

# Phase Transition of Multivariate Polynomial Systems

**Giordano Fusco** and **Eric Bach**

Computer Sciences Department  
University of Wisconsin-Madison  
{fusco, bach}@cs.wisc.edu

Theory and Applications of Models of Computation  
May 25, 2007

# Outline

- 1 Introduction
- 2 Main Results
- 3 Experimental Results
- 4 Applications

# Outline

- 1 Introduction
- 2 Main Results
- 3 Experimental Results
- 4 Applications

# The problem

## Random multivariate quadratic system

- $m$  equations in  $n$  variables
- each equation has the form

$$a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{22}x_2^2 + a_{23}x_2x_3 + \cdots + b_1x_1 + \cdots + b_nx_n = c$$

- coefficients independently and uniformly distributed on  $GF(p)$
- in general coefficients up to degree  $d$

## The questions

- What is the probability that the system has no solutions?
- What is the probability that the system has exactly 1 solution?
- What is the probability that the system has exactly  $s$  solutions?

# The problem

## Random multivariate quadratic system

- $m$  equations in  $n$  variables
- each equation has the form

$$a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{22}x_2^2 + a_{23}x_2x_3 + \cdots + b_1x_1 + \cdots + b_nx_n = c$$

- coefficients independently and uniformly distributed on  $GF(p)$
- in general coefficients up to degree  $d$

## The questions

- What is the probability that the system has no solutions?
- What is the probability that the system has exactly 1 solution?
- What is the probability that the system has exactly  $s$  solutions?

# Motivation

## Cryptographic systems transformed into quadratic systems

- Many cryptographic systems can be transformed into large quadratic systems.
- The solution of this quadratic system is unique because it represents the decoded text.
- One of the methods to solve quadratic systems is called XL, and it was first proposed by Courtois, Klimov, Patarin, and Shamir.
- The authors of XL argue that their method takes advantages of the uniqueness of the solution.
- Quadratic systems from cryptography are not perfectly random, but in absence of a better theory we are assuming that they are.
- Knowing the probability of exactly one solution, allows to understand how often XL has the claimed advantage.

# The phase transition

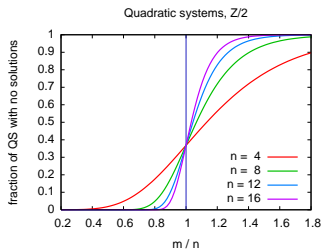
$1/e$  in a prime field ( $m$  equations in  $n$  variables)

## Probability of no solutions

- $m < n \Rightarrow P[\text{no solutions}] \rightarrow 0$
- $m = n \Rightarrow P[\text{no solutions}] = 1/e$
- $m > n \Rightarrow P[\text{no solutions}] \rightarrow 1$

## Probability of exactly 1 solution

- $m < n \Rightarrow P[1 \text{ solution}] \rightarrow 0$
- $m = n \Rightarrow P[1 \text{ solution}] = 1/e$
- $m > n \Rightarrow P[1 \text{ solution}] \rightarrow 0$



# The phase transition

$1/e$  in a prime field ( $m$  equations in  $n$  variables)

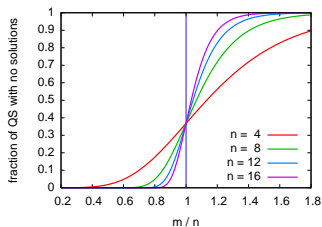
## Probability of no solutions

- $m < n \Rightarrow P[\text{no solutions}] \rightarrow 0$
- $m = n \Rightarrow P[\text{no solutions}] = 1/e$
- $m > n \Rightarrow P[\text{no solutions}] \rightarrow 1$

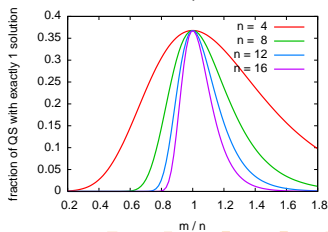
## Probability of exactly 1 solution

- $m < n \Rightarrow P[1 \text{ solution}] \rightarrow 0$
- $m = n \Rightarrow P[1 \text{ solution}] = 1/e$
- $m > n \Rightarrow P[1 \text{ solution}] \rightarrow 0$

Quadratic systems,  $\mathbb{Z}/2$



Quadratic systems,  $\mathbb{Z}/2$



# Outline

- 1 Introduction
- 2 Main Results**
- 3 Experimental Results
- 4 Applications

# Probability of no solutions

The phase transition

## Theorem 1

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $GF(p)$  (where  $p$  is a prime number),  
 $\Rightarrow$  the probability that the system has no solution is  $e^{-p^{-\alpha}}$ , asymptotically in  $n$ .

