2017 中国互联网安全大会
China Internet Security Conference

同态密码技术发展与应用

张振峰

中国科学院软件研究所　研究员、副总工
可信计算与信息保障实验室　主任

ISC 2017 中国互联网安全大会

360互联网安全中心

目录

一、同态加密技术简介

二、同态加密的发展

三、多密钥全同态加密

四、同态加密的实现与应用

一、同态加密技术简介

加密数据的同态计算（Rivest et al 1978）

用户

数据 x

$Enc(x)$

$f(Enc(x))$

云服务器

函数 f

目标	• 用户希望得到$f(x)$
要求	• 用户不希望泄露数据x，（服务器不希望泄露函数f）
方案	• $Dec\left(f\left(Enc(x)\right)\right) = f(x)$

一、同态加密技术简介

Public-Key Homomorphic Encryption [RAD78]

- 密钥生成$KeyGen()$：
  - 输入安全参数，产生公钥pk，私钥sk和同态计算密钥evk
- 加密$Enc()$：
  - 输入公钥pk和消息m，输出密文c
- 解密$Dec()$：
  - 输入私钥sk和密文c，输出消息m

- 同态计算Eval()：
  - 输入pk，函数f，计算密钥evk，密文$c_1, c_2, \dots, c_k$，输出密文c^*

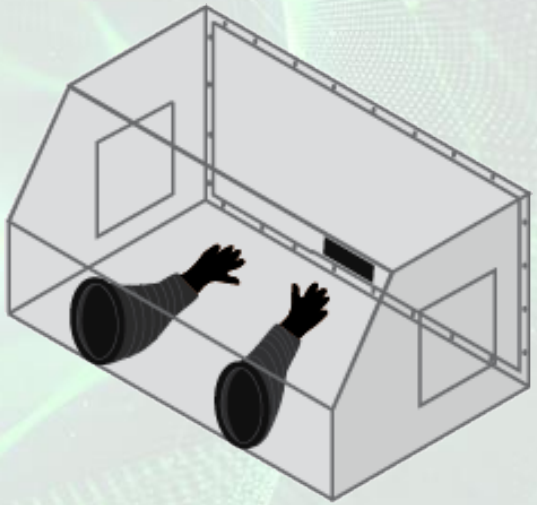
  - 逐比特同态运算布尔电路，只需构造同态加法门和同态乘法门

一、同态加密技术简介

- 正确性:

$$
\begin{array}{ccc}
c_1,\cdots,c_k & \xrightarrow{Eval(pk,evk,f,\cdot,\ldots,\cdot)} & c^* \\
\Big\downarrow Dec(sk,\cdot,\ldots,\cdot) & & \Big\downarrow Dec(sk,\cdot) \\
m_1,\cdots,m_k & \xrightarrow[f(\cdot,\cdots,\cdot)]{} & m^*
\end{array}
$$

- 紧致性：c^*的解密复杂度与f无关
- 安全性：
  - IND-CPA安全性
  - IND-CCA2不可能！ IND-CCA1？
- FHE: From private-key to public-key. [R11, TCC]



一、同态加密技术简介

- 美国“PROCEED”研究项目
  - 2010年DARPA推出“加密数据可编程计算”—PROCEED
  - 2012年白宫推出“大数据研发倡议”，PROCEED纳入其中

- 欧盟2015年启动的“HEAT”研究计划，“同态加密应用与技术”
  - HE设计与实现，HE应用
  - 工业界同态加密标准的制定和推广

二、同态加密的发展

- 1978年Rivest等人提出了同态加密的概念
- 2009年IBM的Craig Gentry在其博士论文中提出了第一个全同态的加密方案
- 2014全同态加密前沿学术论坛


