right choice [again].

And Cryptanalysis too!

Since the early 2000s, algebra is "conquering" cryptanalysis of ciphers, in order for:

- Algebraic public-key, like HFE [late 90s].
- Symmetric ciphers with algebraic components:
  - stream ciphers, KeeLoq, AES.
- Now, algebraization of ciphers that have no algebraic structure AT ALL, such as DES [Courtois-Bard, IMA Cryptography and Coding 2007 and eprint.iacr.org/2006/402/].

Any Progress?

Not many block ciphers are broken so far…

Some are:

KeeLoq, used by millions of people every day to open their cars, can be broken by an Algebraic Attack in practice.

[the full 528 rounds cipher, appears in FSE 2008]

The Role of Finite Fields

They allow to encode any cryptographic problem as problem of solving Boolean equations.

**The Role of NP-hard Problems

Guarantee “hardness” in the worst case.

Many are not that hard in practice…

There is hope and many concrete problems can be solved.

– Multiple reductions allow to use algorithms that solve one problem to solve another.

Algebraization:

- Algebraic Topology
- Algebraic Geometry
- Etc…

Works both ways, algebraic problems can also be viewed in geometric terms.

Example: Theory of T-functions is actually about ultra-metric Non-Archimedean geometry over 2-adic integers.

So maybe the “connection” will strike back!

Algebraization => Geometry-isation?!

Maybe now geometry may help to bring the topic of solving algebraic equations forward?

- Interesting new topics in cryptanalysis of symmetric ciphers to be studied now.

Maybe it is probably all already known in mathematics and we [cryptanalysts] just didn't realise it was there and can be applied to build efficient algorithms to solve systems of equations…

This is already done in number-theory based crypto: LLL is the "geometry of numbers" approach.

Symmetric Cryptanalysis:

From what one can observe:

bad news:

number of ciphers “broken w.r.t. claims”:

O(effort).

good news:

number of ciphers “broken in practice”:

o(effort).

Frontiers

3. New Territory: Algebraic Attacks on Ciphers

Propose New Ciphers ?

Foolish, requires lots of courage:

Ciphers “broken w.r.t. claims” = O(effort).

Algebraic Cryptanalysis of Block Ciphers ?

Also foolish, requires lots of courage:

so far EXCESSIVELY POOR results,
progress is slow. o(effort) ?

Algebraic Cryptanalysis [Shannon]

Breaking a « good » cipher should require:

"as much work as solving a system of simultaneous equations in a large number of unknowns of a complex type"

[Shannon, 1949]

Motivation

Linear and differential cryptanalysis usually require huge quantities of known/chosen plaintexts.

Q: What kind of cryptanalysis is possible when the attacker has

only one known plaintext (or very few) ?

Claim: This question did not receive sufficient attention. Misguided focus on LC and DC.

Two Worlds:

- **The "approximation" cryptanalysis:**
  - Linear, differential, high-order differential, impossible differential, Jakobsen-Knudsen approximation, etc..
  - All are based on probabilistic characteristics true with some probability.
  - Consequently, the security will grow exponentially with the number of rounds, and so does the number of required plaintexts in the attacks (main limitation in practice).
- **The "exact algebraic" approach:**
  - Write equations to solve, true with probability 1.
  - Very small number of known plaintexts required.

What's New ?

CLAIM:

The two worlds CANNOT be compared.

They are going in a very different direction:
what these two CAN ACHIEVE in practice are two very rich sets of cryptanalytic results that are rather disjoint.

Terra Incognita

…two sets of cryptanalytic results that are rather disjoint.

=> So we are really discovering a new frontier for the whole of symmetric cryptanalysis.

Symmetric Cryptanalysis:

Problem:

current metrics for achievement in symmetric cryptanalysis is deeply flawed. For example:

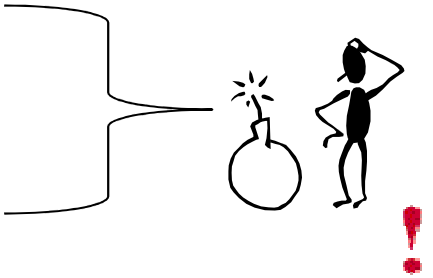
2^{43} KP is NOT better than 2^{56} and 1 KP.

DES was never really broken.

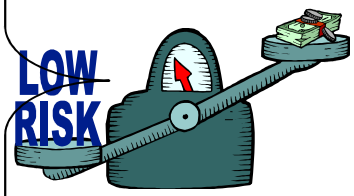
[Don Coppersmith, Crypto 2000].

Algebraic Attacks vs. DC/LC/etc..

- Algebraic attack in 2^{70} operations
  => the only feasible in the real life !



- Attacks with 2^{50} memory – infeasible.
- LC in 2^{30} operations – infeasible.
  – Hard to get 2^{30} KP !
- DC in 2^{20} operations – infeasible.
  – Hard to get 2^{20} CP !



Therefore:

- Computing power is the CHEAPEST resource. Should NO LONGER BE be the comparison metrics.
- Running time comparison with LC/DC is dishonest, makes little sense and should be avoided.

**Real-life Security Metrics:

$2^{70} = 2^{20}$:

An attack with 2^{70} is worth as much as with 2^{20} operations as both are feasible (!).

Compare these two attacks ONLY on:

- the number of required plaintexts
- KP/CP/CPCA etc.

=> Then, an algebraic attack in 2^{70} is worth as much as a differential attack in 2^{20} operations…

****Major Fallacy:

Gets worse - Remark:

by assuming that 2^{43} KP is feasible (it isn't) block ciphers have too many rounds.

Paranoid approach.

As a consequence, attacks that are really feasible, e.g. 2^{70} and 4 KP are never studied.

What to Expect from Algebraic Cryptanalysis

As much as from LC/DC/Etc.:

- Drop hope for practical attacks on AES for now…
- Goal: Just to advance research in symmetric cryptanalysis: what ciphers can be broken, how, and why.

A Strategic Problem

Are-they dead bodies in the closet?

- Expect that a couple of insecure ciphers exists under the cover of industrial/military secret.
  - Lightweight ciphers, designed some time ago, etc..
  - Don't expect me to break them. I don't have the spec.
- These will eventually come out… but for now we need to find substitutes to break – so that there is some progress in cryptanalysis!
- ECRYPT ESTREAM project:
  ciphers grown under "glass house".
  - Goal: make sure that ciphers are broken before being used, and not the opposite…

This was Stream Ciphers.

What About Block Ciphers?

"Glass House" for Algebraic Attacks of Block Ciphers

Web site dedicated to cooperation in algebraic cryptanalysis:

- Publish a system of equations that describe important practical ciphers [e.g. DES].
- Make other researchers compete in solving these.
- See where is the frontier:

  limitations of these attacks
  => new effective measure of security.

www.cryptosystem.net/aes/toyciphers.html

Research in Symmetric Cryptanalysis

I will now be “inspired” by some notations (C0,C1,C2) from Phil Rogaway: Formalizing Human Ignorance: Collision-Resistant Hashing without the Keys, eprint.iacr.org/2006/281

Question (C0): what ciphers can be broken.

Claim: This is an incorrect and misleading question. Existence doesn’t mean we can find them…

Phil Rogaway Talk:

"obviously we cannot found a scientific theory based on what people DO NOT know"

Later he says:

⇒ "Can take a human-ignorance approach for formalising properties of […] blockciphers, etc. "

~~Belief in hardness~~ (classical)
may be replaced by assuming ignorance ?

Maybe P = NP (there are fast algorithms)
but they are hard to find/invent, and not hard to run.

Research in Symmetric Cryptanalysis

(C0): can be broken.

incorrect and misleading

Better Formulation (Code Constructive or C1): What ciphers we (with our ignorance, background and available tools), can break in the next 50 years (and how ?).

What Ciphers We Can Break

(with our ignorance, background and available tools…) in the next 50 years ?

Well, it depends also what ciphers we WANT/TRY to break.

- The most precious resource is time and attention of clever people. Results will greatly depend on how this resource is being allocated. Ciphers that get attention are much more likely to be broken.
- Another scarce resource: CPU time and willing to experiment a lot… Maybe hardness is not where you think.

