技術を社会へ
Integration for Innovation

National Institute of Advanced Industrial Science and Technology
AIST
独立行政法人
産業技術総合研究所

筑波大学 University of Tsukuba

東京大学 THE UNIVERSITY OF TOKYO

理化学研究所 RIKEN

加法準同型暗号を用いたデータベースの秘匿検索プロトコル

***縫田 光司** (産総研), 清水 佳奈 (産総研), 荒井 ひろみ (理化学研究所), 浜田 道昭 (東京大学), 津田 宏治 (産総研), 広川 貴次 (産総研), 花岡 悟一郎 (産総研), 佐久間 淳 (筑波大学), 浅井 潔 (産総研)

2012/11/16　@統計数理研究所

著者一覧(再掲)

縫田 光司 (産業技術総合研究所セキュアシステム研究部門)
清水 佳奈 (産業技術総合研究所生命情報工学研究センター)
荒井 ひろみ (理化学研究所生命情報基盤研究部門)
浜田 道昭 (東京大学大学院新領域創成科学研究科)
津田 宏治 (産業技術総合研究所生命情報工学研究センター)
広川 貴次 (産業技術総合研究所生命情報工学研究センター)
花岡 悟一郎 (産業技術総合研究所セキュアシステム研究部門)
佐久間 淳 (筑波大学大学院システム情報工学研究科)
浅井 潔 (産業技術総合研究所生命情報工学研究センター)

背景

例：新薬の研究開発

- まず、研究者は薬として効果がありそうな化合物組成候補を（理論や実験で）見出す
- 次に、その候補と**似た組成の化合物**（の構成法や性質など）が既に知られているか、市販の**化合物データベースで検索**する
  - 既知の化合物を利用できれば、開発コストがぐっと削減できる！

