Attacks Against the $\mathsf{IND\text{-}CPA^D}$ Security of Exact FHE Schemes.

Jung Hee Cheon[1,2], **Hyeongmin Choe**[1], Damien Stehlé[2],
Alain Passelègue[2], Elias Suvanto[2,3]

[1]**Seoul National University**, [2]CryptoLab Inc., [3]University of Luxembourg

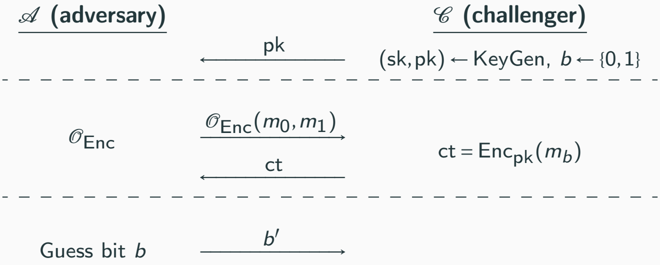
2024 KMS Spring Meeting
April, 19th, 2024.

IND-CPA^D **Security of FHEs:**

Security Definitions

IND-CPA **Security**

General security notion for (F)HE is IND-CPA security:

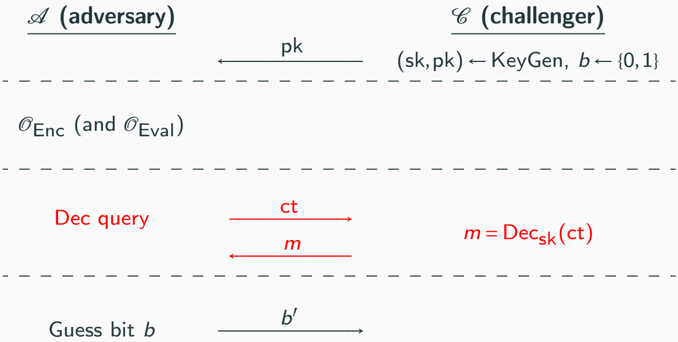


$\mathscr{A}$ **(adversary)**		$\mathscr{C}$ **(challenger)**
	$\xleftarrow{\mathsf{pk}}$	$(\mathsf{sk}, \mathsf{pk}) \leftarrow \mathsf{KeyGen},\ b \leftarrow \{0,1\}$
$\mathscr{O}_{\mathsf{Enc}}$	$\xrightarrow{\mathscr{O}_{\mathsf{Enc}}(m_0, m_1)}$ $\xleftarrow{\mathsf{ct}}$	$\mathsf{ct} = \mathsf{Enc}_{\mathsf{pk}}(m_b)$
Guess bit b	$\xrightarrow{b'}$	

Figure 1: IND-CPA security game.

IND-CCA **Security**

FHE cannot achieve IND-CCA2 security, with Dec oracle:



$\mathscr{A}$ **(adversary)**

$\mathscr{C}$ **(challenger)**

$\xleftarrow{\mathsf{pk}}$ $(\mathsf{sk},\mathsf{pk}) \leftarrow \mathsf{KeyGen},\ b \leftarrow \{0,1\}$

$\mathscr{O}_{\mathsf{Enc}}$ (and $\mathscr{O}_{\mathsf{Eval}}$)

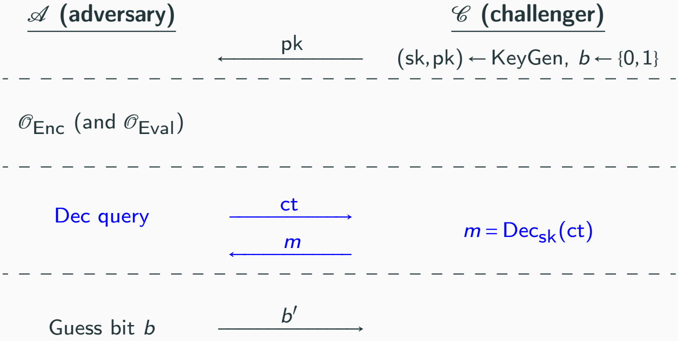
Dec query $\xrightarrow{\mathsf{ct}}$ $\xleftarrow{m}$ $m = \mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct})$

Guess bit b $\xrightarrow{b'}$

Figure 2: IND-CCA2 security game.