Craig Gentry

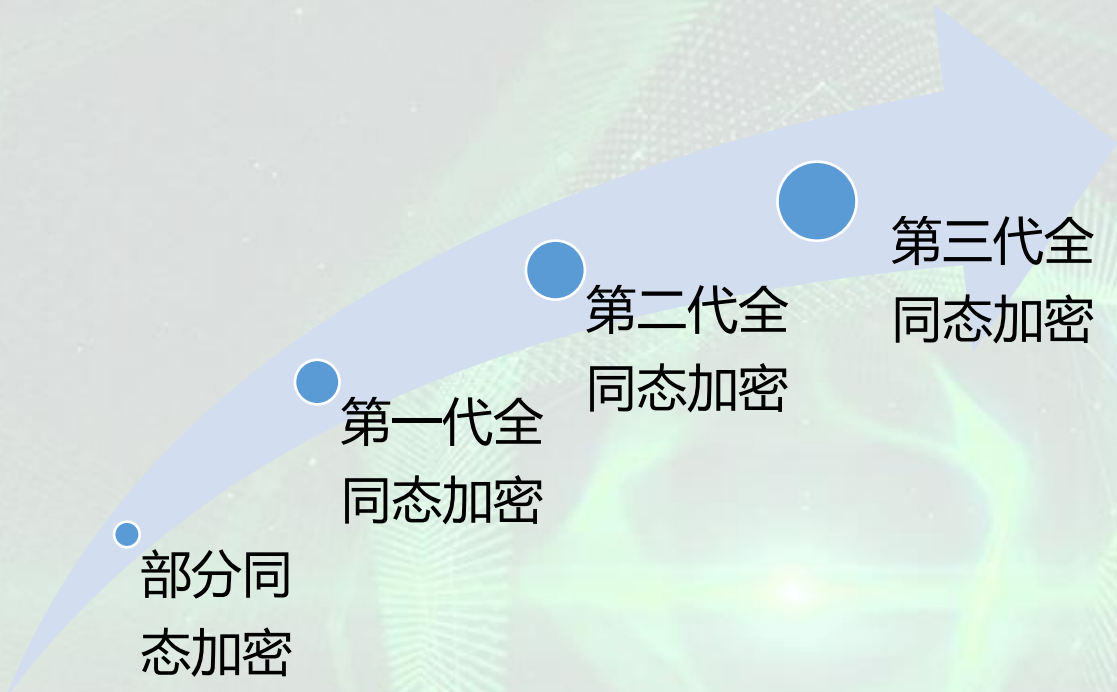
Gentry和Halevi访问TCA实验室


张振峰研究员致开幕词

Craig Gentry研究员做报告

Shai Halevi研究员做报告

二、同态加密的发展



二、同态加密的发展：部分同态

- RSA公钥加密算法 [RSA77]
  - 公钥$(e，N)$，其中$N = pq$，e满足$gcd(e, \phi(n)) = 1$
  - 加密：$Enc(m) = m^e \ mod \ N$
  - 乘法同态：$Enc(m_1) \cdot Enc(m_2) = (m_1 m_2)^e mod \ N$
    $= Enc(m_1 m_2)$
- Paillier公钥加密算法 [P99]
  - 公钥(n, g)，
  - 加密: $Enc(m) = g^m \cdot r^n mod \ n^2$
  - 加法同态：$Enc(m_1) \cdot Enc(m_2) = g^{m_1+m_2} \cdot (r_1 r_2)^n mod \ n^2$
    $= Enc(m_1 + m_2)$

二、同态加密的发展：FHE一代

Gentry's Somewhat FHE（整数版本）

- 密钥生成
  - 选择大的奇数p作为私钥
- 加密操作
  - 选择随机数r保证 $|r| \ll p$
  - 加密明文比特$m \in \{0,1\}$，密文为$c = pq + 2r + m$
- 解密操作
  - 计算$m = (c \bmod p) \bmod 2$
- 同态计算：
  - $c^{+} = (q_1 + q_2)p + 2(r_1 + r_2) + (m_1 + m_2)$
  - $c^{\times} = (q_1 c_2 + q_2 c_1 - p q_1 q_2)p + 2(m_1 r_2 + r_1 m_2 + 2 r_1 r_2) + m_1 m_2$

Idea of Bootstrapping

- 近似同态加密算法在进行同态运算时会使噪音增长，噪音超过上界会导致密文无法正常解密
  - SFHE可以同态计算深度不超过l的电路
  - 当电路深度超过l时？
- 自举算法：假设SFHE可以同态计算自己的解密电路$Dec(sk, c)$
  - 输入$Enc_{pk}(sk)$和$Enc_{pk}(\mathrm{c})$，同态计算$f = Dec(sk, c)$，得到密文$c^* = Enc_{pk}(Dec(sk, c)) = Enc_{pk}(m)$
  - c^*用来同态计算，不断循环