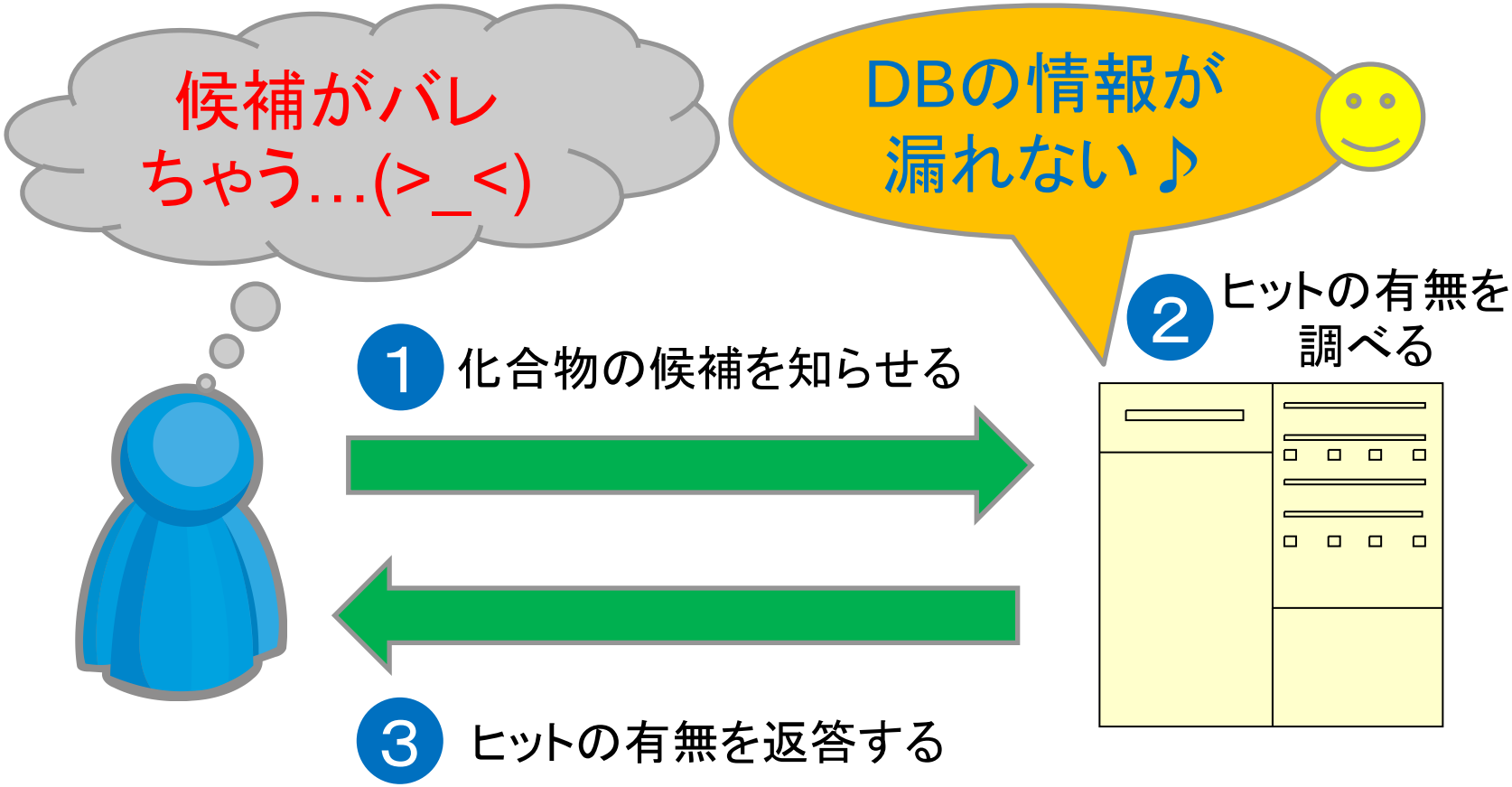
背景

ユーザ（研究者）側は、市販の化合物DB購入前に、自分の欲しいデータがDBに載っているかどうか予め確認したい

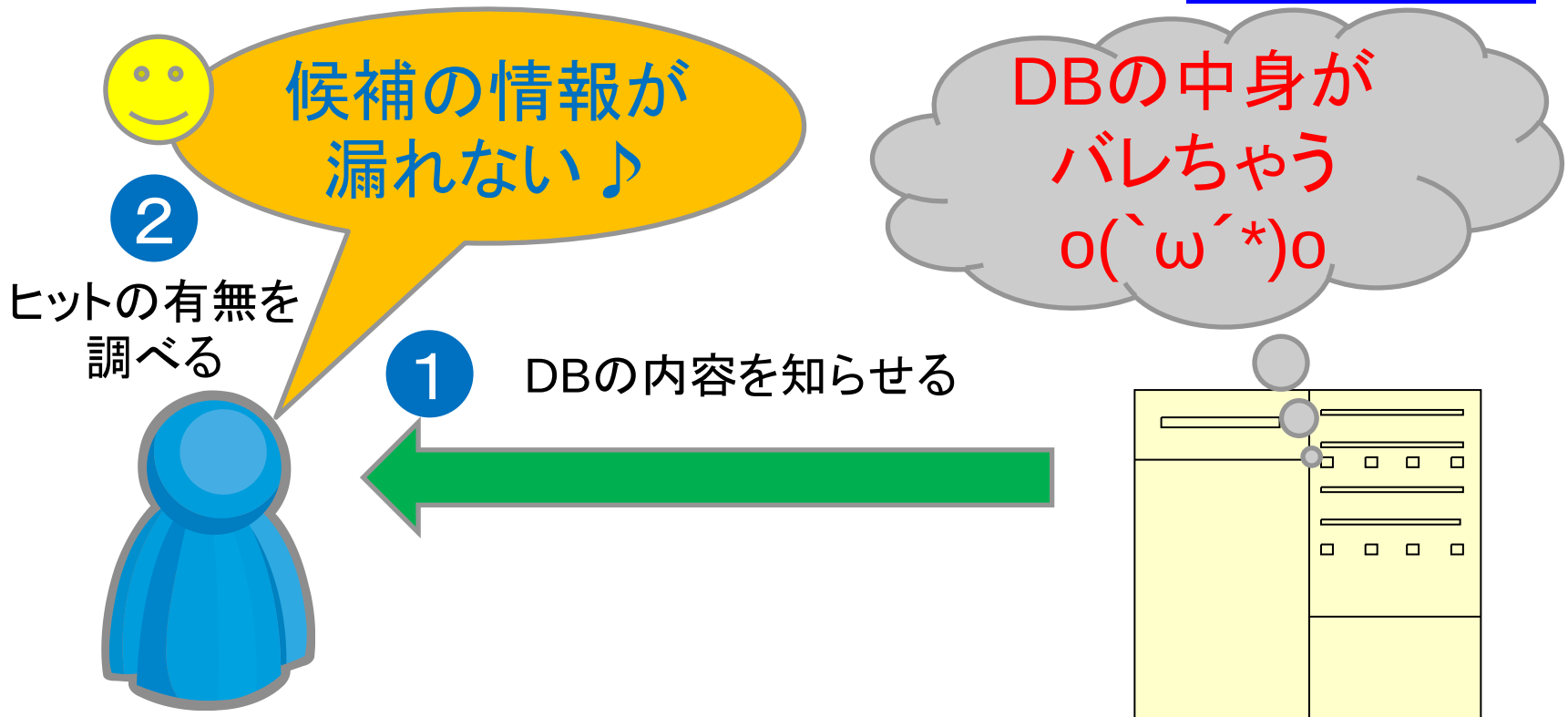
- ユーザ側の要望：**検索内容の秘匿**
  - 知財関連、個人情報・プライバシー、等
- サーバ（DB）側の要望：（検索対象物の有無以外の余分な）**DB内容の秘匿**
  - 膨大なDB構築コストの確実な回収