IND-CPA^D **Security?**

IND-CPA^D [LM21]: IND-CPA + Dec oracle, but only to legitimate ct's.



$\mathscr{A}$ **(adversary)**

$\mathscr{C}$ **(challenger)**

$\xleftarrow{\text{pk}}$ $(\text{sk}, \text{pk}) \leftarrow \text{KeyGen},\ b \leftarrow \{0,1\}$

$\mathcal{O}_{\text{Enc}}$ (and $\mathcal{O}_{\text{Eval}}$)

Dec query $\xrightarrow{\text{ct}}$ $\xleftarrow{m}$ $m = \text{Dec}_{\text{sk}}(\text{ct})$

Guess bit b $\xrightarrow{b'}$

Figure 3: IND-CCA2 security game.

To check the legitimacy, we additionally need:

$$\text{Shared state: } S \in (\mathscr{M} \times \mathscr{M} \times \mathscr{C})^*$$

IND-CPA$^{\mathrm{D}}$ Security

$\mathscr{A}$ (adversary)		**$\mathscr{C}$ (challenger)**
	$\xleftarrow{\mathsf{pk}}$	$(\mathsf{sk},\mathsf{pk}) \leftarrow \mathsf{KeyGen},\ b \leftarrow \{0,1\}$
$\mathcal{O}_{\mathsf{Enc}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Enc}}(m_0,m_1)}$ $\xleftarrow{\mathsf{ct}}$	$S.\mathsf{append}(\mathsf{m}_0,\mathsf{m}_1,\mathsf{ct}=\mathsf{Enc}_{\mathsf{pk}}(\mathsf{m}_{\mathsf{b}}))$
$\mathcal{O}_{\mathsf{Eval}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Eval}}(C,i_1,\cdots,i_k)}$ $\xleftarrow{\mathsf{ct}^*}$	$S.\mathsf{append}(\mathsf{m}_0^*,\mathsf{m}_1^*,\mathsf{ct}^*)$, where $m_i^* = C(S[i_1].m_i,\ldots,S[i_k].m_i)$ for $i=0,1$, $\mathsf{ct}^* \leftarrow \mathsf{Eval}_{\mathsf{pk}}(C,S[i_1].\mathsf{ct},\cdots,S[i_k].\mathsf{ct})$
$\mathcal{O}_{\mathsf{Dec}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Dec}}(j)}$ $\xleftarrow{m}$	if $S[j].m_0 = S[j].m_1$, $m \leftarrow \mathsf{Dec}_{\mathsf{sk}}(S[j].\mathsf{ct})$
Guess bit b	$\xrightarrow{b'}$	

Figure 4: IND-CPA$^{\mathrm{D}}$ security game [LM21].

KR^{D} Security

KR^{D} security can be similarly defined, with $S \in (\mathcal{M} \times \mathcal{C})^*$:

$\mathcal{A}$ (adversary)		**$\mathcal{C}$ (challenger)**
	$\xleftarrow{\quad \mathsf{pk} \quad}$	$(\mathsf{sk},\mathsf{pk}) \leftarrow \mathsf{KeyGen}$
$\mathcal{O}_{\mathsf{Enc}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Enc}}(m)}$ $\xleftarrow{\quad \mathsf{ct} \quad}$	$S.\mathsf{append}(\mathsf{m},\mathsf{ct} = \mathsf{Enc}_{\mathsf{pk}}(\mathsf{m}))$
$\mathcal{O}_{\mathsf{Eval}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Eval}}(C, i_1, \cdots, i_k)}$ $\xleftarrow{\quad \mathsf{ct}^* \quad}$	$S.\mathsf{append}(\mathsf{m}^*, \mathsf{ct}^*)$, where $m^* = C(S[i_1].m, \ldots, S[i_k].m)$, $\mathsf{ct}^* \leftarrow \mathsf{Eval}_{\mathsf{pk}}(C, S[i_1].\mathsf{ct}, \cdots, S[i_k].\mathsf{ct})$
$\mathcal{O}_{\mathsf{Dec}}$	$\xrightarrow{\mathcal{O}_{\mathsf{Dec}}(j)}$ $\xleftarrow{\quad m \quad}$	$m \leftarrow \mathsf{Dec}_{\mathsf{sk}}(S[j].\mathsf{ct})$
Guess sk	$\xrightarrow{\quad \mathsf{sk}' \quad}$	

Figure 5: KR^{D} security game [LM21].