Weakness of Cryptographic Research Community

In fundamental physics, there are people that do the theory, and other people that design and handle experiments for their whole life.

Claim: we need this in symmetric crypto.

Otherwise we are not doing a lot of progress and are lying to everybody about some systems being not broken…

Research in Algebraic Cryptanalysis

Wishful thinking:

Theory

Proven Attack

The theory is almost never complete, [e.g. the complexity of XL or F5] and many algebraic attacks many attacks IMPREDICTABLE (much better / much worse than your theory).

Theory

Experimentation

Claim: this is not enough.

More Powerful Approach

Claim: many attacks will never be discovered if you do not experiment.

Fact

There are powerful tools that break certain ciphers without anybody knowing exactly why and when they work. (Cumulative effect of different phenomena that we can study separately).

The source code is usually not public. Non trivial implementation problems. E.g.

tiny subset(F4) >>
one version of F4
>>
another F5.

Tools – black-box (cryptanalytic oracles).

Reformulate the Goal Then (C2)

Black-box Constructive (C2):

what ciphers can be broken if I'm allowed to try my equations with

- Magma F4
- Faugère F5
- ElimLin [today]
- ANF-to-CNF and MiniSat [today]
- tools known to the NSA ???

Symmetric Crypto

Statistical Cryptanalysis:

Successful => More rounds are considered = >
Scarcity of attacks as only few combinations of biases give sufficiently strong overall bias.

Algebraic Cryptanalysis:

At present time: a handful of rounds, yet over-abundance of attacks to try.
MANY degrees of freedom.

Frontiers

4. Sources of Algebraic Vulnerability

Paradigm Shift

[Shannon, Jakobsen-Knudsen, Patarin, Pieprzyk-Courtois et al]

Look at multivariate algebraic relations (implicit equations).

Claim: This is the most general formulation of algebraic attacks [Carlet's Algebraic Immunity (AI) is a very, very restricted one].

Unified view of Algebraic Attacks

Non-existence of small multivariate relations between inputs/outputs.

- Applies to multivariate public key cryptosystems: Sflash, Quartz
- Applies to the non-linear part of a stream cipher, even if stateful.
- Applies to the S-boxes of a block cipher

Nicolas Courtois: General Principles of Algebraic Attacks and New Design Criteria for Components of Symmetric Ciphers,
In AES 4 Conference, LNCS 3373, Springer.

Def: "I / O Degree" = "Graph AI"

Consider function $f : GF(2)^n \rightarrow GF(2)^m$,
$f(x) = y$, with $x = (x_0, \ldots, x_{n-1})$, $y = (y_0, \ldots, y_{m-1})$.

Definition [The I/O degree] The I/O degree of f is the smallest degree of the algebraic relation

$$g(x_0, \ldots, x_{n-1}; y_0, \ldots, y_{m-1}) = 0$$

that holds with certainty for every couple (x, y) such that $y = f(x)$.

Design of Ciphers

When people design block cipher they usually study “ALL KNOWN ATTACKS” on it, then claim that the system is resistant to them.

My conjecture: it has become HARD to know and maybe THERE IS NO WAY to know, if a given system is resistant to all known attacks [particularly difficult for algebraic attacks].

What Can be Done ?

Algebraic Cryptanalysis:

- Very special ciphers: 1 M rounds [Courtois'AES4].
- General ciphers, key size=block size: SMALL number of rounds, 4,5,6 rounds.
  - Nobody can break CTC2(255,255,7).
- But no fundamental limitation of combinatorial nature [like in e.g. multiset attacks], more like:
  - 128-bit simple cipher – broken for 5-6 rounds.
  - 256-bit simple cipher – broken for 6-7 rounds.
  - 512-bit simple cipher – broken for 7-8 rounds. Etc.
- If key size > block size – more rounds.
  - CTC2(96,256,10) can be broken. 

- If many solutions (Hash functions, MACs) => expected to be still easier.

5. Wishful Thinking vs. Recent Results

"XSL is not an attack, it is a dream"

Vincent Rijmen

Def: "I / O Degree" = "Graph AI"

Consider function $f : GF(2)^n \rightarrow GF(2)^m$,
$f(x) = y$, with $x = (x_0, \ldots, x_{n-1})$, $y = (y_0, \ldots, y_{m-1})$.

Definition [The I/O degree] The I/O degree of f is the smallest degree of the algebraic relation

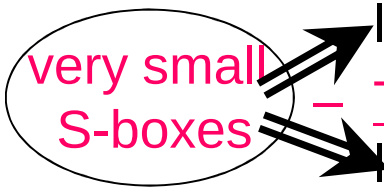
$$g(x_0, \ldots, x_{n-1}; y_0, \ldots, y_{m-1}) = 0$$

that holds with certainty for every couple (x, y) such that $y = f(x)$.

Early Work on Algebraic Attacks on Ciphers

- [2002] XSL paper: Are block ciphers [AES, Serpent] secure at all ?

3 “crazy” conjectures:



- I/O Degree Hypothesis (**IOH**): all ciphers with low I/O degree and lots of I/O relations may be broken ?
- The Very Sparse Hypothesis (**VSH**): ciphers with very low gate count broken ?
- The SubExponential Hypothesis (**SEH**): the security of block ciphers under IOH or VSH would NOT grow exponentially with the number of rounds ?
- XSL is one attack [presumably correct in principle, and presumably bad in practice] that might work and confirm these. But in fact; we need to find a better one…

**Early Work on Algebraic Attacks on Ciphers

- Work on XL, F4, F5 and relationship. (SEH)☠ for dense MQ.
- [1999-2004] HFE cryptanalysis [Courtois, Joux, Faugère] - IOH !.
- [2003-4]: Several LFSR-based stream ciphers badly broken: IOH* !.
- [2004] Weak ciphers broken for 1 M of rounds based on Inv S-box. (Strong results for IOH+SEH, but only for VERY special ciphers).
- [2004] Forget XSL do F5 instead. Extraordinarily poor results.
  - Yet shows the importance of diffusion/WTS. Nobody believes (IOH or VSH) is relevant for block ciphers.

**More Recent Work on Algebraic Attacks on Block Ciphers

- I/O degree Hypothesis (**IOH**).
- The Very Sparse Hypothesis (**VSH**).
- The SubExponential Hypothesis (**SEH**): ☠ for now…

- [2005] Work on T' method on extra-simplified versions of F4. Break CTC with more than 500 S-boxes. IOH +VSH !
- [2006] Attacks on DES, discovery of SAT solvers. VSH !.
- [2007] Some ciphers with small S-boxes are MUCH easier to break than others [ToySerpent << ToyRijndael]. New criteria on S-boxes that makes attacks easier/more difficult. First step to build a theory of algebraic attacks on block ciphers…

Summary: What I Do

Ciphers:

CTC, CTC2, ToySerpent, ToyRijndael, DES, KeeLoq.

research ciphers, NOT carefully designed…

real-life

programming / experimentation

Methods:

XSL+T', XL+T', F4 versions, ElimLin, ANF-to-SAT.

Algebraic Attacks on Block Ciphers

1. Write +
2. Solve [key recovery].

In fact, very early formulated as 3 stages.

XSL Attacks - Summary

Algebraic attacks on block ciphers work in 3 stages:

1. Write good equations – overdefined, sparse or both.
2. Expand - to obtain a very overdefined system.
3. Final "in place" elimination method – completely solve.

Historically 3. [XSL+T' method] was done separate because I did not know that 2. + 3. was already handled simultaneously by known GB techniques… Distinction not needed ?

Reinvent It:

Algebraic attacks on block ciphers today:

1. Write good equations – overdefined, sparse or both.
   - LESS TRIVIAL than expected [new tricks: higher degree, add variables, etc.].

   - avoid / minimise impact of…

3. Final "in place" deduction / inference / elimination method.
   - ElimLin alone and T' method. Amazingly powerful.
   - New tools [SAT solvers]. Amazingly powerful.

