Relinearization attack on LPN over $\mathbb{F}_p$
CFAIL 2022

Paul Lou, Amit Sahai, Varun Sivashankar

UCLA

August 2022

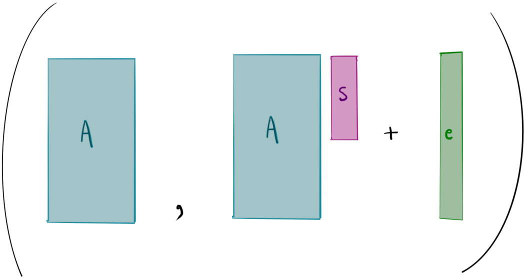
Why do we care about attacking LPN over large fields?

- LPN over large fields [IPS09] is an important assumption in current indistinguishability obfuscation constructions [JLS21].
- Important to understand its security: so far a naive sub-exponential guessing algorithm is still the state-of-the-art.

Will Gröbner Bases Work?

Does a linearization/Gröbner bases attack work? So far, nope :(

LPN over $\mathbb{F}_p$



- $\mathbf{A} \leftarrow \mathbb{F}_p^{m \times n}$, $\mathbf{s} \leftarrow \mathbb{F}_p^n$ where p is a λ-bit prime (sec. param λ).
- For sparsity constant γ, for $i \in [m]$, $\mathbf{e}_i \leftarrow \begin{cases} \mathbb{F}_p & \text{with prob. } n^{-\gamma} \\ 0 & \text{otherwise} \end{cases}$
- Number of equations $m = n^{1+\alpha}$ (Think constant $\alpha < 1$).

Goal: Recover $\mathbf{s}$ from $(\mathbf{A}, \mathbf{As} + \mathbf{e})$ (unique $\mathbf{s}$ w.h.p.)

Some use cases of LPN over $\mathbb{F}_p$ in cryptography

Using the decisional variant (there's a search-to-decision reduction):

- Public-key encryption [Ale03; DP12; AAB15] (when sparsity $\gamma \geq 1/2$)
- Vector oblivious linear-function evaluation (VOLE) generators [Boy+18]
- Indistinguishability obfuscation [JLS21]

Known attack landscape for search LPN over $\mathbb{F}_p$

- No known reductions between LPN over $\mathbb{F}_p$ and LWE (different error distributions).
- **Folklore attack** (low noise rate $n^{-\gamma}$): repeatedly take n samples, assume error-free, and solve for **s** via Gaussian elimination [Car+09; EKM17].
  - Expected runtime: $1/(1-n^{-\gamma})^n$.
    - If $\gamma \geq \frac{1}{2}$, this is $O\left(e^{n^{1-\gamma}}\right)$.
    - If $\gamma < 1/2$, then it's $e^{O(n^{1-\gamma})}$.
  - Sample complexity: $O(n^{1+\gamma})$.
- Information set decoding and variants [Pra62; CS16].
- For high noise rate (e.g. constant): BKW algorithm with runtime, memory, and sample complexity $O\left(2^{n/\log n}\right)$ [BKW]. Scaled-down version works with polynomial sample size but worse runtime [Lyu05].
- What about Gröbner basis attacks?

Our objective and contributions

- **Our regime**: low noise rate $n^{-\gamma}$ and sample complexity $m = n^{1+\alpha}$ for $\alpha \in (0, 1)$.
  - No known attack better than the folklore attack.
- **Objective**: Find a better subexponential attack via a Gröbner basis approach.

We didn't succeed.

- Our approach only yields an exponential time attack, assuming a widely believed conjecture about "semi-regularity".
- We discuss the approaches we tried and some open questions.

What is linearization?

Linearization technique [KS99; AG11]: replace all the monomials with a new set of variables to obtain a linear system

$$x_1 \mapsto y_1$$
$$x_1 x_2 \mapsto y_{1,2}$$
$$x_1 x_2 + x_1 + 3 \mapsto y_{1,2} + y_1 + 3$$

- Starting with m degree-d equations, the number of monomials present is the number of new variables. At most $n' = \binom{n+d}{d}$ many.
- If initially there was a unique solution and the number of equations m is sufficiently larger than n', then the linearized system has the same unique solution with high probability.
- Solving the resulting polynomial system takes time approximately $O\left((n')^{\omega}\right)$ for linear algebra constant $2 \leq \omega \leq 3$.

An example: linearization attack on Binary LWE

Binary LWE setting: each error $e_i \in \{0, 1\}$ where $e_i \sim \text{Ber}(\tau)$. Given $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e})$, recover $\mathbf{s}$.