IND-CPA$^{\text{D}}$ and KR$^{\text{D}}$ Security

To summarize,

- IND-CPA$^{\text{D}}$ security
  - Track the messages (correspond to $b = 0, 1$) in the ciphertexts during HE operations.
  - Decrypt only the ciphertexts that are tracked & the two tracked messages (correspond to $b = 0, 1$) are the same.
- KR$^{\text{D}}$ security
  - Track the ciphertexts during HE operations.
  - Decrypt only the ciphertexts that are tracked.

Prior Works

Easy check: for (F)HEs,

- IND-CPA$^{\mathsf{D}}$ security ⇒ IND-CPA security, KR$^{\mathsf{D}}$ security.
- KR$^{\mathsf{D}}$ security ⇒ KR security

Li and Micciancio [LM21]:

- For exact FHEs, IND-CPA security ⇒ IND-CPA$^{\mathsf{D}}$ security.
- This is not the case for approximate FHEs: CKKS KR$^{\mathsf{D}}$ attack.

Li, Micciancio, Schultz, and Sorrel [LMSS22]:

- IND-CPA$^{\mathsf{D}}$-secure CKKS with noise flooding & DP.

Guo, Nabokov, Suvanto, and Johansson [GNSJ24]:

- KR$^{\mathsf{D}}$ attack against [LMSS22].

Prior Works

In the community:

"Exact" FHEs, such as BFV/BGV,
and DM/CGGI schemes, are IND-CPA^D secure!

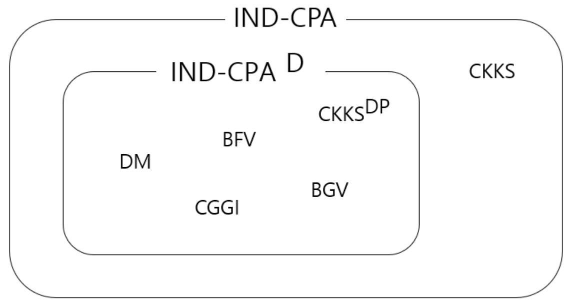


Figure 6: The belief.

*CKKS^{DP} denotes a correction of [LMSS22] against [GNSJ24].

Really?

Theoretical Results 1:

IND-CPA^{D} Security of Exact (F)HEs

What if?

What if $\mathsf{Dec}_{\mathsf{sk}}(\mathsf{Eval}_{\mathsf{pk}}(C, \mathsf{ct}_1, \cdots, \mathsf{ct}_k))$
$\neq C(\mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_1), \ldots, \mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_k))$, a.k.a. decryption fails?

$\Rightarrow \mathcal{O}_{\mathsf{Enc}}/\mathcal{O}_{\mathsf{Eval}}$ will record $(m_0, m_1, \mathsf{ct}_{\mathsf{result}})$, with

$$\mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_{\mathsf{result}}) \neq m_b.$$

Can we make $m_0 = m_1$, but $\mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_{\mathsf{result}}) \neq m_0$?

Can we make $\mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_{\mathsf{result}})$ to depend on b?

Yes,

with the failing probability!

Generic IND-CPA$^{\mathsf{D}}$ Attack

$\mathscr{A}$ **(adversary)**	$\mathscr{C}$ **(challenger)**
m_0, m_1	$b \leftarrow \{0,1\}$
$\xrightarrow{\mathscr{O}_{\mathsf{Enc}}(\mathtt{T},\mathtt{F})}$	$S[0] = (\mathtt{T},\mathtt{F},\mathsf{ct}_0)$
$\xrightarrow{\mathscr{O}_{\mathsf{Enc}}(\mathtt{F},\mathtt{F})}$	$S[1] = (\mathtt{F},\mathtt{F},\mathsf{ct}_1)$
$\xrightarrow{\mathscr{O}_{\mathsf{Eval}}(\mathsf{AND}(\cdot,\cdot),0,1)}$	$S[2] = (\mathtt{T} \wedge \mathtt{F} = \mathtt{F}, \mathtt{F} \wedge \mathtt{F} = \mathtt{F}, \mathsf{ct}_2)$
$\xrightarrow{\mathscr{O}_{\mathsf{Dec}}(2)}$	$m_{\mathsf{res}} \leftarrow \mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_2)$
$\xleftarrow{m_{\mathsf{res}}}$	