## Corollary 1

- For a system like in theorem 1, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )  
 $\Rightarrow$  the probability of no solutions is  $e^{-1}$ .

# Probability of no solutions

The phase transition

## Theorem 1

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $GF(p)$  (where  $p$  is a prime number),  
 $\Rightarrow$  the probability that the system has no solution is  $e^{-p^{-\alpha}}$ , asymptotically in  $n$ .

## Corollary 1

- For a system like in theorem 1, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )  
 $\Rightarrow$  the probability of no solutions is  $e^{-1}$ .

# Probability of exactly $s$ solutions

The general theorem for  $p$  prime

## Theorem 2

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $GF(p)$  (where  $p$  is a prime number),  
 $\Rightarrow$  the probability that the system has exactly  $s$  solutions follows the Poisson distribution  $\lambda^s e^{-\lambda} / s!$ , asymptotically in  $n$ , where  $\lambda = e^{-\alpha \log p}$ .

## Corollary 2

- For a system like in theorem 2, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )  
 $\Rightarrow$  the probability that the system has exactly  $s$  solutions is  $e^{-1} / s!$ .

# Probability of exactly $s$ solutions

The general theorem for  $p$  prime

## Theorem 2

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $GF(p)$  (where  $p$  is a prime number),  
 $\Rightarrow$  the probability that the system has exactly  $s$  solutions follows the Poisson distribution  $\lambda^s e^{-\lambda} / s!$ , asymptotically in  $n$ , where  $\lambda = e^{-\alpha \log p}$ .

## Corollary 2

- For a system like in theorem 2, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )  
 $\Rightarrow$  the probability that the system has exactly  $s$  solutions is  $e^{-1} / s!$ .

# Key idea of the proof for theorem 1

## Probability of no solutions

- Let  $P_1, \dots, P_t$  be a set of inputs, each representing a possible assignment to the variables.
- Consider the events of satisfying all equations with each  $P_i$ .
- The proof would be easy if these events were independent, but they are not in general.
- For enough of the  $k$ -subsets we do have independence, so asymptotically the result is like if we had independence.

# Probability of no solutions in $\mathbb{Z}/(pq)$

The phase transition in  $\mathbb{Z}/(pq)$

## Theorem 3

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $\mathbb{Z}/(pq)$  with  $p$  and  $q$  distinct primes.

⇒ the probability that the system has no solution is  $e^{-p^{-\alpha}} + e^{-q^{-\alpha}} - e^{-(p^{-\alpha} + q^{-\alpha})}$ , asymptotically in  $n$ .

## Corollary 3

- For a system like in theorem 3, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )

⇒ the probability of no solutions is  $2e^{-1} - e^{-2}$ .

# Probability of no solutions in $\mathbb{Z}/(pq)$

The phase transition in  $\mathbb{Z}/(pq)$

## Theorem 3

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $\mathbb{Z}/(pq)$  with  $p$  and  $q$  distinct primes.

⇒ the probability that the system has no solution is  $e^{-p^{-\alpha}} + e^{-q^{-\alpha}} - e^{-(p^{-\alpha}+q^{-\alpha})}$ , asymptotically in  $n$ .

## Corollary 3

- For a system like in theorem 3, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )

⇒ the probability of no solutions is  $2e^{-1} - e^{-2}$ .

# Probability of exactly $s$ solutions in $\mathbb{Z}/(pq)$

The general theorem in  $\mathbb{Z}/(pq)$

## Theorem 4

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $\mathbb{Z}/(pq)$  with  $p$  and  $q$  distinct primes.
- $\Rightarrow$  the probability that the system has exactly  $s$  solutions follows the Poisson distribution  $e^{-\lambda - \mu} \sum_{u,v \geq 1}^{uv=s} \frac{\lambda^u \mu^v}{u! v!}$ , asymptotically in  $n$ , where  $\lambda = e^{-\alpha \log p}$  and  $\mu = e^{-\alpha \log q}$ .