Algebraic Attacks on Block Ciphers

Gröbner Bases:

- Optimising the expansion step 2. at high degree.
- Mostly the dense case is understood and implemented.
- Then either AES-128 is broken at up to say 4 [Gwenolé Ars thesis: maybe it is?]. AND if not at this degree, it must be secure (!).

Fast Algebraic Attacks [will just explain]:

- EFFORT on 1. and 3.
- Avoid 2., start with BIGGER initial systems but never allow any expansion or increase in the degree.
- Sparse case ! Essential problems: heuristics to preserve sparsity, memory management.

Algebraic Attacks on Block Ciphers

Gröbner Bases, XL:

- How to avoid reduction to 0 while increasing the degree of polynomials.
- Mostly infeasible and impractical attacks…

Claim: A lot of research in a totally wrong direction. There so many much better methods to break ciphers. They are NOT more advanced/more sophisticated. On the contrary, they are much simpler.

Gröbner bases soon to be forgotten ?

NOT AT ALL, but attention must be shifted from high degree [all work on F5] to handling MUCH BIGGER systems but at VERY LOW DEGREE. Degrees between 1 and 3. Close to 1.

Powerful competitor: SAT Solvers + conversion.

Random sparse MQ:

- When both work, Magma F4 is faster (except uses 100 times as much RAM !!!).
- In many other cases our conversion + MiniSat just breaks in seconds systems that [it seems that] nobody would ever dream about solving.

6. Fast Algebraic Attacks On Block Ciphers

Fast Algebraic Attacks on Block Ciphers

Definition [informal on purpose] Methods to lower the degree of equations that appear throughout the computations… [e.g. max deg in F4]

(more generally need to substantially lower the memory requirements of algebraic attacks compared to their running time).

⇒ Very rich galaxy of attacks to be studied in the next 20 years…

How to lower the degree ?

- by having several P/C pairs (bigger yet much easier !)
- by CPA, CPCA, etc…
- by fixing internal variables (Guess-then-Algebraic).
- by finding [approximate] equations on bigger blocks
  - by interpolation [cf. W. Meier's talk]
  - by guessing equations that have strong bias
    - Linear-Algebraic or Bi-Linear-Algebraic Cryptanalysis
    - Differential-Algebraic.
- by clever choice of representation
- by introducing new variables (oh yes !)
- by having a larger key
- new tricks to be invented ?

cumulative effect !!!

How to Evaluate the Quality of Alg. Attacks

Compare ONLY to other similar attacks:

- Straightforward algebraic approach. Write + solve.
- Other attacks that work given VERY SMALL quantity of plaintexts.

- NEVER compare to DC/LC etc. Doesn't make sense. Two independent areas of research that have no intersection.

7. Solving Methods…

Fact

- Before 2005, I could break ciphers 12 S-boxes on 3 bits, key size 6 bits.
- Carlos Cid simulations with Magma: 10 S-boxes on 4 bits. 4 bit key.

In 2005-2006 huge progress have been made.

- Up to 510 S-boxes broken on a laptop:

Fast Algebraic attacks on block ciphers <= Cumulative effect of improvements in all these directions !

3.1. XSL + The T' method - Bad

Timings (old old implementation)
on a 1GHz laptop.

4 S-boxes:
Few minutes.

6 S-boxes:
40 minutes.

8 S-boxes:
Few hours.

12 S-boxes:
Few days
•
•

Example of how the rank grows:
(4 S-boxes).

7329 + 28
7329 + 52
7329 + 56
7329 + 96
7329 + 147
7329 + 165
7329 + 172
7329 + 173
7329 + 174

A unique solution found.
249.7 seconds

3.2. XL + T' method - Better

Gives better results already.
Closer to F4, F5 etc.

T' method (and extensions of it)

- Not easy to implement. I did 100 implementations and still not happy.
- Clearly less powerful than F4 in general, BUT seems more powerful assuming memory is very scarce…
- VERY powerful, there is nearly no system derived from symmetric crypto on which it really fails, just progresses again and again (but takes lots of time). Two open problems:
  - How to avoid reductions to 0 in the T' method => Big speed-up expected.
  - Better heuristics to preserve sparsity. => Feasibility extended ?

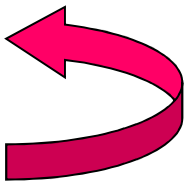
What's New

The biggest discoveries in Science are the simplest.

3.3. ElimLin – The Most Surprising.

Complete description:

- Find linear equations in the linear span.
- Substitute, and repeat.



Amazingly powerful, (Surprisingly) VERY HARD TO IMPLEMENT: 20 implementations and still not happy. Now able of handling systems of 1 M nonlinear equations (!) on a PC. Millions of monomials. Issues:

- Heuristics to preserve sparsity. Local optimization.
- Data Representation and Memory Management vs. Speed.

3.3. ElimLin – Remark:

In a way it is:

An ultra-light and super-simplified

version of F4 operating

at ”degree 1.2” or ”2.1”

(makes sense: relatively small number of higher-degree monomials, and certain types of monomials much more likely to ever appear).

3.4. ANF-to-CNF - The Outsider

Before we did try,

we actually never believed it could work…

☺ ☺ ☺

Convert MQ to a SAT problem.

(both are NP-hard problems)

3.4. ANF-to-CNF - The Outsider

Principle 1:

each monomial = one dummy variable.

$$a = wxyz$$

$$\Updownarrow$$

$$a \iff (w \wedge x \wedge y \wedge z)$$

$$\Updownarrow$$

$$(w \vee \bar{a})(x \vee \bar{a})(y \vee \bar{a})(z \vee \bar{a})(a \vee \bar{w} \vee \bar{x} \vee \bar{y} \vee \bar{z})$$

d+1 clauses for each degree d monomial

Also

Principle 2:

Handling XORs – Not obvious. Long XORs known to be hard problems for SAT solvers.

$$a \oplus b \oplus c \oplus d = 0$$

$$\Updownarrow$$

$$(\bar{a} \vee b \vee c \vee d)(a \vee \bar{b} \vee c \vee \bar{d})(a \vee b \vee \bar{c} \vee d)(a \vee b \vee c \vee \bar{d})$$
$$(\bar{a} \vee \bar{b} \vee \bar{c} \vee d)(\bar{a} \vee \bar{b} \vee c \vee \bar{d})(\bar{a} \vee b \vee \bar{c} \vee \bar{d})(a \vee \bar{b} \vee \bar{c} \vee \bar{d})$$

- Split longer XORs in several shorter with more dummy variables.
- About 4 h clauses for a XOR of size h.

ANF-to-CNF

This description is enough to produce a working version.

Space for non-trivial optimisations. See:

Gregory V. Bard, Nicolas T. Courtois and Chris Jefferson:

"Efficient Methods for Conversion and Solution of Sparse Systems of Low-Degree Multivariate Polynomials over GF(2) via SAT-Solvers".

Solving SAT

What are SAT solvers?

Heuristic algorithms for solving SAT problems.

- Guess some variables.
- Examine consequences.
- If a contradiction found, I can add a new clause saying “In this set of constraints one is false”.

Very advanced area of research.

Introduction for “dummies”:

Gregory Bard PhD thesis.

MiniSat 2.0.

Winner of SAT-Race 2006 competition.