両立したい

ダメな方法その1


候補がバレちゃう…(>_<)
DBの情報が漏れない♪
1 化合物の候補を知らせる
2 ヒットの有無を調べる
3 ヒットの有無を返答する

ダメな方法その2

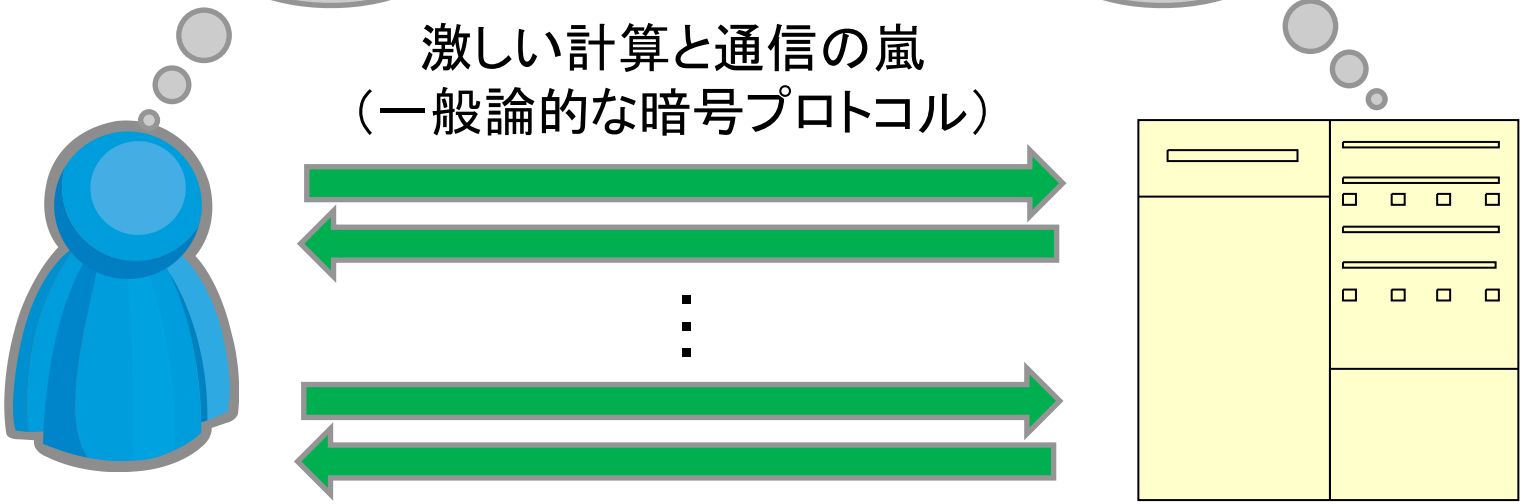


（＊「Private Information Retrieval」の原理）

ダメな方法その3

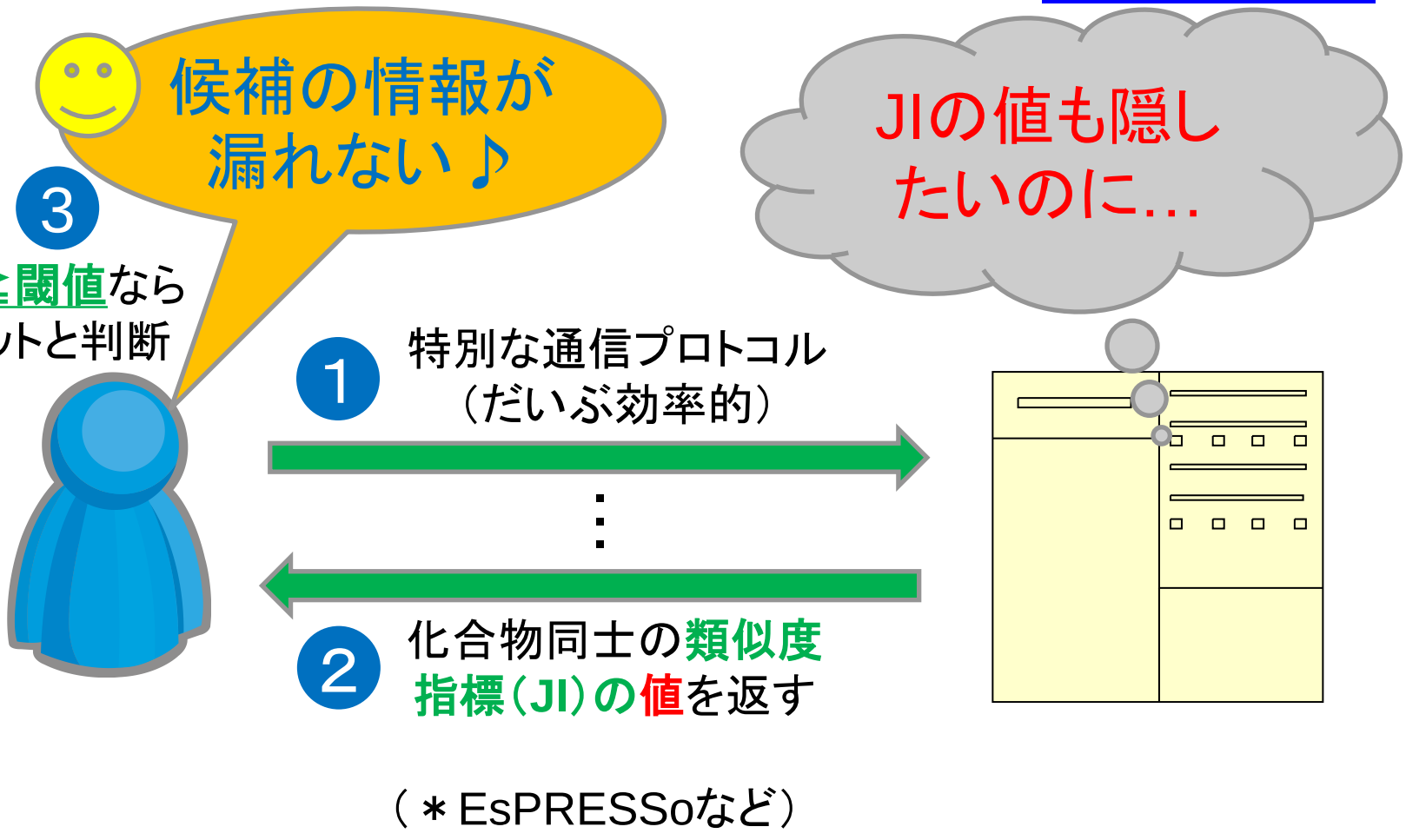
どんだけ待たせるの…orz

どんだけ計算させるの…orz