二、同态加密的发展：FHE一代

- 第一代全同态加密思想
  - 利用特殊的代数结构——多项式环或者整数环的理想
  - 不同理想表示不同的明文消息
  - 理想本身具有加法和乘法的同态性
- 特点：
  - 基于非标准困难假设
  - 噪音指数增长：进行d次同态操作，需要模数 $q > B^d$
- 代表性工作
  - Gentry09方案——Gentry 发表于 STOC 09

二、同态加密的发展：FHE二代

- 第二代FHE思想：分层的同态加密——每层密文有不同的解密密钥，第i层密文经过同态计算得到第$i+1$层密文，引入计算密钥

- 特点
  - 标准困难假设——LWE问题或者ring LWE问题
  - 噪音亚指数增长：进行d次同态操作，需要模数 $q > B^{O(\log d)}$
  - Overhead=(同态加密计算时间)/(明文计算时间)可达polylog
  - 同态操作需要计算密钥

- 代表性工作
  - BV11方案—Brakerski 和 Vaikuntanathan，FOCS 2011
  - BGV12 — Brakerski，Gentry，Vaikuntanathan，ITSC12

二、同态加密的发展：FHE二代

BV11方案 [Brakerski-Vaikuntanathan，FOCS 2011]

- 密钥生成：$\boldsymbol{s} \leftarrow \mathbb{Z}_q^n$
- 加密：
  - 选择随机向量$\boldsymbol{a} \leftarrow \mathbb{Z}_q^n$和噪音$e \leftarrow \chi$
  - 对消息$m \in \{0,1\}$，计算$\tilde{c} = \langle \boldsymbol{a}, \boldsymbol{s} \rangle + 2e + m$，密文$\boldsymbol{c} = (\tilde{c}, \boldsymbol{a})$
- 解密
  - 计算$m = (\tilde{c} - \langle \boldsymbol{a}, \boldsymbol{s} \rangle) \bmod q \bmod 2$
  - 令$\bar{\boldsymbol{s}} = (1, -\boldsymbol{s}) \in \mathbb{Z}_q^n$，则解密算法满足$\langle \boldsymbol{c}, \bar{\boldsymbol{s}} \rangle = 2e + m$
- 同态加法
  - $\boldsymbol{c}^+ = (\tilde{c}_1 + \tilde{c}_2, \boldsymbol{a}_1 + \boldsymbol{a}_2)$
  - 其中$\tilde{c}_1 + \tilde{c}_2 = \langle \boldsymbol{a}_1 + \boldsymbol{a}_2, \boldsymbol{s} \rangle + 2(e_1 + e_2) + (m_1 + m_2)$

二、同态加密的发展：FHE二代

BV11方案 — 乘法同态

- 令计算张量积 $\tilde{\boldsymbol{c}}_3 = \boldsymbol{c}_1 \otimes \boldsymbol{c}_2$ ，$\tilde{\boldsymbol{s}} = \bar{\boldsymbol{s}} \otimes \bar{\boldsymbol{s}}$ ，乘法同态性

  $\langle \tilde{\boldsymbol{c}}_3, \tilde{\boldsymbol{s}} \rangle = (2e_1 + m_1)(2e_2 + m_2) = 2(2e_1e_2 + m_1e_2 + m_2e_1) + m_1m_2$
- 问题：$\tilde{\boldsymbol{c}}_3$的维数为$(n+1)^2$ ，如何避免维数的指数增长？

- 解决方案：使用维数转换 技术将$\tilde{\boldsymbol{c}}_3$转换为$n+1$维密文$\boldsymbol{c}_3$
- 计算密钥：$evk = \left(\boldsymbol{A} \ , \boldsymbol{B} = \boldsymbol{A}\boldsymbol{s}' + 2\boldsymbol{e} + \text{Powerof2}(\tilde{\boldsymbol{s}}) \in \mathbb{Z}_q^{nl}\right)$
- 转换后密文$\boldsymbol{c}_3 = (\text{BitDecomp}(\tilde{\boldsymbol{c}}_3)^T \boldsymbol{B}, \text{BitDecomp}(\tilde{\boldsymbol{c}}_3)^T \boldsymbol{A}) \in \mathbb{Z}_q^{n+1}$

$$\begin{aligned} \text{BitDecomp}(\tilde{\boldsymbol{c}}_3)^T \boldsymbol{B} &= \text{BitDecomp}(\tilde{\boldsymbol{c}}_3)^T \boldsymbol{A}\boldsymbol{s}' + 2\boldsymbol{e}' + \langle \tilde{\boldsymbol{c}}_3, \tilde{\boldsymbol{s}} \rangle \\ &= \text{BitDecomp}(\tilde{\boldsymbol{c}}_3)^T \boldsymbol{A}\boldsymbol{s}' + 2\boldsymbol{e}'' + m_1m_2 \end{aligned}$$