An open-source SAT solver package,
by Niklas Eén, Niklas Sörensson,

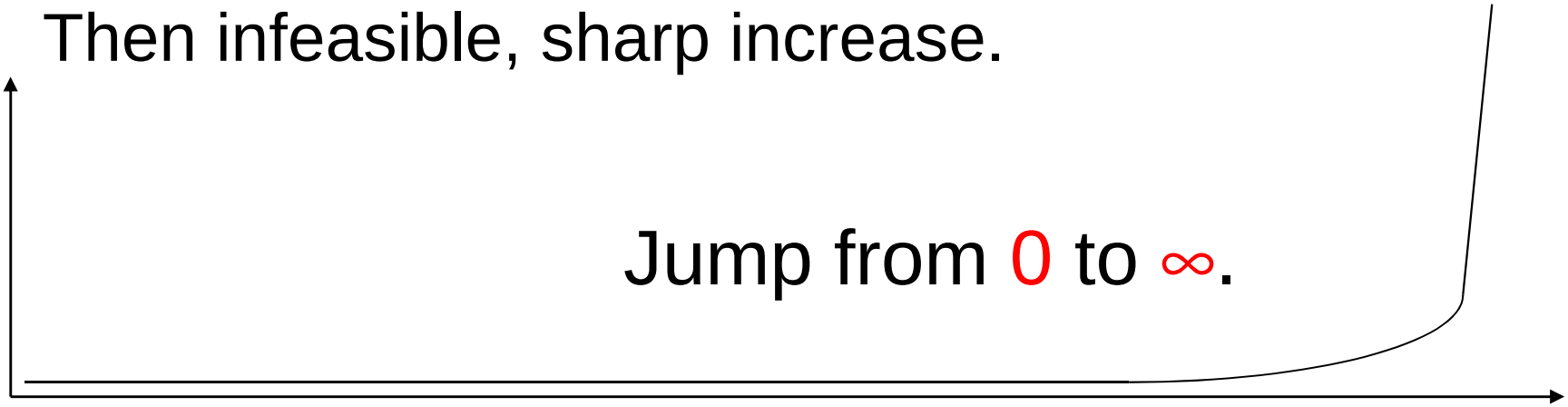
http://www.cs.chalmers.se/Cs/Research/FormalMethods/MiniSat/

Compiles with gcc under both Unix and Windows.

ANF-to-CNF + MiniSat 2.0.

Gives amazing results in algebraic cryptanalysis of just any (not too complex/not too many rounds) cipher, cf. (VSH). Also for random sparse MQ.

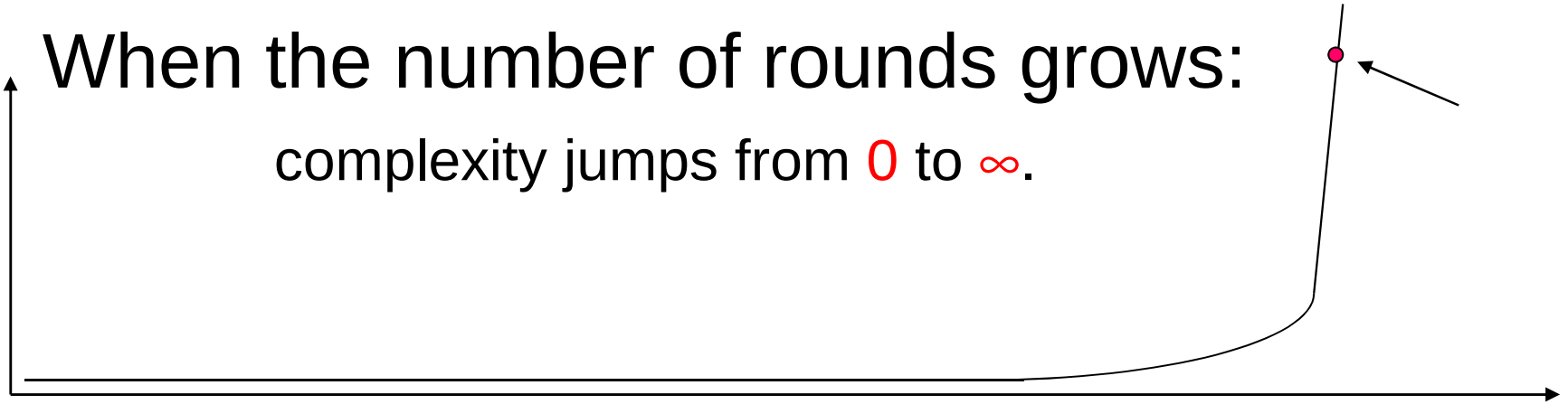
- Certain VERY large systems solved in seconds on PC (thousands of variables !).
- Few take a couple hours/days…
- Then infeasible, sharp increase.



What Are the Limitations of Algebraic Attacks ?

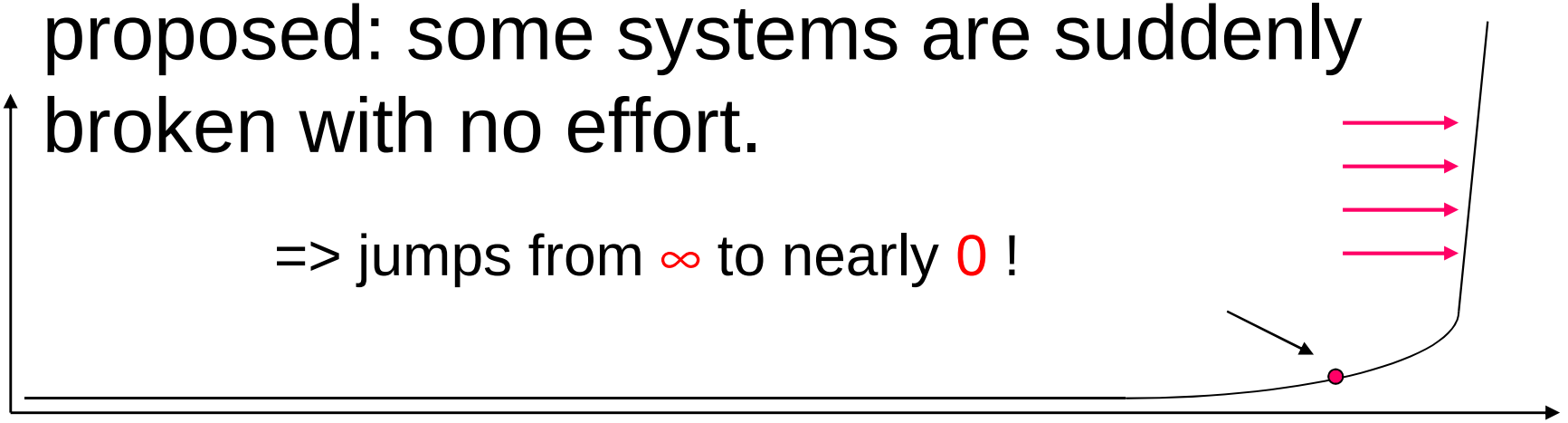
- When the number of rounds grows:

  complexity jumps from 0 to ∞.



- With new attacks and new “tricks” being proposed: some systems are suddenly broken with no effort.

  => jumps from ∞ to nearly 0 !



What Can Be Done with SAT Solvers ?

- Clearly it is not the size of the system but the nature of it.
- Sometimes more powerful than GB, sometimes less.

Paradoxes:

- If you guess some variables, can become much slower ☺.
- Great variability in results (hard to compute an average running time, better to look at 20 % faster timings).
- Memory:
  - For many cases tiny: 9 Mbytes while Magma hangs at > 2Gbytes for the same system.
  - For some working cases: 1.5 Gbytes and substantial time. Then terminates with the solution as well.

8. Toy Ciphers…

Summary – What I Do

Ciphers:

CTC, CTC2, ToySerpent, ToyRijndael, DES.

research ciphers, NOT carefully designed…

broken also by many other attacks

Design Summary: S-box + fast diffusion [avalanche effect].
Not much more, lazy designers and not ashamed of it.
Moreover we claim that it does not matter whether the cipher is secure w.r.t. LC or DC…

CTC = "Courtois Toy Cipher" [eprint]

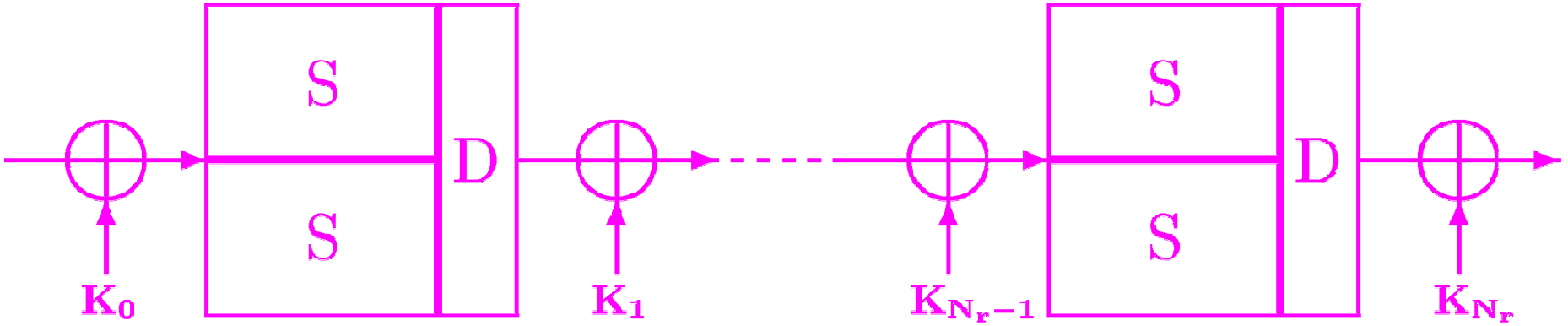


Fig. 1. A toy cipher with $B = 2$ S-boxes per round

- 3-bit S-boxes.
- Diffusion D: permuting wires (as DES P-box !).
- 1,2,4,8,... S-boxes per round.
- 1,2,3,...,10,...,30,... rounds.
- Key size == Block size.
- Simple key schedule: bit permutation (as in DES !)