（＊「マルチパーティ計算」「完全準同型暗号」）

惜しい方法（既存方式）



（＊EsPRESSoなど）

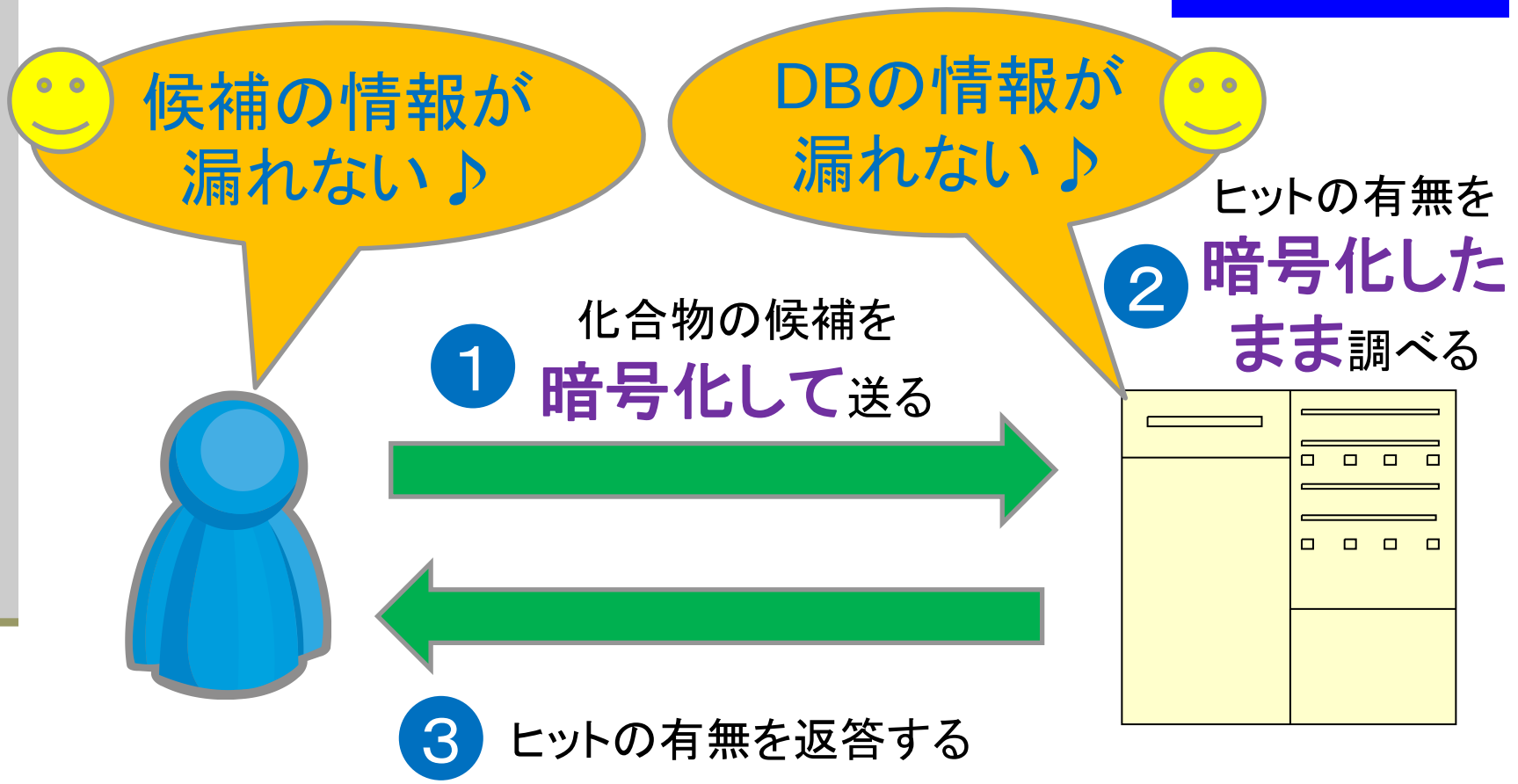
既存方式との比較

方式	出力	通信量	ラウンド数	計算量	TPP
提案法	JIの大小	小	1	小	無
Singhら	JI	小	2以上	小	必要
EsPRESSo	JI	小	1	小	無
Zhangら	JI	小	1	小	無
マルチパーティ計算（一般論）	JIの大小	大	2以上	大	無
完全準同型暗号	JIの大小	大	1	大	無