$$b' = \begin{cases} 0 & \text{if } m_{\mathsf{res}} = \mathtt{T} \quad , \\ 1 & \text{else.} \end{cases}$$

Figure 7: Generic and passive IND-CPA$^{\mathsf{D}}$ attack on binary FHE.

Generic IND-CPA$^{\mathsf{D}}$ Attack

$$\begin{array}{ccc}
\underline{\mathscr{A}\ \textbf{(adversary)}} & & \underline{\mathscr{C}\ \textbf{(challenger)}} \\
 & m_0, m_1 & b \leftarrow \{0,1\} \\
 & \xrightarrow{\mathscr{O}_{\mathsf{Enc}}(\mathtt{T},\mathtt{F})} & S[0] = (\mathtt{T},\mathtt{F},\mathsf{ct}_0) \\
 & \xrightarrow{\mathscr{O}_{\mathsf{Enc}}(\mathtt{F},\mathtt{F})} & S[1] \overset{?}{=} (\mathtt{T},\mathtt{T},\mathsf{ct}_1) \\
 & \xrightarrow{\mathscr{O}_{\mathsf{Eval}}(\mathsf{AND}(\cdot,\cdot),0,1)} & S[2] \overset{?}{=} (\mathtt{T} \land \mathtt{T} = \mathtt{T}, \mathtt{F} \land \mathtt{T} = \mathtt{F}, \mathsf{ct}_2) \\
 & \xrightarrow{\mathscr{O}_{\mathsf{Dec}}(2)} & m_{\mathsf{res}} \leftarrow \mathsf{Dec}_{\mathsf{sk}}(\mathsf{ct}_2) \\
 & \xleftarrow{m_{\mathsf{res}}} &
\end{array}$$

$$b' = \begin{cases} 0 & \text{if } m_{\mathsf{res}} = \mathtt{T} \quad , \\ 1 & \text{else.} \end{cases}$$

Figure 8: Generic and passive IND-CPA$^{\mathsf{D}}$ attack on binary FHE.

Generic IND-CPA$^{\text{D}}$ Attack

- Advantage (success probability) of $\mathscr{A}$: $\Pr[b = b'] \approx \frac{1}{2} + \frac{p_{\text{fail}}}{2}$.
- Can be boosted by compositions, e.g.,

$$C(x_0, \ldots, x_N) = (x_0 \wedge x_1) \vee \cdots \vee (x_0 \wedge x_N).$$

- Can be extended to general integer(F)HEs.

Thus, the failing probability of any λ–IND-CPA$^{\text{D}}$-secure (F)HE ciphertext should be $\lesssim 2^{-\lambda}$ along the whole homomorphic operations.

Generic IND-CPA^D **Attack**

HOWEVER,

- **DM/CGGI**
  - TFHE-rs: $p_{\text{fail}} = 2^{-40}$ (DEFAULT_PARAMETERS set)
  - Concrete-Python: $p_{\text{fail}} = 2^{-17}$ (default setting)
  - Dahl et al. [DDK+23]: $p_{\text{fail}} = 2^{-13.9}$
- **BFV/BGV**: Recent average-case approaches try to lower the correctness for more multiplicative levels:
  - Murphy and Player [MP19]: $p_{\text{fail}} = 0.001 \approx 2^{-10}$
  - Biasioli et al. [BMCM23]: $p_{\text{fail}} = 2^{-36} \sim 2^{-80}$

IND-CPA^D-secure? :(

Theoretical Results 2:

IND-CPA$^{\text{D}}$ **Security of Threshold (F)HE**

Threshold (F)HE

In a multi-party setting, Threshold (F)HE [JRS17] allows decrypting a ciphertext in a distributed manner via partial decryption.

