Transaction Processing on Confidential Data using Cipherbase

Arvind Arasu, Ken Eguro, Manas Joglekar*
Raghav Kaushik, Donald Kossmann, Ravi Ramamurthy
Microsoft Research
Stanford University*
Cloud Data Security Concerns

Data in the cloud vulnerable to:

- Snooping administrators
- Hackers with illegal access
  - Compromised servers
Database Encryption

- Client App

- SQL

- SELECT BranchId, Balance FROM Account WHERE AccountId = 4

- Account Table
<table>
<thead>
<tr>
<th>AccountId</th>
<th>BranchId</th>
<th>Balance</th>
<th>SSN</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>000-00-0000</td>
<td>Edges</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1000</td>
<td>100-10-1000</td>
<td>Mike</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2000</td>
<td>200-20-2000</td>
<td>Jim C</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3000</td>
<td>300-30-3000</td>
<td>Phil</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4000</td>
<td>400-40-4000</td>
<td>Davis</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5000</td>
<td>500-50-5000</td>
<td>Mohit</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>6000</td>
<td>600-60-6000</td>
<td>Davis</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>7000</td>
<td>700-70-7000</td>
<td>Seng</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>8000</td>
<td>800-80-8000</td>
<td>Hector</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>9000</td>
<td>900-90-9000</td>
<td>Rajesh</td>
</tr>
</tbody>
</table>
Database Encryption

```
SELECT BranchId, Balance
FROM Account
WHERE AccountId = 0x8de526fdec17ed9ca4ba...
```
Cipherbase Summary

• Data Confidentiality:
  – Strong column-level encryption
  – Decoupled from functionality
  – *Lightweight “trusted module” in secure hardware

• Functionality:
  – Industrial Strength Database system (SQL Server)
  – Concurrency, Recovery, Stored Procedures.

• Performance on TPCC
  – 85% of plaintext for typical encryption
  – 40% of plaintext for “worst case” encryption

No prior work with this {Confidentiality, Functionality, Performance} characteristics
Organization

• Introduction

• Solution Landscape & Design Choices

• Cipherbase Design & Engineering

• Evaluation
What Makes Encryption Challenging?

Select Sum (Score)
From Assignment
Where StudentId = 1
Solution Landscape

• Two fundamental techniques
  – Directly compute over encrypted data
    • Special *homomorphic* encryption schemes
Deterministic Encryption

\[ \sigma_{\text{StudentId}=1} \]

```
select *
from assignment
where studentid = 1
```

<table>
<thead>
<tr>
<th>StudentId</th>
<th>AssignId</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>99</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Deterministic Encryption

\[ \sigma_{\text{StudentId\_det}=bd6...} \]

```
select *
from assignment
where studentid_det = bd6e7c3df2b5779e0b61216e8b10b689
```
Homomorphic Encryption Schemes

**Impractical**

- Fully Homomorphic Encryption
  - [G09, G10]
  - (Any function)

**Expensive**

- Partial Homomorphic Encryption
- Paillier Cryptosystem [P99]
- ElGamal Cryptosystem [E84]

**Practical**

- Order-Preserving Encryption
  - [BCN11, PLZ13]
  - (≤)

- Deterministic Encryption
  - (==)

- Non-Deterministic Encryption
  - (∅)
PHE Limitations

• Limited Server Functionality
  – \( \text{SUM}(\text{L\_EXTENDEDPRICE} \times (1-\text{L\_DISCOUNT}) \times (1+\text{L\_TAX})) \)

• Data Security tied to functionality

• Lack of Composability
  – \( A + B = C \)

• Performance
  – \( \approx \text{msec} \) for a single addition under Paillier

CryptDB [PRZ+11], Monomi [TFM 13], [HMH08]
Solution Landscape

• Two fundamental techniques
  – Directly compute over encrypted data
    • Special homomorphic encryption schemes
    • Challenge: limited class of computations
    • Challenge: Not composable
  – Use a “secure” location
    • Hardware provisioned isolation and protection
    • Computations on plaintext
Secure Location
Secure Hardware Landscape

• Long history
  – Banking, Defense Applications

• Becoming mainstream and commoditized

• Players:
  – Crypto co-processors
  – FPGAs
  – Intel SGX
  – TPM, HSM
Intel Software Guard Extensions

• Extensions to Intel Architecture
• Isolation to code + data within a designated region called *enclave*
  – Confidentiality
  – Integrity

Virtual Addr Space

Enclave

code/data

Physical Memory

Ack: Andrew Baumann

[MAB+ 13, AGJ+ 13, HLP+ 13]
Design Choice: Trusted Functionality

- Haven [MPH14]: Larger Trusted Computing Base (TCB)
- TrustedDB [BS11]: Smaller TCB
- Cipherbase

Secure h/w

Commodity h/w

DBMS

Library OS

OS

Smaller TCB
Design Choice: Trusted Functionality

Haven [MPH14]
- OS
- Commodity h/w
- DBMS
- Secure h/w
- Library OS

TrustedDB [BS11]
- OS
- Commodity h/w
- DBMS
- Secure h/w
- Embedded OS

Cipherbase
- OS
- Commodity h/w
- DBMS
- Secure h/w
- Expr Eval

Expr
Eval

More secure
Less secure
Design Choice: Trusted Functionality

Minimal software engg.