Due to Arora-Ge [AG11]:

- Since the errors are in $\{0, 1\}$, we have m degree 2 equations in $s_1, \ldots, s_n$:

$$\left\{(b_i - \mathbf{a}_i \cdot \mathbf{s} - 1) \cdot (b_i - \mathbf{a}_i \cdot \mathbf{s}) = 0 \bmod p\right\}_{i \in [m]}$$

- Linearize $s_i \mapsto y_i$, $s_i s_j \mapsto y_{i,j}$. Number of linearized variables is $O(n^2)$.
- This gives polynomial time recovery if $m = \Omega(n^2)$.
- Sample-time tradeoff for samples $m = n^{1+\alpha}$ characterized by Sun et al. [STA20].

Smaller sample complexity $m \sim n^{1+\alpha}$ (1/2)

- **Issue**: When $m \sim n^{1+\alpha}$ for $\alpha \in (0, 1)$, there are not enough equations for the linearized system to have a unique solution.
- **Goal**: Generate more equations.

Smaller sample complexity $m \sim n^{1+\alpha}$ (2/2)

Degree-d **Macaulay expansion**: multiply every equation by all monomials up to degree d (can view as a matrix of coefficients, the Macaulay matrix):

$$
\begin{aligned}
&\left\{ f_i(x_1, \ldots, x_n) \right\}_{i \in [m]} \\
&\quad \cup \left\{ x_1 \cdot f_i(x_1, \ldots, x_n) \right\}_{i \in [m]} \\
&\quad \cup \left\{ x_2 \cdot f_i(x_1, \ldots, x_n) \right\}_{i \in [m]} \\
&\quad \cup \left\{ x_1 x_2 \cdot f_i(x_1, \ldots, x_n) \right\}_{i \in [m]} \\
&\quad \vdots
\end{aligned}
$$

Macaulay expansion finds our unique solution

- **Intuition**: if there is a unique solution to $\{f_1(\mathbf{x}) = 0, \ldots, f_m(\mathbf{x}) = 0\}$, say $\mathbf{s}$, then Hilbert's Nullstellensatz says ideal

$$\langle f_1, \ldots, f_m \rangle$$

is equivalent to the ideal (whose generators are our Gröbner basis)

$$\langle x_1 - s_1, \ldots, x_n - s_n \rangle.$$

Therefore, there exist some polynomials (WLOG of minimal degree) $\{g_{i,j}\}_{i \in [m], j \in [n]}$ such that for all $j \in [n]$

$$x_j - s_j = \sum_{i \in [m]} g_{i,j} \cdot f_i$$

Punchline: Expand until we can recover the Gröbner basis $(x_1 - s_1, \ldots, x_n - s_n)$.

Macaulay expansion intimately related with Gröbner bases

Computing a Gröbner basis for a homogeneous polynomial system $(f_1, \ldots, f_m)$ is equivalent to performing Gaussian elimination on Macaulay matrices [Laz83].

Recall LPN over $\mathbb{F}_p$

Recall the setup: Our input is $(\mathbf{A}, \mathbf{b})$ where

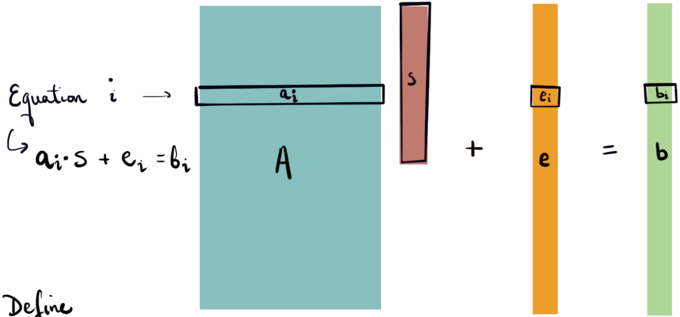
- $\mathbf{A} \leftarrow \mathbb{F}_p^{m \times n}$ where $m = n^{1+\alpha}$ samples, $\alpha \in (0, 1)$ constant.
- $\mathbf{b} = \mathbf{A} \cdot \mathbf{s} + \mathbf{e}$ where $\mathbf{s} \leftarrow \mathbb{F}_p^n$ and $\mathbf{e} = (e_1, \ldots, e_m)$ such that for constant sparsity parameter $\gamma \in (0, 1)$

$$\mathbf{e}_i \stackrel{\$}{\leftarrow} \begin{cases} \mathbb{F}_p & \text{with probability } n^{-\gamma} \\ 0 & \text{otherwise} \end{cases}$$

To solve for $\mathbf{s}$, we'll construct a quadratic system of equations.