## Corollary 4

- For a system like in theorem 4, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )
- $\Rightarrow$  the probability that it has exactly  $s$  solutions is  $e^{-2} \sum_{u,v \geq 1}^{uv=s} \frac{1}{u! v!}$ .

# Probability of exactly $s$ solutions in $\mathbb{Z}/(pq)$

The general theorem in  $\mathbb{Z}/(pq)$

## Theorem 4

- Given a multivariate polynomial system of  $n + \alpha$  random equations of degree  $d$  (with  $d \geq 2$ ) in  $n$  variables over  $\mathbb{Z}/(pq)$  with  $p$  and  $q$  distinct primes.
- $\Rightarrow$  the probability that the system has exactly  $s$  solutions follows the Poisson distribution  $e^{-\lambda - \mu} \sum_{\substack{uv=s \\ u, v \geq 1}} \frac{\lambda^u \mu^v}{u! v!}$ , asymptotically in  $n$ , where  $\lambda = e^{-\alpha \log p}$  and  $\mu = e^{-\alpha \log q}$ .

## Corollary 4

- For a system like in theorem 4, if the number of equations equals the number of variables (i.e.  $\alpha = 0$ )
- $\Rightarrow$  the probability that it has exactly  $s$  solutions is  $e^{-2} \sum_{\substack{uv=s \\ u, v \geq 1}} \frac{1}{u! v!}$ .

# Outline

- 1 Introduction
- 2 Main Results
- 3 Experimental Results**
- 4 Applications

# Experimental Results

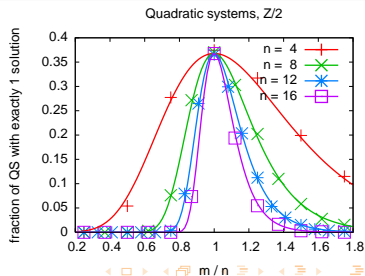
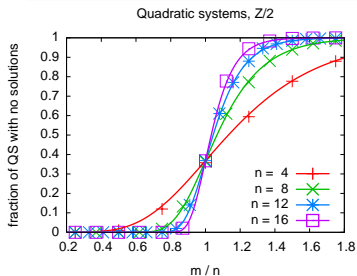
## Why do we need experimental results?

- The proofs cover the case in which the number of equations is close to the number of variables (and only asymptotically).
- Experimental results confirm that the formulas have a wider range of application.
- Variance of the order of  $10^{-5}$ .
- Good because uniform at random values would give  $10^{-2}$ .

# Experimental Results

## Why do we need experimental results?

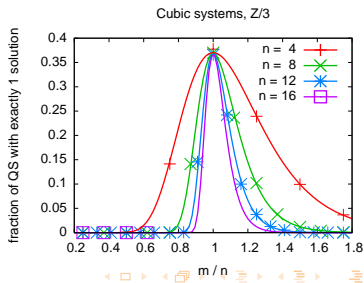
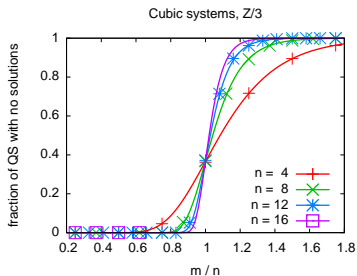
- The proofs cover the case in which the number of equations is close to the number of variables (and only asymptotically).
- Experimental results confirm that the formulas have a wider range of application.
- Variance of the order of  $10^{-5}$ .
- Good because uniform at random values would give  $10^{-2}$ .



# Linearly independent equations

The formulas are confirmed in this case

- Quadratic systems derived from cryptography have only linearly independent equations.
- The equations of a random polynomial system are linearly independent with very high probability.
- Also in this case the variance is of the order of  $10^{-5}$ .



# Sparse systems

Quadratic systems from cryptography are sparse

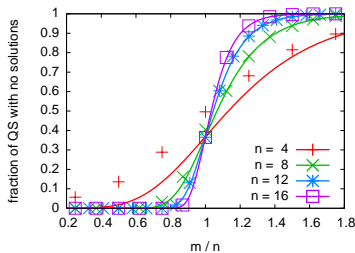
## Three type of sparseness

- 1 Each coefficient is 0 with probability  $z$  and non zero with probability  $1 - z$ .
- 2 Each equation contains exactly a fraction  $f$  of the variables.
- 3 Bi-affine equations (this is the case of Rijndael). The variables are partitioned in two sets of equal size. Each quadratic term is composed of a variable from the first set and one from the second set.