- 比特分解：$x \in \mathbb{Z}_q$ ，$\text{BitDecomp}(x) = (x_0 \quad \cdots \quad x_{l-1}) \in \{0,1\}^l$
- 二次幂乘：$y \in \mathbb{Z}_q$ ，$\text{Powerof2}(y) = (y, 2y, \quad \cdots \quad 2^i \cdot y_{l-1}) \in \mathbb{Z}_q^l$

二、同态加密的发展：FHE二代

- 模数转换技术
  - Brakerski-Gentry-Vaikuntanathan 在 ITSC 2012提出的BGV方案
  - 选择$p < q$，对$\frac{p}{q} \cdot \boldsymbol{c}$取整，得到模$p$下的合法密文，噪音减小!
  - 选取一系列模数$q_L, \ldots, q_0$，满足$q_i/q_{i-1} > O(B)$；当噪音上界由B变成B^2时，将模数q_{i-1}转换成q_i，使得转换后的噪音小于B
  - 进行d次同态操作，需计算$O(\log d)$层电路,只要模数 $q_L > B^{O(\log d)}$
- 基于ring-LWE的全同态加密
  - 明文为一个环元素而非一比特，密文/明文扩展比为$O(\log n)$
  - 公钥大小为$O(n)$而非$O(n^2)$
  - 同态乘法可以用快速傅里叶变换实现，复杂度为$\tilde{O}(n)$而非$\tilde{O}(n^3)$
- 批量密文技术：密文对应于消息向量，一次同态操作对应于向量中所有元素的批量操作，ring-LWE BGV方案$Overhead = poly(L, \log k, \log w)$

二、同态加密的发展：FHE三代

➢第三代FHE思想：使用近似特征向量的结构

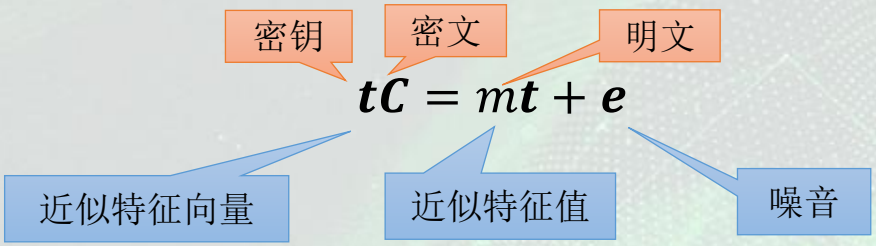
- 特点
  - 相比于FHE二代，第三代全同态加密在进行同态计算时不需要额外的evk而可以直接进行同态计算
  - 乘法噪音积累为伪线性：
    进行d次同态操作，需要保证模数 $q > poly(n) \cdot \tilde{O}(d) \cdot B$
  - 即使基于ring-LWE方案，明文空间为一个比特
  - 不支持基于中国剩余定理的批量密文技术

- 代表工作：
  - GSW13——Gentry Sahai 和 Waters 发表于 CRYPTO 2013

二、同态加密的发展：FHE三代

近似特征向量

- 定义：



- 性质：
  - $t(C_1 + C_2) = (m_1 + m_2)t + (e_1 + e_2)$
  - $t(C_1 \cdot C_2) = (m_1 t + e_1) \cdot C_2 = m_1 m_2 t + (e_1 C_2 + m_1 e_2)$
- 问题：如何保证新噪音$e_1 C_2$足够小？

- 工具矩阵[MP12]：存在$n \times nl$二次幂乘矩阵$\mathbf{G}$及可计算逆函数$\mathbf{G}^{-1}$，
  对任意的$C \in \mathbb{Z}_q^{n \times nl}$: $\mathbf{G}\mathbf{G}^{-1}(C) = C$