Our approach: guess whether an equation has error

There's no bound on the error size, so instead we'll guess whether an equation has error:



Define

$$\alpha_i = \begin{cases} 1 & \text{if } e_i = 0 \\ 0 & \text{if } e_i \neq 0 \end{cases}$$

Gives equations:

$$\{ \alpha_i (a_i \cdot s) = \alpha_i b_i \}_{i=1}^{m}$$

Our system of equations for LPN over $\mathbb{F}_p$ (2/2)

- Variables:
  - $\mathbf{x} = (x_1, \ldots, x_n)$ for the secret.
  - $\alpha_1, \ldots, \alpha_m$ will be indicator variables for error-free equations so that $\alpha_i = 1$ if ith equation is error-free, 0 otherwise.
  - Number of initial variables is $N := n + m$.
- Equations:
  - Guess the number of error-ridden equations t where $t \in [m]$.

$$\mathcal{F} \triangleq \left\{ \alpha_i \mathbf{a}_i \cdot \mathbf{x} = \alpha_i b_i \right\}_{i \in [m]} \cup \left\{ \alpha_i(\alpha_i - 1) = 0 \right\}_{i \in [m]} \cup \left\{ t = m - \sum_{i \in [m]} \alpha_i \right\}$$

  - Number of initial equations is $2m + 1$.

Initial hopes for a subexponential attack

Initially, $N = n + m$ variables and $2m + 1$ equations.

After d-degree Macaulay expansion,

- The number of variables is at most the number of monomials of degree at most $d + 2$: $V_d = \binom{N+d+2}{d+2}$
- The number of equations is $E_d = (2m+1)\binom{N+d}{d}$.
- $E_d \geq V_d$ when $d = \Omega\left(\sqrt{m}\right)$.

Assuming full rank of the expanded system, we see that Gaussian elimination on $O\left(\sqrt{m}\right)$-degree expanded system takes time $O\left(\binom{n+m+\sqrt{m}}{\sqrt{m}}^{\omega}\right) = e^{O(\sqrt{m}\ln m)}$.

The issue of rank of the expanded Macaulay matrix

Is the full rank assumption with an $O(\sqrt{m})$ expansion justified?

What degree of Macaulay expansion do we actually need so that the linearized expanded system of polynomials has full rank?

“Semi-regularity” implies exponential time attack

Main Problem: If our initial polynomial system is “semi-regular”, then $O(m)$-degree expansion is necessary (the runtime therefore is exponential).

What is a semi-regular polynomial system?

For our purposes:

- Semi-regular polynomial systems are sequences for which we can estimate a runtime upper bound for computing the Gröbner basis. (i.e. via a characterization for the Hilbert polynomial w.r.t grevlex order)
- Random overdetermined ($m > n$) polynomial systems are conjectured to be semi-regular (related to Fröberg's conjecture (1985), an open algebraic-geometric question).

Estimating runtime with the degree of regularity

Assuming a polynomial system is semi-regular, characterizing the attack complexity reduces to computing the degree of semi-regularity.

Lemma ([BFS15; Alb+15])

Let $f_1, \ldots, f_m \in \mathbb{F}_p[x_1, \ldots, x_n]$ where $m > n$. If $(f_1, \ldots, f_m)$ semi-regular, then the number of field operation required to compute a Gröbner basis of the ideal $\langle f_1, \ldots, f_m \rangle$ for any graded monomial ordering is bounded by

$$O\left(m \cdot d_{reg} \binom{n + d_{reg} - 1}{d_{reg}}^{\omega}\right), \text{ as } d_{reg} \to \infty$$

where ω is the linear algebra constant and d_{reg} is the degree of regularity of $\langle f_1, \ldots, f_m \rangle$.

Semi-regularity for homogeneous polynomials

Definition ([Alb+15])

Let $m \geq n$, let $(f_1, \ldots, f_m) \in \mathbb{F}_p[x_1, \ldots, x_n]$ be homogeneous polynomials of degree $d_1, \ldots, d_m$ resp. and let $\mathcal{I}$ be the ideal generated by these polynomials. The system is said to be a semi-regular sequence if the Hilbert polynomial associated to $\mathcal{I}$ w.r.t. to the grevlex order is

$$H(z) = \left[\frac{\prod_{i=1}^{m}(1 - z^{d_i})}{(1 - z)^n}\right]_+$$

where $[S]_+$ is the polynomial obtained by truncating the series S before the index of its non-positive coefficient.