CTC2

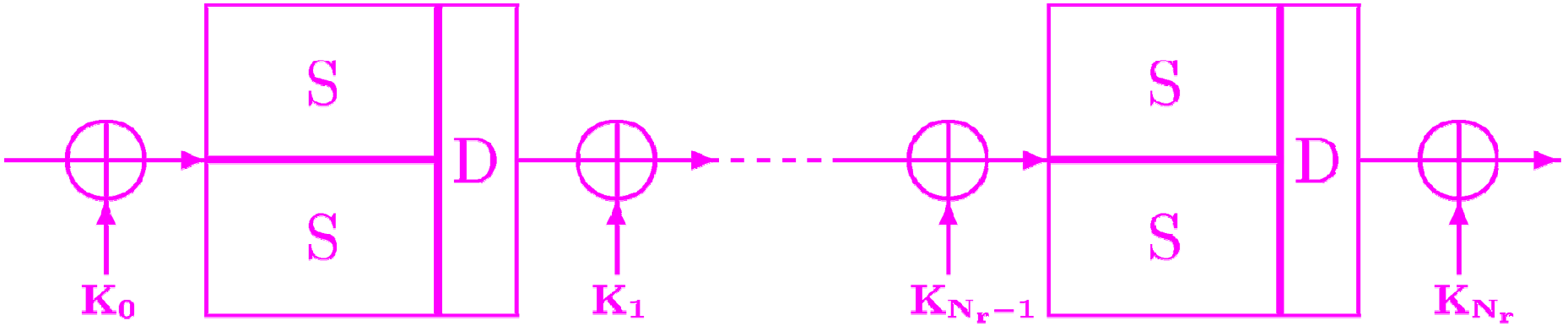


Fig. 1. A toy cipher with $B = 2$ S-boxes per round

- Virtually no difference
  - Different D-box but difference only at 1 bit position (!).
  - Changes <u>everything</u> w.r.t. linear cryptanalysis.
  - Changes <u>nothing</u> w.r.t. algebraic cryptanalysis.
    - In both cases 6 rounds are broken, 7 rounds maybe this year…

CTC vs. CTC2

$$\begin{cases} Z_{i\ (j\cdot 1987+257 \mod Bs)} = Y_{i\ j} & \text{if } j = 0 \\ Z_{i\ (j\cdot 1987+257 \mod Bs)} = Y_{i\ j} \oplus Y_{i\ (j+137 \mod Bs)} & \text{otherwise} \end{cases}$$

CTC2: Just remove one “weak” bit:

$$\begin{cases} Z_{i\ (j\cdot 1987+257 \mod Bs)} = Y_{i\ j} \oplus Y_{i\ (j+137 \mod Bs)} \oplus Y_{i\ (j+274 \mod Bs)} & \text{if } j = 257 \mod Bs \\ Z_{i\ (j\cdot 1987+257 \mod Bs)} = Y_{i\ j} \oplus Y_{i\ (j+137 \mod Bs)} & \text{otherwise} \end{cases}$$

No other difference. Same for “99 % of positions”.

CTC S-box:

Random on 3 bits without linear equations.

Theorem [Courtois]: 14 MQ Equations:

$$\begin{cases}
0 = x_1x_2 + y_1 + x_3 + x_2 + x_1 + 1 \\
0 = x_1x_3 + y_2 + x_2 + 1 \\
0 = x_1y_1 + y_2 + x_2 + 1 \\
0 = x_1y_2 + y_2 + y_1 + x_3 \\
0 = x_2x_3 + y_3 + y_2 + y_1 + x_2 + x_1 + 1 \\
0 = x_2y_1 + y_3 + y_2 + y_1 + x_2 + x_1 + 1 \\
0 = x_2y_2 + x_1y_3 + x_1 \\
0 = x_2y_3 + x_1y_3 + y_1 + x_3 + x_2 + 1 \\
0 = x_3y_1 + x_1y_3 + y_3 + y_1 \\
0 = x_3y_2 + y_3 + y_1 + x_3 + x_1 \\
0 = x_3y_3 + x_1y_3 + y_2 + x_2 + x_1 + 1 \\
0 = y_1y_2 + y_3 + x_1 \\
0 = y_1y_3 + y_3 + y_2 + x_2 + x_1 + 1 \\
0 = y_2y_3 + y_3 + y_2 + y_1 + x_3 + x_1
\end{cases}$$

Courtois, Indocrypt 2008

ToyRijndael and ToySerpent:

Basically a 4-bit version of CTC…

ToyRijndael S-box [4 bits]

Inv+Affine a in AES, borrowed from Carlos Cid.

Theorem [Courtois]: 21 MQ equations.

ToySerpent S-box [4 bits]

Sbox number 2 [chosen at random] stolen from Serpent [without permission from the authors].

Theorem [Courtois]: 21 MQ equations.

ToySerpent vs. ToyRijndael:

Both cases: 21 MQ equations.

Same degree, same number, yet TOTALLY DIFFERENT results (and we can explain why !).

Bad news for the idea (IOH) that I/O degree implies the existence of algebraic attacks.

- For some equations – good attacks [for 5 rounds].
- For some equations – little hope.

Rijndael S-box shows unexpected resistance w.r.t. our fast algebraic attack on block ciphers. [ElimLin].

Weakness in Serpent S-box 2:

4 / 21 equations of types

- 2 are "Linear+ X^2".

$$x2 + x3 + x4 + x1x3 + y1$$
$$x1 + x2 + x4 + x1x4 + y2 + y3 + y4 + 1$$

and

- 2 are "Linear+ Y^2".

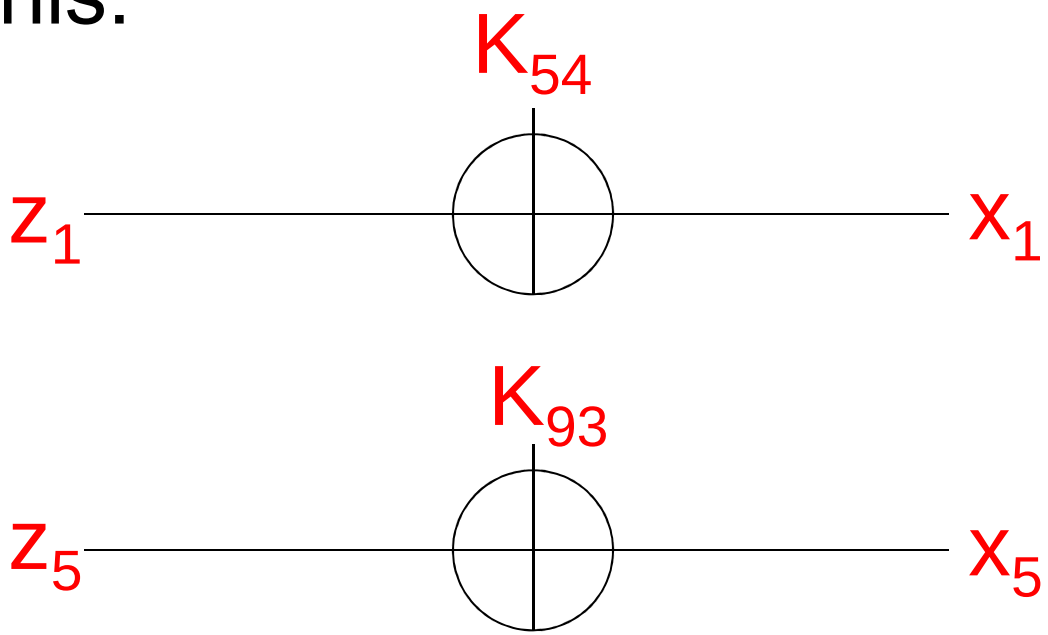
$$x1 + y1 + y2 + y3 + y2y3 + y2y4$$
$$x1 + x2 + x3 + y3 + y4 + y2y3 + y3y4 + 1$$

---

0 / 21 such equations for 4-bit Rijndael S-box !