二、同态加密的发展：FHE三代

GSW方案 [Gentry-Sahai-Waters， CRYPTO 2013]

- 密钥生成：
  - $\boldsymbol{s} \leftarrow \mathbb{Z}_q^{n-1}$，私钥$\boldsymbol{t} = (1, -\boldsymbol{s}^T)$，
  - $\boldsymbol{A} \leftarrow \mathbb{Z}_q^{(n-1)\times nl}$，$\boldsymbol{e} \leftarrow \chi^{nl}$，公钥$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{s}^T\boldsymbol{A} + \mathbf{e} \\ \boldsymbol{A} \end{pmatrix} \in \mathbb{Z}_q^{n\times nl}$
- 加密：$\boldsymbol{R} \leftarrow \{0,1\}^{nl\times nl}$，密文$\boldsymbol{C} = \boldsymbol{P}\boldsymbol{R} + m\boldsymbol{G} \in \mathbb{Z}_q^{n\times nl}$
- 解密：$\boldsymbol{t}\boldsymbol{C} = m\boldsymbol{t}\boldsymbol{G} + \boldsymbol{e}'$
- 同态计算：
  - $\boldsymbol{t}(\boldsymbol{C}_1 + \boldsymbol{C}_2) = (m_1 + m_2)\boldsymbol{t}\boldsymbol{G} + (\boldsymbol{e}_1 + \boldsymbol{e}_2)$
  - $\boldsymbol{t}(\boldsymbol{C}_1 \cdot \mathbf{G}^{-1}(\boldsymbol{C}_2)) = (m_1\boldsymbol{t}\boldsymbol{G} + \boldsymbol{e}_1) \cdot \mathbf{G}^{-1}(\boldsymbol{C}_2) = m_1 m_2 \boldsymbol{t}\boldsymbol{G} + (\boldsymbol{e}_1 \mathbf{G}^{-1}(\boldsymbol{C}_2) + m_1\boldsymbol{e}_2)$
- 乘法噪音的积累是非对称的，是（近似）相加而不是相乘；
- 明文空间为1比特，通过NAND门同态实现一般电路

二、同态加密的发展：FHE三代

	LWE BGV	ring-LWE BGV	LWE GSW	ring-LWE GSW
密文/明文比	$\tilde{O}(n)$	$O(\log n)$	$\tilde{O}(n^2)$	$\tilde{O}(n)$
同态操作复杂度	$\tilde{O}(n^3)$	$\tilde{O}(n)$	$\tilde{O}(n^{2.373})$	$\tilde{O}(n)$
自举噪音积累	$SuperPoly(n)$	$SuperPoly(n)$	$Poly(n)$	$Poly(n)$
同态计算代价	$\tilde{O}(n^2)$	$Polylog(n)$	$\tilde{O}(n^{1.373})$	$\tilde{O}(n)$

[CZ17]：在同态计算一类特殊的电路（如内积函数）时，基于ring-LWE的BGV方案可以同时实现$Poly(n)$噪音积累下自举且同态计算代价不超过$Polylog(n)$

L.Chen, Z.Zhang, Bootstrapping Fully Homomorphic Encryption with Ring Plaintexts within Polynomial Noise, PROVSEC 2017

三、多密钥全同态加密

- 参与方$1, 2, \ldots, N$有其各自的密钥
- 不同公钥 $pk_1, pk_2, \ldots, pk_N$ 加密的密文之间可以进行同态运算，得到一个共同公钥pk^*下加密的密文
- 共同公钥pk^*加密的密文可以通过共同解密的方式解密
- 支持门限解密的多密钥全同态加密可以构造对两轮的任意函数的多方安全计算协议

三、多密钥全同态加密

MKFHE-GSW方案 [Clear-McGoldrick，CRYPTO 2014]

- 参与方$1, 2, \ldots, N$有各自的GSW密钥
  - 用户i拥有私钥$\boldsymbol{t}_i \in \mathbb{Z}_q^n$
  - 扩展私钥$\boldsymbol{t}^* = (\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_N)$