The degree of regularity of a semi-regular sequence is $1 + \deg(H(z))$.

Homogenization of arbitrary polynomials

Definition ([Alb+15])

Let $f_1, \ldots, f_m \in \mathbb{F}_p[x_1, \ldots, x_n]$ be arbitrary (possibly inhomogeneous) polynomials. Let $f_1^h, \ldots, f_m^h$ be their respective homogeneous components of highest degree. A sequence $(f_1, \ldots, f_m)$ is semi-regular if the sequence $(f_1^h, \ldots, f_m^h)$ is semi-regular.

- e.g. if $f = 1 + x_1 + x_1x_2 + x_1^2$, then $f^h = x_1x_2 + x_1^2$.

Assuming semi-regularity in our setting

- First, a simplification:

$$\alpha_1 = m - t - \sum_{i \neq 1} \alpha_i$$

eliminate the variable α_1 by substitution to obtain $E = 2m$ equations (all are degree 2), $V = n + m - 1$ variables.

- After the simplification, our Hilbert series assuming semi-regularity is

$$H_{E,V}(z) = \frac{(1 - z^2)^E}{(1 - z)^{V+1}} = \sum_{d=0}^{\infty} h_d z^d$$

- Degree of regularity, d_{reg} is the first d such that h_d is non-positive.

Computing the degree of regularity in our setting

Sun et al. [STA20] perform the same computation for a different polynomial system for Binary LWE:

- Saddle point approximation to estimate the behavior of the coefficients of the Hilbert series:

$$d_{reg} + 1 = E - \frac{V+1}{2} - \sqrt{E(E-V)}$$

Theorem

Consider an $\mathrm{LPN}(n, m, \gamma)$ *instance with* $m = n^{1+\alpha}$. *Assuming semi-regularity, the degree of regularity of our system* $\mathcal{F}$ *behaves asymptotically as*

$$d_{reg} \approx 0.09n^{1+\alpha} + 0.2n + 0.18n^{1-\alpha} + o(n^{-2\alpha}) = O(m)$$

Can we directly increase the rank of the Macaulay matrix?

Observation: α_i variables are indicators for *sparse* errors. The product $\alpha_{i_1} \cdots \alpha_{i_d} = 0$ with high probability for large d.

- How many of these equations can we add? Subexponentially many.

Theorem

Consider an $\mathrm{LPN}(n, m, \gamma)$ *instance with* $m = n^{1+\alpha}$*. We assume that the number of instances with errors is* $t = \frac{m}{n^\gamma}$*. Pick* $\delta \in (0, 1)$ *sufficiently small and* $d \in \mathbb{Z}^+$ *such that* $d = \lceil n^{\gamma+\gamma'} \rceil$ *where* $\gamma' < 1 + \alpha$*. Then we can introduce up to* $k = \left\lfloor -\ln(1-\delta) 2^{n^{\gamma'}} \right\rfloor$ *equations of the form* $\alpha_{i_1} \cdots \alpha_{i_d} = 0$ *where the* i_j *are distinct for each equation, and all* k *equations hold with probability* $1 - \delta$*.*

- We don't know how these equations affect the rank of the Macaulay matrix.

Difficulty of Estimating Rank

- Estimating the rank of even the standard Macaulay matrix is quite challenging. Semi-regular assumptions only provide a rough heuristic.
- Introducing high degree equations might boost the rank, but is now even harder to analyze.
- Experiments are difficult to run due to sub-exponential blow-up in the size of the Macaulay matrix.

Recap and reflections

- **Recap**: we formulate a quadratic system of equations for LPN. Falsely assuming the Macaulay matrix is full rank suggests $O(\sqrt{m})$-expansion on this system is sufficient. Assuming semi-regularity suggests an upper bound of $O(m)$-expansion is required.
- Question: Is there some clever way to increase the rank of a Macaulay matrix at lower degrees of expansion?
- Question: We proposed adding random high degree equations that hold with high probability, but how does one analyze the rank of the matrix?
- Question: Is there a better system of equations for LPN over $\mathbb{F}_p$?

References I

[AAB15] Benny Applebaum, Jonathan Avron, and Christina Brzuska. "Arithmetic Cryptography: Extended Abstract". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*. Ed. by Tim Roughgarden. ACM, 2015, pp. 143–151. DOI: 10.1145/2688073.2688114. URL: https://doi.org/10.1145/2688073.2688114.