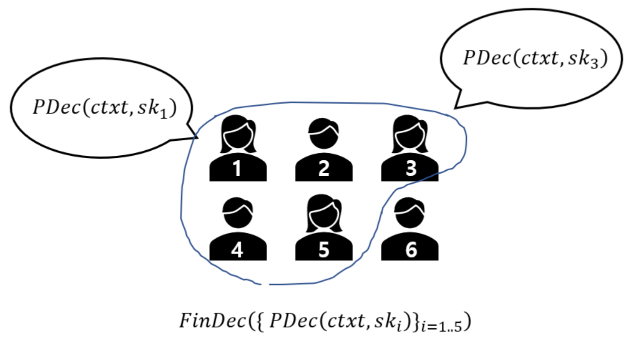


Figure 9: Threshold FHE.

Basically, each party has a Dec oracle for legitimate ciphertexts!

Threshold Security Reduction

Concretely, for

- Threshold FHE scheme Π (Setup, Enc, Eval, PDec, FinDec),
- FHE scheme Π^* (Setup, Enc, Eval, Dec),

where $\Pi^*.\mathsf{Dec} = \Pi.\mathsf{FinDec}_{pk}\left(\{\Pi.\mathsf{PDec}_{sk_i}(\cdot)\}_i\right)$,

Threshold IND-**security of** $\Pi \Rightarrow$ IND-CPA$^{\mathsf{D}}$ **security of** Π^*.

Threshold KR-**security of** $\Pi \Rightarrow$ KR$^{\mathsf{D}}$ **security of** Π^*.

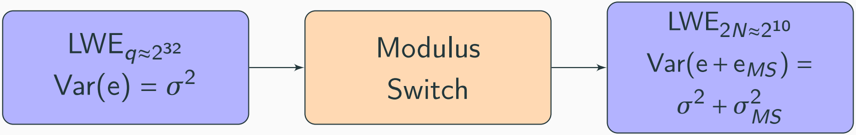
Practically,

- Π^* for Noah's Ark [DDK$^+$23], a CGGI-based Threshold FHE scheme, has $\mathsf{p_{fail}} = 2^{-40}$:(

Practical Results 1:

KR^{D} attack on CGGI

Recap: CGGI Modulus Switch



$\mathrm{LWE}_{q\approx 2^{32}}$
$\mathrm{Var(e)} = \sigma^2$

Modulus Switch

$\mathrm{LWE}_{2N\approx 2^{10}}$
$\mathrm{Var}(\mathrm{e}+\mathrm{e}_{MS}) = \sigma^2+\sigma^2_{MS}$

$$\mathrm{ct} \mapsto \lfloor \mathrm{ct}\cdot 2N/q \rceil$$

Modulus Switching error: $\mathrm{e}_{MS} := \langle \vec{e}_{MS}, \mathrm{sk}\rangle$, $\mathrm{Var}(\vec{e}_{MS}) = \sigma^2_{MS} \gtrsim \sigma^2$.

Fails when $\mathrm{e} + \langle \vec{e}_{MS}, \mathrm{sk}\rangle > t$, for some threshold t, with probability

$$\mathrm{p_{fail}} = \mathrm{erfc}\left(t/\sqrt{2(\sigma^2+\sigma^2_{MS})}\right) \approx 2^{-40}.$$

KR^D attack on CGGI

Key idea[1]:

- When decryption fails, the inequality $\mathrm{e} + \langle \vec{e}_{MS}, \mathrm{sk} \rangle > t$ holds.
- sk is likely parallel to $\vec{e}_{MS}$.
- $\vec{e}_{MS}$ is public!

The distribution of $(\vec{e}_{MS} \mid \text{decryption fail})$ will reveal sk.

---

[1] the key idea is somewhat similar to IND-CCA attacks against KEMs based on the decryption failures.