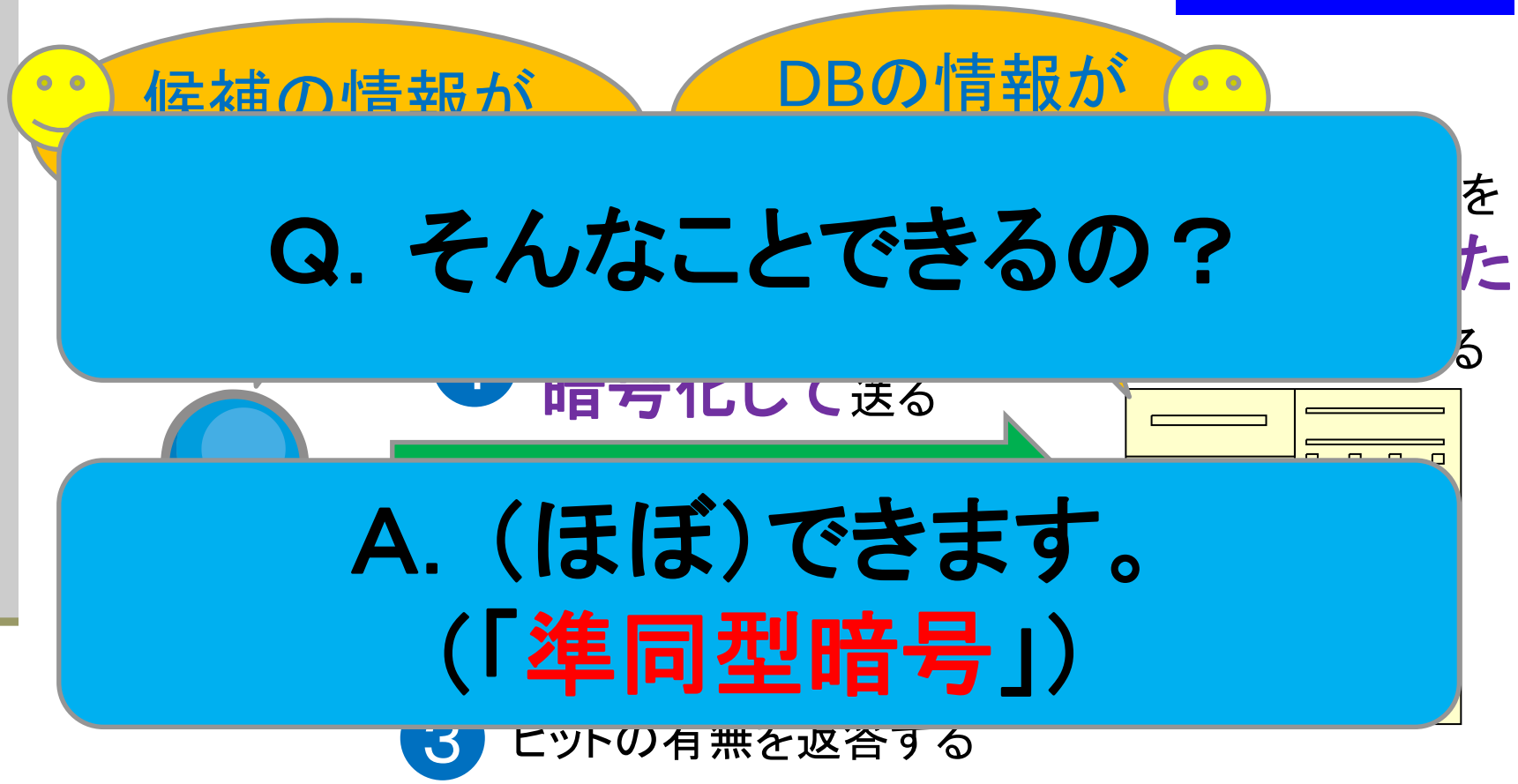
本提案手法は、（計算・通信コストも配慮しつつ）類似度指標JIの値まで隠せるような初の手法