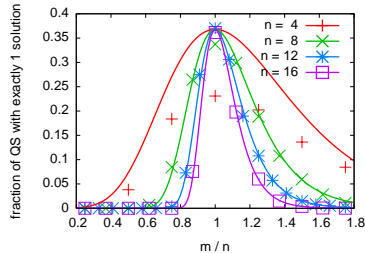
# Coefficients have higher probability to be 0

- Each coefficient is 0 with probability  $z$  and non zero with probability  $1 - z$ .
- We tried values of  $z$  from 0.5 to 0.9.
- Variance of the order of  $10^{-5}$  for  $z$  up to 0.7.

Quadratic systems,  $Z/2$ , sparse, coefficients are 0 with probability  $z$ :



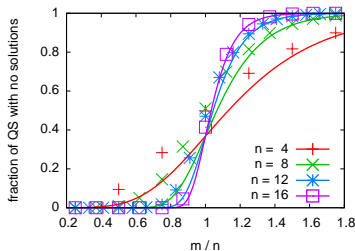
Quadratic systems,  $Z/2$ , sparse, coefficients are 0 with probability :



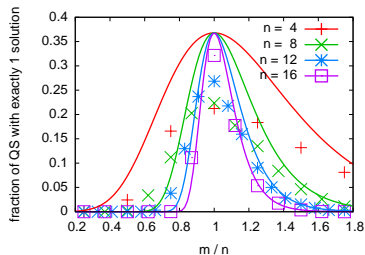
# Each equations has only a fraction of variables

- Each equation contains exactly a fraction  $f$  of the variables.
- We tried values of  $f$  from 0.1 to 0.5.
- Variance of the order of  $10^{-3}$  for  $f = 0.5$ .
- A possible explanation is that the coefficients are not uniformly independently distributed.

Quadratic systems,  $Z/2$ , sparse, 50% variables/equation



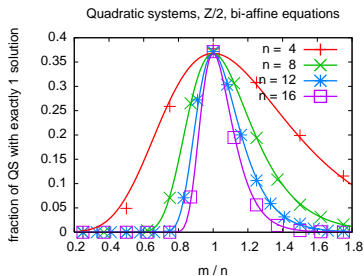
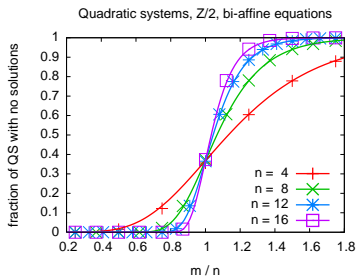
Quadratic systems,  $Z/2$ , sparse, 50% variables/equation



# Bi-affine equations

These are the equations of Rijndael

- The variables are partitioned in two sets of equal size. Each quadratic term is composed by a variable from the first set and one from the second set.
- Variance of the order of  $10^{-5}$



# Outline

- 1 Introduction
- 2 Main Results
- 3 Experimental Results
- 4 Applications**

# Quadratic systems from cryptography

- The positive experimental results give us confidence in using the formulas also in cases not strictly covered by the proofs.

<i>Cryptosystem</i>	$n$	$m$	$\alpha$	Total # of systems	Pr[1 solution]
Khazad	6464	7664	1200	$6.86 \cdot 10^{6249185}$	$5.81 \cdot 10^{-362}$
Misty1	3856	3856	0	$1.68 \cdot 10^{2239709}$	$1/e$
Kasumi	4264	4264	0	$4.20 \cdot 10^{2738543}$	$1/e$
Camellia-128	3584	6224	2640	$1.64 \cdot 10^{1934992}$	$1.91 \cdot 10^{-795}$
Rijndael-128	3296	6296	3000	$5.40 \cdot 10^{1636625}$	$8.13 \cdot 10^{-904}$
Serpent-128	16640	17680	1040	$3.58 \cdot 10^{41683551}$	$8.49 \cdot 10^{-314}$

- For all but 2 cryptosystems, the chance of having exactly one solution is extremely small, but the number of systems with exactly one solution is not that small, because the number of systems is huge.

# Open problems

- Prove that the formulas are valid in general, not only near the phase transition.
- Find more precise formulas for sparse systems, in particular when each equation contains exactly a fixed number of variables.

# Questions ?

Giordano Fusco  
fusco@cs.wisc.edu