KR^D attack on CGGI

For $Y_i = \langle \vec{e}_{MS}, \mathrm{sk}\rangle - e_{MS,i} \cdot \mathrm{sk}_i$, the pdf f of $e_{MS,i} \mid$ "$\langle \vec{e}_{MS}, \mathrm{sk}\rangle + e > t$" satisfies,

$$f(x) = \begin{cases} \frac{\Pr[x + Y_i + e > t]}{\Pr[\langle \vec{e}_{MS}, \mathrm{sk}\rangle + e > t]} & \text{if } \mathrm{sk}_i = 1, \\ 1 & \text{if } \mathrm{sk}_i = 0. \end{cases}$$

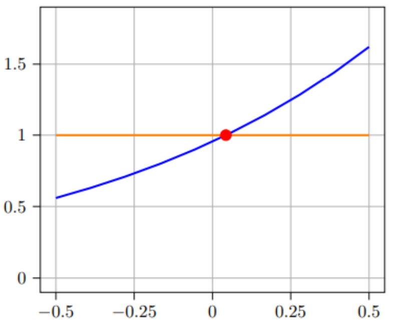


Figure 10: Distribution of $e_{MS,i}$ conditioned on decryption failures.

KR^D attack on CGGI

By collecting $\vec{e}_{MS}$ of the failing ciphertexts, we can estimate sk:

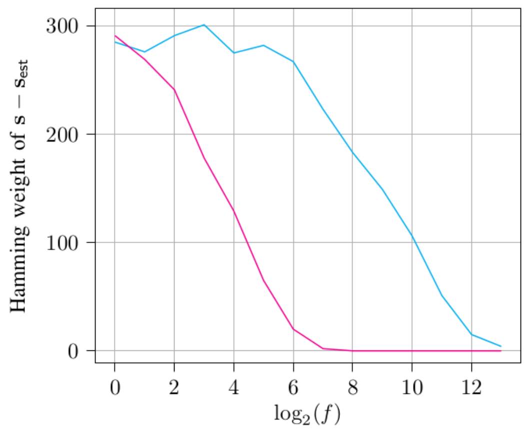


Figure 11: Accuracy of the attack ($\|\mathrm{sk} - \mathrm{sk}_{\mathrm{est}}\|_1$) with "$f$" failing ctxts, based on experimental results for a custom parameter set and simulated results for TFHE-rs DEFAULT_PARAMETERS set is given.

Practical Results 2:

KR^{D} attack on BFV

Recap: Average-case Error Analysis on BGV/BFV

RLWE ciphertext $\mathsf{ct} = (\mathsf{a}, \mathsf{b} = \mathsf{as} + \Delta\mathsf{m} + \mathsf{e})$ has $\mathsf{p}_{\mathsf{fail}} = \mathsf{erfc}\left(\frac{t}{\sqrt{2 \cdot \mathsf{Var}(e)}}\right)$.

For $\mathsf{ct} + \mathsf{ct}'$,

- Average-case analysis: $\mathsf{Var}(e + e') = 2\sigma^2$, assuming i.i.d.
- But in the worst case $(\mathsf{ct} = \mathsf{ct}')$, $\mathsf{Var}(2e) = 4\sigma^2$.

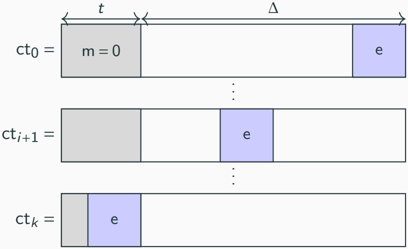
$\Rightarrow \mathsf{p}_{\mathsf{fail}}^{\mathsf{worst}} \gg \mathsf{p}_{\mathsf{fail}}^{\mathsf{avg}}$, which leads to IND-CPA$^{\mathsf{D}}$/KR$^{\mathsf{D}}$-insecurity.

KR^{D} Attack on BFV

Iterative addition attack[2]: After k iterative additions, error e blows up to $2^k e$, which will be decrypted to

$$\left\lfloor (2^k e \bmod q)/\Delta \right\rceil \approx e,$$

if $2^k \approx \Delta$.



Note, average-case approach will allow this, since $\mathrm{p}_{\mathrm{fail}}^{\mathrm{worst}} \gg \mathrm{p}_{\mathrm{fail}}^{\mathrm{avg}}$.