技術的なポイント

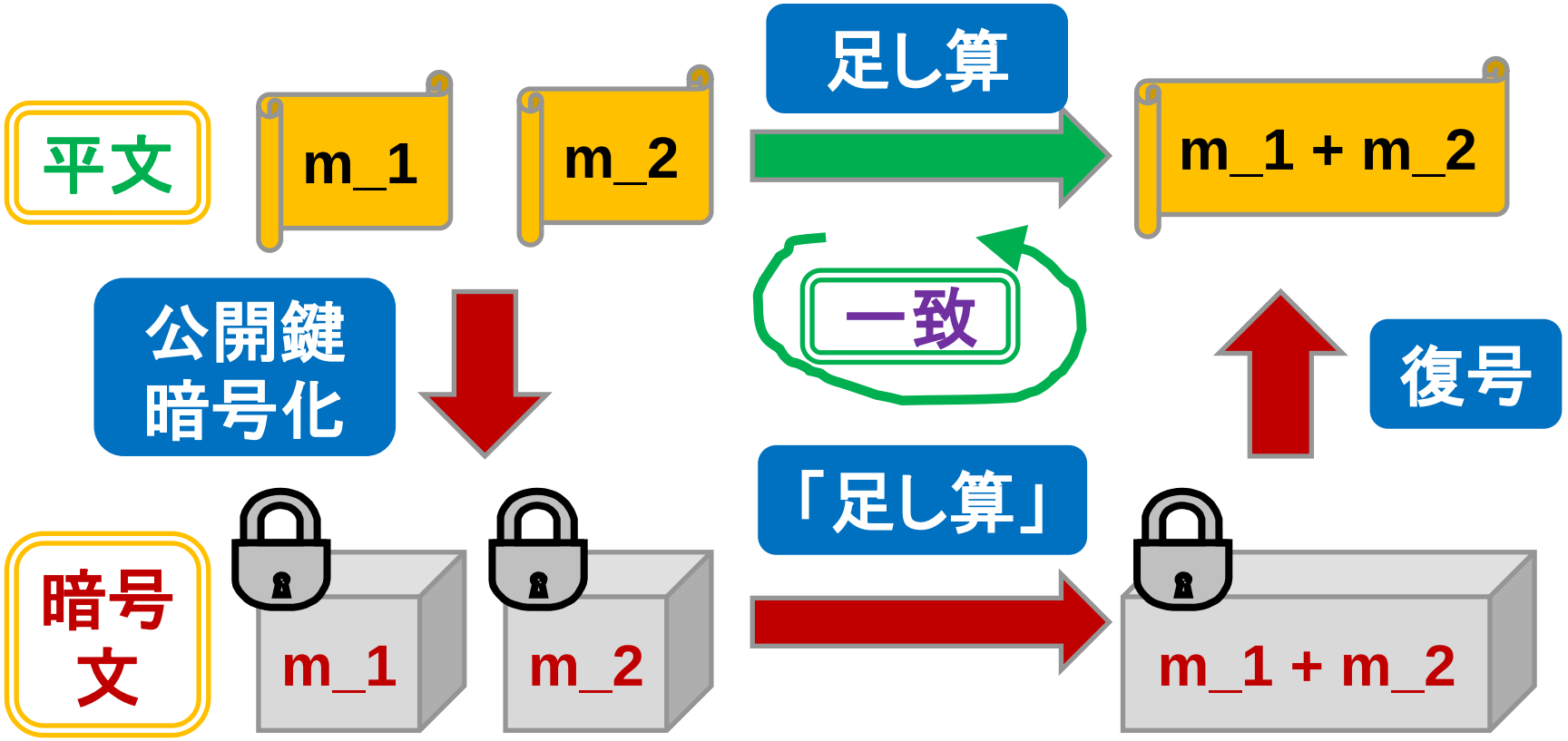
理想的な方法


候補の情報が漏れない♪
DBの情報が漏れない♪
ヒットの有無を
2 暗号化したまま調べる
化合物の候補を
1 暗号化して送る
3 ヒットの有無を返答する

理想的な方法


候補の情報が
DBの情報が
を
た
る
暗号化して送る
Q. そんなことできるの？
A.（ほぼ）できます。
（「準同型暗号」）
3
ヒットの有無を返答する

（加法）準同型暗号



- 例：Paillier暗号、(lifted) ElGamal暗号

類似度指標の計算式再考

条件「JI（の一般形*TI*）が閾値以上」を変形

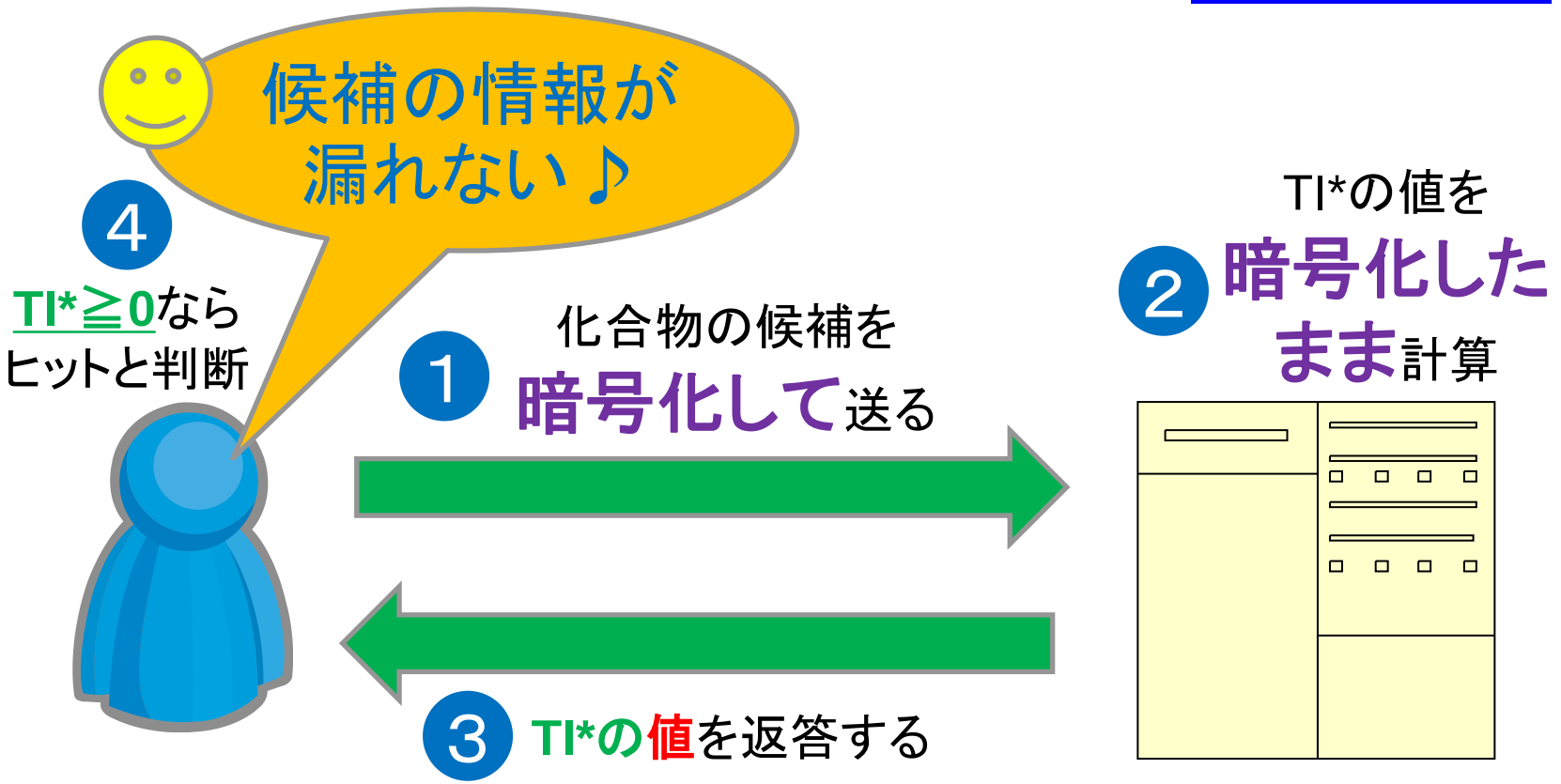
「フィンガープリント」
（特徴の有無を表すベクトル）

$$\overline{\mathrm{TI}}(\vec{p},\vec{q}) := \Gamma|\vec{p}\cap\vec{q}| - \theta_{\mathrm{n}}(\mu_{\mathrm{a}}|\vec{p}| - \mu_{\mathrm{b}}|\vec{q}|) \geq 0$$