- 密文扩展：
  - 参与方i的密文$\boldsymbol{C} \in \mathbb{Z}_q^{n \times nl}$满足$\boldsymbol{t}_i \boldsymbol{C} \approx m\boldsymbol{t}_i \boldsymbol{G}$
  - （公开计算）扩展密文$\boldsymbol{C}^* \in \mathbb{Z}_q^{Nn \times Nnl}$满足$\boldsymbol{t}^* \boldsymbol{C}^* \approx m\boldsymbol{t}^* \boldsymbol{G}^*$，$\boldsymbol{G}^*$为扩展的工具矩阵

$$\boldsymbol{G}^* = \begin{pmatrix} \boldsymbol{G} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{G} \end{pmatrix}$$

  - 对扩展密文进行GSW形式同态操作

三、多密钥全同态加密

MKFHE-BGV方案[Chen-Zhang-Wang，TCC 2017]

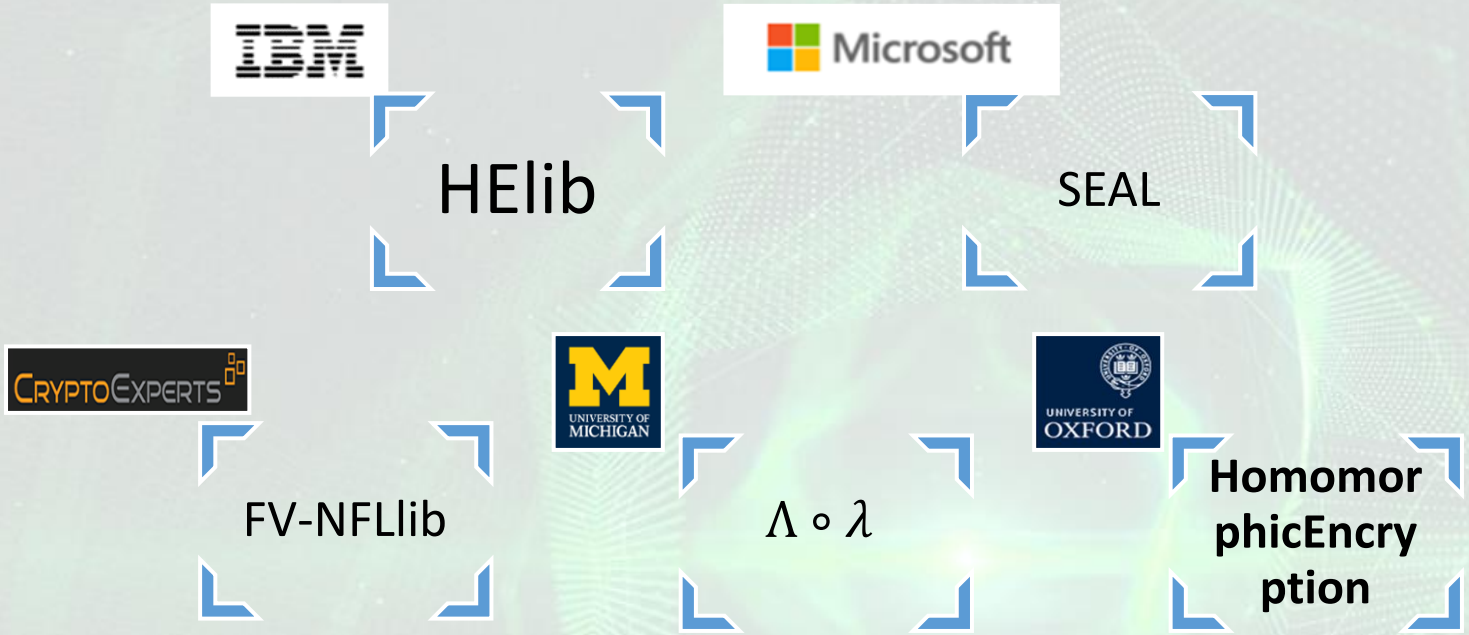
- 参与方$1, 2, \dots, N$有各自的BGV密钥
  - 用户i拥有私钥$\boldsymbol{t}_i \in \mathbb{Z}_q^{n+1}$；扩展私钥$\boldsymbol{t}^* = (\boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_N)$
- 密文扩展：
  - 参与方i的密文$\boldsymbol{c}_i$满足$\langle \boldsymbol{t}_i, \boldsymbol{c}_i \rangle = m_i + 2e_i$
  - 扩展密文$\boldsymbol{c}^* = (0, \dots, \boldsymbol{c}_i, \dots, 0) \in \mathbb{Z}_q^{N(n+1)}$，满足$\langle \boldsymbol{t}^*, \boldsymbol{c}^* \rangle = \langle \boldsymbol{t}_i, \boldsymbol{c}_i \rangle$
- evk生成：通过GSW密文的前l列构造BGV方案的计算密钥！