What Is Special About These Equations ?

Look at this:



Observe that [combined syzygy]:

$$z_1 z_5 \oplus x_1 x_5 = K_{54} z_5 \oplus K_{93} z_1 \oplus K_{54} K_{93}$$

Combined Effect of These:

They allow to “avoid” / “lower the relative rank of” the set of higher degree monomials in the x_i in algebraic equations that can be written for several rounds.

In other words, some quadratic monomials / some linear combinations of monomials can be systematically eliminated:

$$x1 + x2 + x4 + x1x4 + y2 + y3 + y4 + 1$$

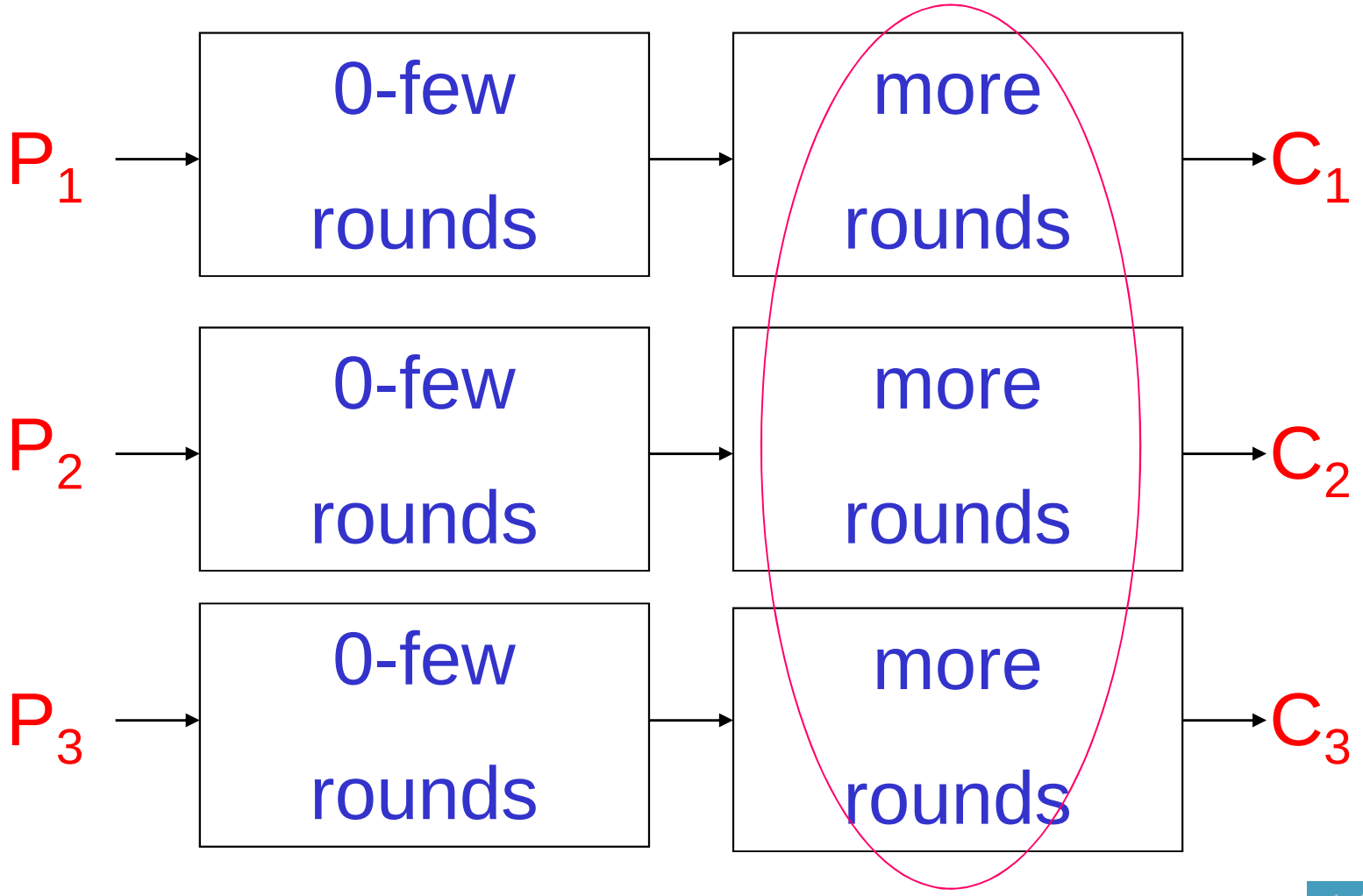
Claim: Will greatly help to compute Gröbner bases at a lower degree !

Now we will test the most optimistic version of this claim:

Replace F4 by ElimLin, how many linear equations can we generate ?

Interesting and WEIRD Question

KPA. How many linear equations true with Pr=1:



Courtois, Indocrypt 2008

Very Surprising and Powerful

Answer 1: They don't exist (cf. LC).

Answer 2: They DO exist when the P_i are fixed !

- Can be recovered by interpolation ? I did program this. Some toy examples take ages… Most relevant cases => infeasible ! Too large matrices.
- Fact: I have found a method to compute these equations VERY EFFICIENTLY given the set of plaintexts P_i. ← Arbitrary = a KPA.

Remark: A whole (big) part of the algebraic attacks that is done for a truncated cipher, i.e. without knowing the ciphertext - pre-computation possible give the spec. of the cipher (Pb. to use: CPA only).

When the P_i are fixed, how many equations ?

Nb. of linear equations found, 5 rounds x 3 S-boxes, KPA truncated (unknown ciphertext) ToySerpent & ToyRijndael.

Number of p-c pairs	1	2	4	8	16	32	64
Linear equations for ToySerpent(5,3)	0	7	27	75	171	748	3149
Number of round concerned	0	2	2	2	2	3	5
Linear equations for ToyRijndael(5,3)	0	0	0	0	0	0	0

Equations with rounds 0-5.

Some totally avoid the first 2 rounds. Rounds 3-5.

More powerful with full cipher (the ciphertexts are known => WORKS FROM both directions !!!! ElimLin even easier !

What About…

Real Life Ciphers ?

DES

At a first glance,

DES seems to be a very poor target:

there is (apparently)

no strong algebraic structure

of any kind in DES

What's Left ?

Idea 1: (IOH)

Algebraic I/O relations.

Theorem [Courtois-Pieprzyk]:

Every S-box has a low I/O degree.

=>3 for DES.