提案法の基本アイデア

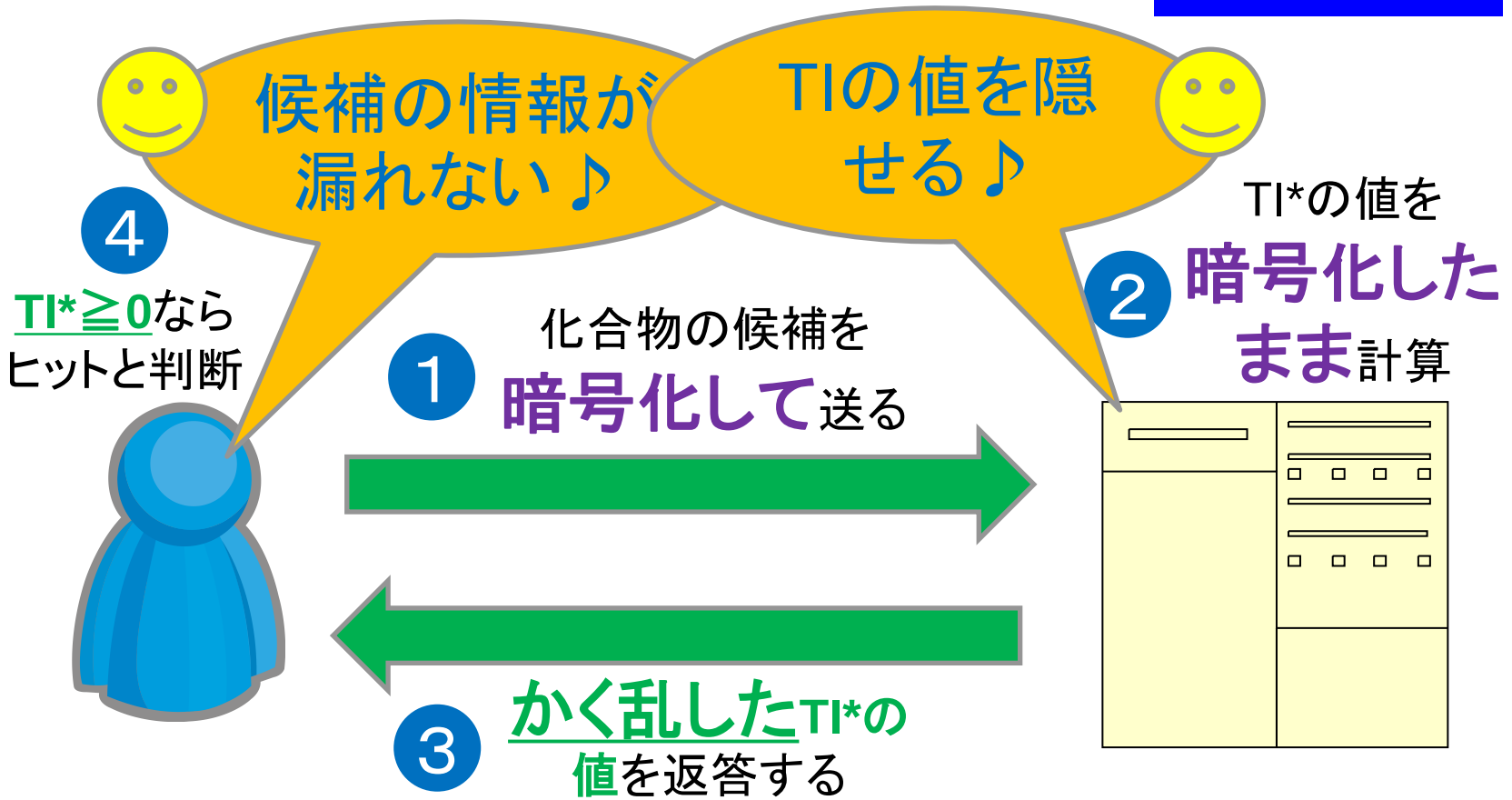
左辺は（加法）準同型暗号で計算可能！

提案手法(の原型)



しかし、これでは既存手法と変わらない効果

提案手法



結果に影響しない範囲で真の値を隠す

TIの値のかく乱方法

- 対策1：**ランダムなアフィン変換**の導入
  - TI → r･TI + s（s < r は非負の2整数）
- 対策2：**ダミー値**の導入
  - ランダムなダミー値（の暗号文）とその中の非負値の個数をTIとともにユーザに返送
  - 全体の非負値の個数から教えられた個数を引けば、真の非負値の個数がわかる

理論と実験の両面からかく乱の効果を定量的評価

ユーザのプロトコル逸脱攻撃

えっ！？何で！？
Σ(￣ロ￣lll)