- 特点
  - 基于ring-LWE时，明文空间为环元素而非一比特
  - 支持基于CRT的批量密文技术，同态计算代价为$poly(N, \log n)$
  - 密文扩展复杂度与密文个数无关，只与安全参数相关

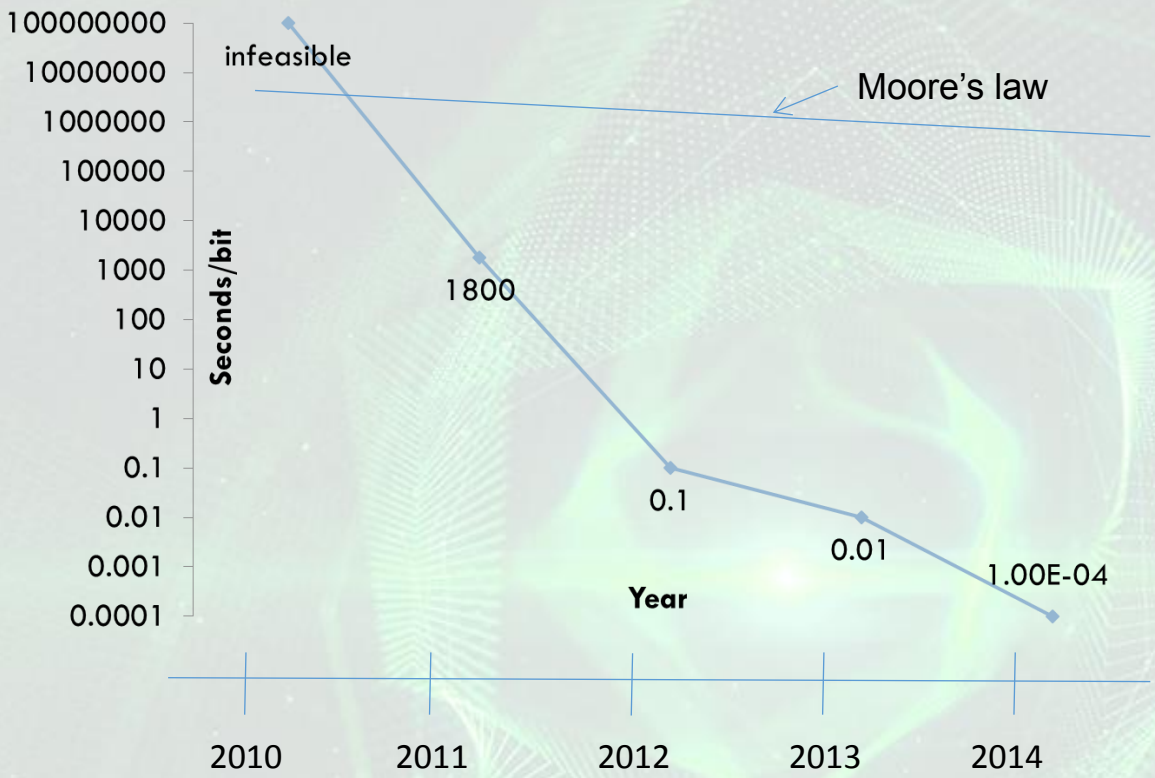
- **L.Chen，Z.Zhang, X.Wang, Batched Multi-hop Multi-key FHE from ring-LWE with Compact Ciphertext Extension, TCC 2017**

四、同态加密的实现与应用

全同态加密的开源实现库



四、同态加密的实现与应用



加密数据计算的速度： from Gentry's talk

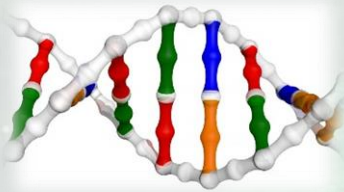
四、同态加密的实现与应用

全同态加密效率实例

- HElib实现了基于ring-LWE的BGV方案
  - 同态实现 “AES-128” 电路：AES 本身大约有 ~30,000 门，HElib在4-15 分钟内处理100-200 个blocks
    - 不进行自举: 245 秒120 blocks
    - 进行自举: 1050 秒180 blocks
  - 同态计算集合交：在五分钟内计算两个加密集合的交集，每个集合大约100,000个元素
- Chillotti et al. ASIACRYPT 2016: “Faster Fully Homomorphic Encryption: Bootstrapping in less than 0.1 Seconds”.
  - 基于GSW方案，明文空间为1比特