---

[2] similar to the attack by Guo et al. [GNSJ24], targeted Li and Micciancio's $\mathrm{CKKS}^{\mathrm{DP}}$, implemented in OpenFHE [LMSS22].

Summary

Summary

Theoretical results:

- Exact FHE schemes are also IND-CPA$^{\mathsf{D}}$-insecure unless it has failure probability $\lesssim 2^{-\lambda}$.[3]
- Threshold FHE schemes are IND/KR-insecure unless the underlying FHE schemes are IND-CPA$^{\mathsf{D}}$/KR$^{\mathsf{D}}$-secure.

Practical results: KR$^{\mathsf{D}}$ attacks on

- BFV[4] (not included),
- CGGI,
- CGGI-based Threshold FHE, Noah's Ark[5] (not included).

---

[3] during the whole homomorphic operations.

[4] similar to [GNSJ24], which also applies to BGV.

[5] may work even when the underlying FHE scheme is perfectly correct!

Summary

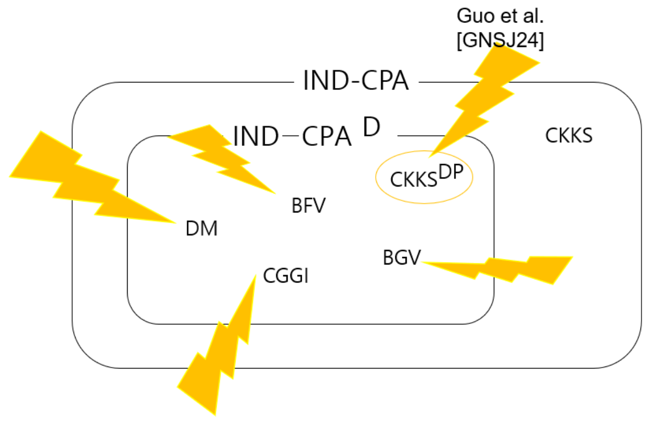


Figure 12: Our attack.

Thank You!

References i

[BMCM23] Beatrice Biasioli, Chiara Marcolla, Marco Calderini, and Johannes Mono.
Improving and automating bfv parameters selection: An average-case approach.
Cryptology ePrint Archive, Paper 2023/600, 2023.
`https://eprint.iacr.org/2023/600`.

[DDK+23] Morten Dahl, Daniel Demmler, Sarah El Kazdadi, Arthur Meyre, Jean-Baptiste Orfila, Dragos Rotaru, Nigel P. Smart, Samuel Tap, and Michael Walter.
Noah's ark: Efficient threshold-fhe using noise flooding.
Cryptology ePrint Archive, Paper 2023/815, 2023.
`https://eprint.iacr.org/2023/815`.

[GNSJ24] Qian Guo, Denis Nabokov, Elias Suvanto, and Thomas Johansson.
Key recovery attacks on approximate homomorphic encryption with Non-Worst-Case noise flooding countermeasures.
In *33rd USENIX Security Symposium (USENIX Security 24)*, Philadelphia, PA, August 2024. USENIX Association.

References ii

[JRS17] Aayush Jain, Peter M. R. Rasmussen, and Amit Sahai.
Threshold fully homomorphic encryption.
Cryptology ePrint Archive, Paper 2017/257, 2017.
`https://eprint.iacr.org/2017/257`.

[LM21] Baiyu Li and Daniele Micciancio.
On the security of homomorphic encryption on approximate numbers.
In Anne Canteaut and François-Xavier Standaert, editors,
EUROCRYPT 2021, Part I, volume 12696 of *LNCS*, pages 648–677.
Springer, Heidelberg, October 2021.

[LMSS22] Baiyu Li, Daniele Micciancio, Mark Schultz, and Jessica Sorrell.
Securing approximate homomorphic encryption using differential privacy.
In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part I*,
volume 13507 of *LNCS*, pages 560–589. Springer, Heidelberg, August 2022.

References iii

[MP19] Sean Murphy and Rachel Player.
A central limit framework for ring-LWE decryption.
Cryptology ePrint Archive, Report 2019/452, 2019.
https://eprint.iacr.org/2019/452.