[AG11] Sanjeev Arora and Rong Ge. "New Algorithms for Learning in Presence of Errors". In: *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Zurich, Switzerland, July 4-8, 2011, Proceedings, Part I*. Vol. 6755. Lecture Notes in Computer Science. Springer, 2011, pp. 403–415.

[Alb+15] Martin R. Albrecht et al. "Algebraic algorithms for LWE problems". In: *ACM Commun. Comput. Algebra* 49.2 (2015), p. 62. DOI: 10.1145/2815111.2815158. URL: https://doi.org/10.1145/2815111.2815158.

[Ale03] Michael Alekhnovich. "More on average case vs approximation complexity". In: *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* IEEE. 2003, pp. 298–307.

[BFS15] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. "On the complexity of the F5 Gröbner basis algorithm". In: *J. Symb. Comput.* 70 (2015), pp. 49–70. DOI: 10.1016/j.jsc.2014.09.025. URL: https://doi.org/10.1016/j.jsc.2014.09.025.

References II

[BKW] Avrim Blum, Adam Kalai, and Hal Wasserman. "Noise-tolerant learning, the parity problem, and the statistical query model". In: *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*. ACM, pp. 435–440.

[Boy+18] Elette Boyle et al. "Compressing vector OLE". In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. 2018, pp. 896–912.

[Car+09] José Carrijo et al. "A novel probabilistic passive attack on the protocols HB and HB+". In: *IEICE transactions on fundamentals of electronics, communications and computer sciences* 92.2 (2009), pp. 658–662.

[CS16] Rodolfo Canto Torres and Nicolas Sendrier. "Analysis of Information Set Decoding for a Sub-linear Error Weight". In: *Post-Quantum Cryptography*. Ed. by Tsuyoshi Takagi. Cham: Springer International Publishing, 2016, pp. 144–161.

[DP12] Ivan Damgård and Sunoo Park. *How Practical is Public-Key Encryption Based on LPN and Ring-LPN?* Cryptology ePrint Archive, Paper 2012/699. https://eprint.iacr.org/2012/699. 2012. URL: https://eprint.iacr.org/2012/699.

[EKM17] Andre Esser, Robert Kübler, and Alexander May. "LPN decoded". In: *Annual International Cryptology Conference*. Springer. 2017, pp. 486–514.

References III

[IPS09] Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. "Secure Arithmetic Computation with No Honest Majority". In: *Theory of Cryptography, 6th Theory of Cryptography Conference, TCC 2009, San Francisco, CA, USA, March 15-17, 2009. Proceedings*. Ed. by Omer Reingold. Vol. 5444. Lecture Notes in Computer Science. Springer, 2009, pp. 294–314. DOI: 10.1007/978-3-642-00457-5_18. URL: https://doi.org/10.1007/978-3-642-00457-5%5C_18.

[JLS21] Aayush Jain, Huijia Lin, and Amit Sahai. "Indistinguishability obfuscation from well-founded assumptions". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 60–73.

[KS99] Aviad Kipnis and Adi Shamir. "Cryptanalysis of the HFE Public Key Cryptosystem by Relinearization". In: *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*. Vol. 1666. Lecture Notes in Computer Science. Springer, 1999, pp. 19–30.

[Laz83] D. Lazard. "Gröbner bases, Gaussian elimination and resolution of systems of algebraic equations". In: *Computer Algebra*. Ed. by J. A. van Hulzen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1983, pp. 146–156. ISBN: 978-3-540-38756-5.

References IV

[Lyu05] Vadim Lyubashevsky. "The Parity Problem in the Presence of Noise, Decoding Random Linear Codes, and the Subset Sum Problem". In: *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th InternationalWorkshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*. Ed. by Chandra Chekuri et al. Vol. 3624. Lecture Notes in Computer Science. Springer, 2005, pp. 378–389. DOI: 10.1007/11538462_32. URL: https://doi.org/10.1007/11538462%5C_32.

[Pra62] Eugene Prange. "The use of information sets in decoding cyclic codes". In: *IRE Transactions on Information Theory* 8.5 (1962), pp. 5–9.

[STA20] Chao Sun, Mehdi Tibouchi, and Masayuki Abe. "Revisiting the hardness of binary error LWE". In: *Australasian Conference on Information Security and Privacy*. Springer. 2020, pp. 425–444.