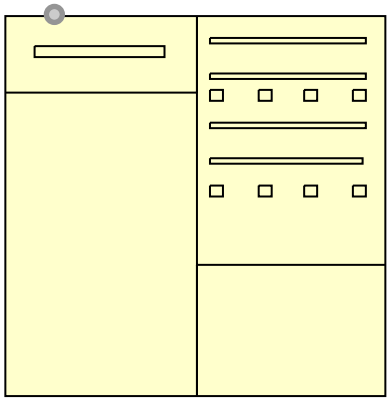
TI*の値から
DB側の**特定のビットを推測可能**

4

異常な範囲の入力値を

1 **暗号化して**送る

2 TI*の値を **暗号化したまま**計算



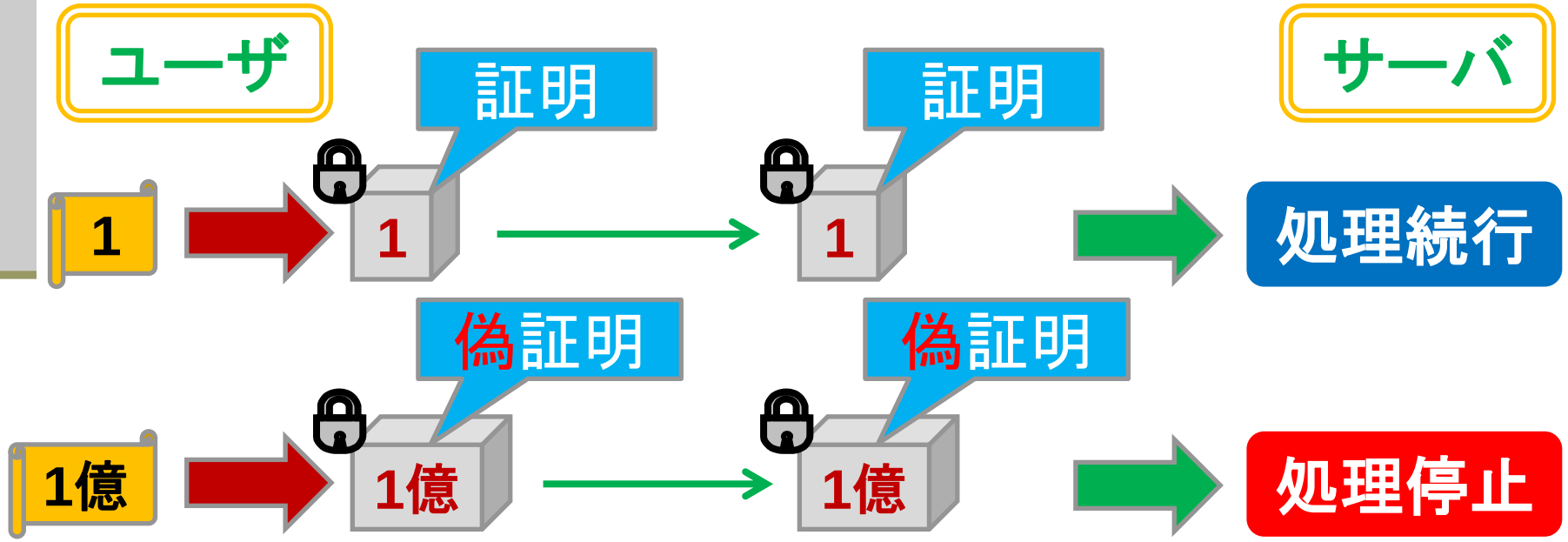
3 **かく乱したTI*の値**を返答する

悪意あるユーザにDBの中身が漏れる問題

（非対話）ゼロ知識証明

- 入力値が正常なことを全員が確認可能
- その際、**入力値の情報が漏れることはない**

という要求を実現する暗号プロトコル



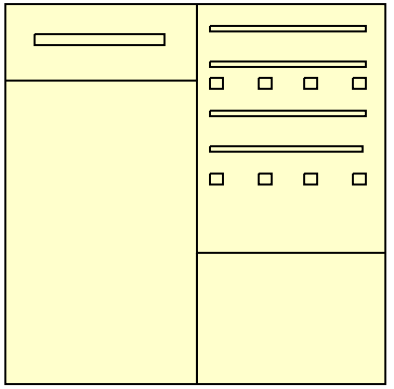
プロトコル逸脱攻撃への対処

（ちっ・・・
失敗か・・・）

異常な範囲の入力値を

1 暗号化して送る

2 ゼロ知識証明の正当性確認
→失敗



3 プロトコルを中断
→ユーザへ送信されない

正常な入力値だけを受理するプロトコル

まとめ

- 化合物データベースの類似化合物検索について、ユーザ側・サーバ側双方の情報秘匿を考慮した検索プロトコルの提案
  - （加法）準同型暗号を利用
- ユーザ側の情報秘匿は暗号化で達成
- サーバ側の情報秘匿を定量的に評価
- 非対話ゼロ知識証明の利用により、異常値を含む検索語を利用した攻撃について対策

宣伝

- 本方式の基本プロトコルは産総研生命情報工学研究センター他が開発
- その後、暗号の知見を求められて産総研セキュアシステム研究部門（RISEC）が加入
  - ダミー値、定量的評価、ゼロ知識証明、…
- 産総研RISEC・次世代セキュリティ研究グループでは、他組織・企業への知見提供を通した研究成果の社会還元を目指します