Idea 2: (VSH)

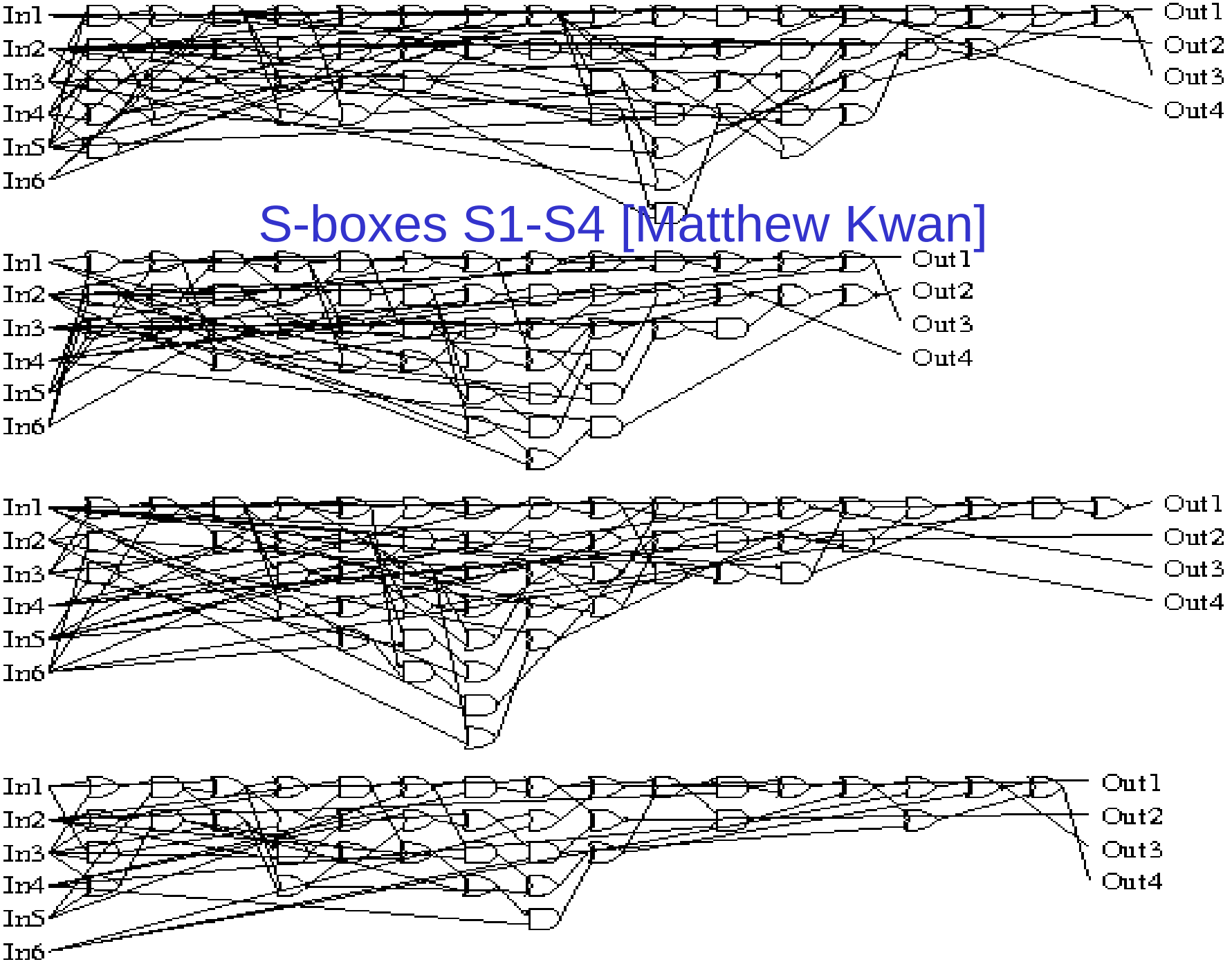
DES has been designed to be implemented in hardware.

=> Very-sparse quadratic equations at the price of adding some 40 new variables per S-box.

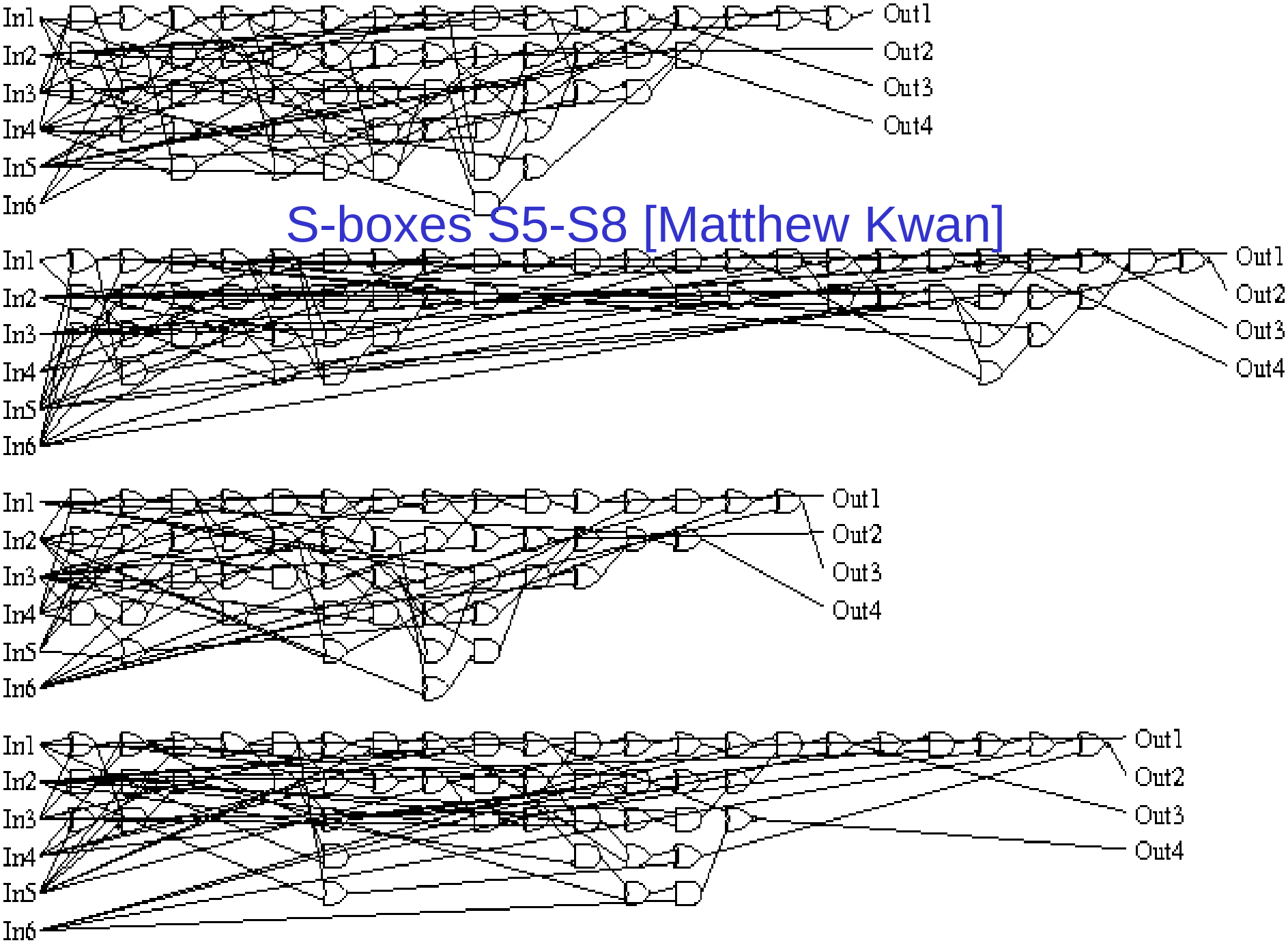
Results ?

Both Idea 1 (IOH) and Idea 2 (VSH)
(and some 20 other I have tried…)
can be exploited in working
key recovery attacks.

S-boxes S1-S4 [Matthew Kwan]



S-boxes S5-S8 [Matthew Kwan]



I / O Degree

Consider function $f : GF(2)^n \rightarrow GF(2)^m$,
$f(x) = y$, with $x = (x_0, \ldots, x_{n-1})$, $y = (y_0, \ldots, y_{m-1})$.

Definition [The I/O degree] The I/O degree of f is the smallest degree of the algebraic relation

$$g(x_0, \ldots, x_{n-1}; y_0, \ldots, y_{m-1}) = 0$$

that holds with certainty for every couple (x, y) such that $y = f(x)$.

Theorem

Theorem[Courtois]

For any $n \times m$ S-box, $F : (x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_m)$, and for any subset $\mathcal{T}$ of t out of 2^{m+n} possible monomials in the x_i and y_j, if $t > 2^n$, there are at least $t - 2^n$ linearly independent I/O equations (algebraic relations) involving (only) monomials in $\mathcal{T}$, and that hold with probability 1, i.e. for every (x, y) such that $y = F(x)$.