四、同态加密的实现与应用

应用实例：iDASH



$Enc(x)$

$f(Enc(x))$

x：病人的DNA信息

病人得知自己的致病基因片段$f(x)$

f：确定潜在致病基因片段

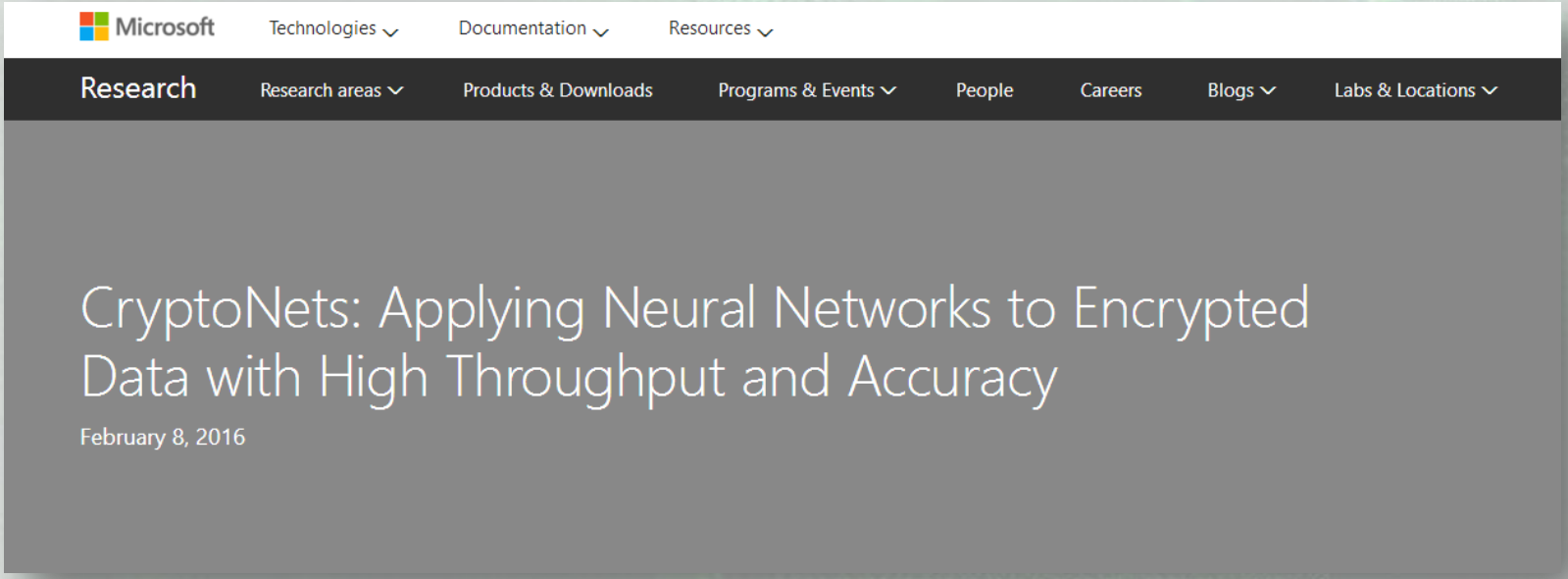
医疗机构不知道任何关于x的信息

- iDASH：美国国家卫生局发起的基因检测中的隐私与安全挑战
- Cheon-Kim-Song: 3.9秒从4M数据中定位并提取询问的基因片段

四、同态加密的实现与应用

应用实例：隐私保护

- Microsoft’ s CryptoNets（使用SEAL库）实现对加密数据的神经网络学习
  - 应用于MNIST手写字识别训练库
  - 99%准确率，单一PC上每小时51,739次预测，570秒延迟



四、同态加密的实现与应用

同态加密标准化

- 同态加密标准研讨会在2017年7月13-14日，由Microsoft Research in Redmond举办
- 白皮书：http://homomorphicencryption.org/

谢 谢