Courtois, Indocrypt 2008

Corollary

Cubic Equations and DES

$$t = 1 + (n+m) + \frac{(n+m)(n+m-1)}{2} +$$

$$+ \frac{(n+m)(n+m-1)(n+m-2)}{6} = 176$$

$$r \geq t - 2^n = 176 - 64 = 112.$$

Exactly 112 for all DES S-boxes.

9. Selected Results: Some Successful Attacks

Results on CTC

Nicolas T. Courtois:

"How Fast can be Algebraic Attacks on Block Ciphers ?". eprint.iacr.org/2006/168/

6 rounds broken: 255-bit key, 510 S-boxes.

ElimLin: 80 hours after 210/255 bits are guessed. 64 CP. About 10 times (slightly) faster than exhaustive search…

Results on CTC2

Much more resistant to LC

[cf. Orr Dunkelman and Nathan Keller : Linear Cryptanalysis of CTC, eprint.iacr.org/2006/250/].

ElimLin still breaks 6 rounds in the same way (no visible difference).

10 rounds broken if block=96, key=256.

Results on ToySerpent

ToySerpent, 5 rounds, 32 S-boxes * 4 bits.

84 first key bits guessed, 44 remain unknown.

4 CP => broken in 32 hours by ElimLin.

6 rounds should be feasible for 256-bit version.
Work in progress.

Results on ToyRijndael

Unexpectedly strong,
the only difference is the S-box:

0/21 “Linear+X^2“ equations...

Results on DES

Nicolas T. Courtois and Gregory V. Bard: “Algebraic Cryptanalysis of the D.E.S.”.

eprint.iacr.org/2006/402/

What Can Be Done ?

Idea 1 (Cubic IOH) + ElimLin:

We recover the key of 5-round DES with 3 KP faster than brute force.

- When 23 variables fixed, takes 173 s.
- Magma crashes > 2 Gb of RAM.

Idea 2 (VSH^{40}) + ANF-to-CNF + MiniSat 2.0.:

Key recovery for 6-round DES. Only 1 KP (!).

- Fix 20 variables takes 68 s.
- Magma crashes with > 2 Gb.

Frontiers

10. Limitations of Algebraic Cryptanalysis

What is Hard ?

The complexity of current attacks does grow exponentially with the number of rounds (but no limitation expected).

- Like 100x for each additional round…

So

- no hope for breaking full Serpent = 32 rounds
- Fact: 5 rounds Serpent is quite weak w.r.t. algebraic attacks, unlike 4-bit Rijndael S-box.

DES – New Frontier:

Break 8 rounds

given 1 KP and in less than 2^{55}.

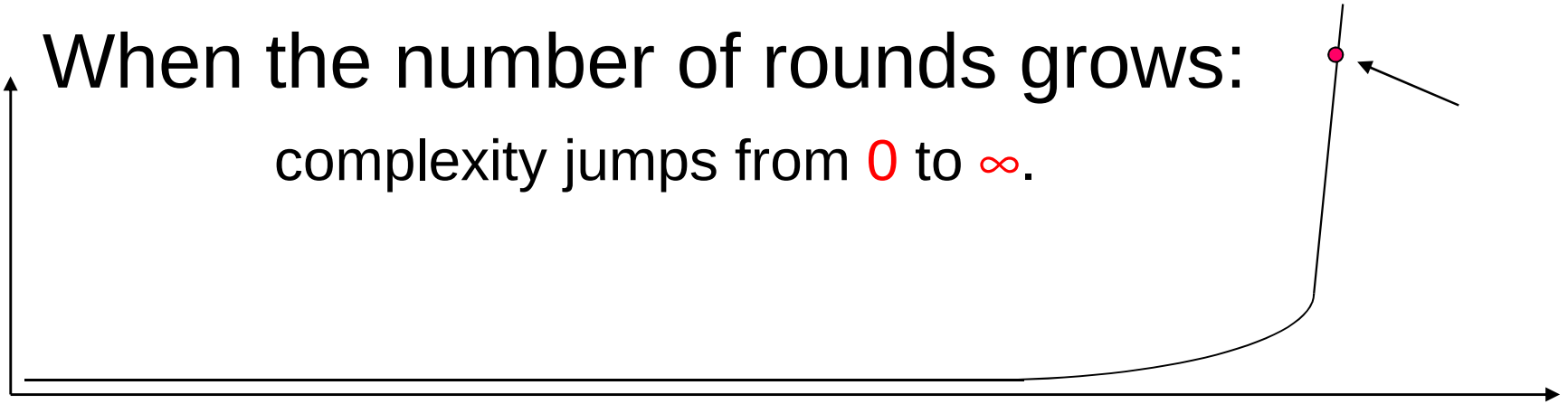
We encourage researchers to try.

We cannot do it so far.

What Are the Limitations of Algebraic Attacks ?

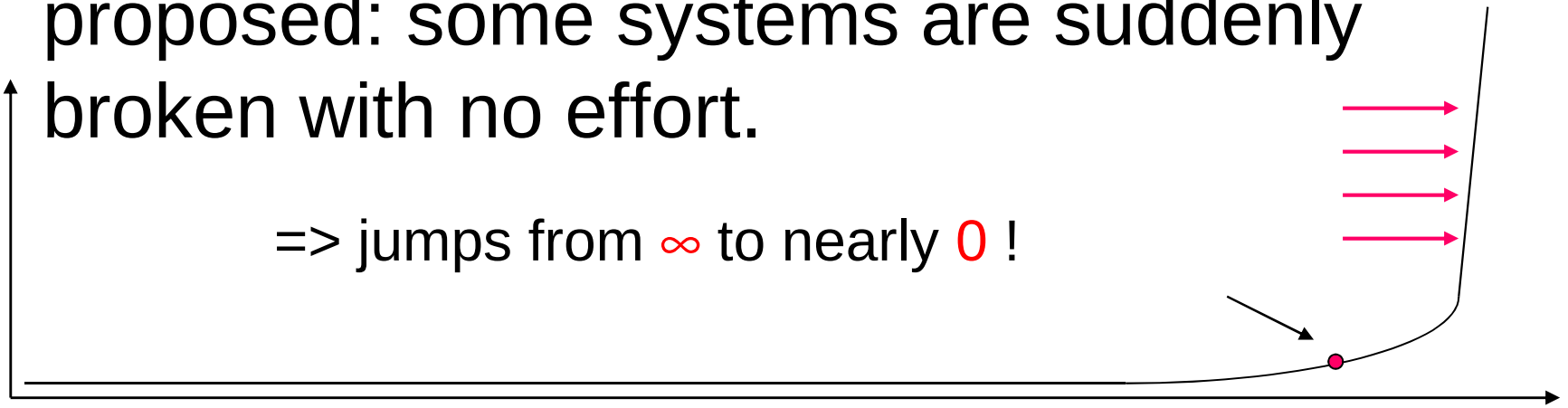
- When the number of rounds grows:

  complexity jumps from 0 to ∞.



- With new attacks and new “tricks” being proposed: some systems are suddenly broken with no effort.

  => jumps from ∞ to nearly 0 !



What About the 3 "Crazy" Conjectures ?

- I/O Degree Hypothesis (**IOH**): ciphers with low I/O degree and lots of relations may be broken ?

  - Positive results, but not as good for as (VSH).

- The Very-sparse Hypothesis (**VSH**): all ciphers with quite small gate-count broken ?

  - EXCELLENT results wit SAT solvers. OPTIMISTIC !

- The SubExponential Hypothesis (**SEH**): the security of vulnerable block ciphers would NOT grow exponentially with the number of rounds ?

  - So far, no working key recovery attack on >10 rounds, not even for CTC.

Final Conclusion

Some limitations of algebraic cryptanalysis are very hard, we “hit the wall” (e.g. when the number of rounds increases).

Some are spectacularly naïve (e.g. maximum degree in Gröbner basis computation) and are easily circumvented.