Haven [MPH14]

TrustedDB [BS11]

Cipherbase
Organization

• Introduction

• Solution Landscape & Design Choices

• Cipherbase Design & Engineering

• Evaluation
Life of a Query in Cipherbase

Cipherbase Server

Modified SQL Server

FPGA

Network

Cipherbase Client Lib

App

Encryption Config

UPDATE Account
SET Balance = Balance + 10
WHERE AccountId = 4

UPDATE Account
SET Balance = TNEval(5, Balance)
WHERE AccountId = 4

AccountId: Plaintext
BranchId: AES-CBC
Balance: AES-CBC

4/15/2015
## Life of a Query in Cipherbase II

### Cipherbase Server

<table>
<thead>
<tr>
<th>AccountId</th>
<th>BranchId</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0x6C56A4C7030DCDF1DA14D228EDE9DA73872FDCDE6E66F6...</td>
<td>0x3CE12E3A288DE208EA286A15F7386AEDFB8EBEBF85C34...</td>
</tr>
<tr>
<td>2</td>
<td>0x8F48B350D0C07F63F90725C1658694E4236D45C703546...</td>
<td>0x8D0705C5212057EB87E8585B9021167D307402BFC04C7...</td>
</tr>
<tr>
<td>3</td>
<td>0x6E0B03443623C0CAFC71B7C7043C0CE1177179D91946135A...</td>
<td>0xE19E2D3292F02AEC74AB048001FD4F502797124411DE6634...</td>
</tr>
<tr>
<td>4</td>
<td>0x20B869A4710CD13E8025AA50D8A5C0E9DD17017D931252...</td>
<td>0x86005BC654341196C3A584008880830BD05C43207232A...</td>
</tr>
<tr>
<td>5</td>
<td>0x4C0D3CB40DF8547298798635D043885FC134202E2D47...</td>
<td>0x87C146568839F5875772D454D011B4F02074530835975D...</td>
</tr>
<tr>
<td>6</td>
<td>0x4C69B4C63140362C6F574DA727250D67FON144D0E6B6...</td>
<td>0xF754A15590C558EB36499964A0F1043948E35592571E56F1...</td>
</tr>
<tr>
<td>7</td>
<td>0x4A150C1A311807C2D03B3B67A8E3257C0480B1C50461...</td>
<td>0x3CC351B3ED221D9B9910FC6581592DAA74E7ACD9E3...</td>
</tr>
<tr>
<td>8</td>
<td>0xC87F929CD435407C06524A2DA3D0C558A987026E35...</td>
<td>0x877CE70BCEE30F2D5003E85423AB3451954D708D850...</td>
</tr>
<tr>
<td>9</td>
<td>0x1DD6B37DA74580C40718DD647C7338E888DF95F7161EC15778F...</td>
<td>0x9F9B590238E7D21C31573CC3195F8ED206A57551D4E...</td>
</tr>
<tr>
<td>10</td>
<td>0x4F5166CDE4F23ADCC55350A552A9F0C63A47384E50C0...</td>
<td>0x2D235C788CB3601D52762E1A9526A527370675372282...</td>
</tr>
</tbody>
</table>

**PK:** AccountId: AES-CBC  
BranchId: AES-CBC  
Balance: AES-CBC
B+-Tree Indexes over Encrypted Data
B+-Tree Indexes over Encrypted Data
Life of a Query in Cipherbase II

UPDATE Account
SET Balance = TNEval(S, Balance)
WHERE AccountId = 0x8de526fdec17ad9ca4ba3de246aa...

App
 Cipherbase Client Lib

Network

Cipherbase Server

Stack Machine
(Expression Evaluation)

FPGA

Modified SQL Server

PCIe

Insecure (x86)

PK: AccountId: AES-CBC
BranchId: AES-CBC
Balance: AES-CBC

4/15/2015
ICDE 2015
B+-Tree Indexes over Encrypted Data

Search key: **8DE526**

\[ \text{comp}(8DE526, 0A183E) \]
B+-Tree Indexes over Encrypted Data

Search key: 8DE526

comp(8DE526, 0A183E)
Life of a Query in Cipherbase II

**UPDATE Account**
SET Balance = TNEval(5, Balance)
WHERE AccountId = 0x8de526fdcf1d7a0ca4ba3de246aa...

<table>
<thead>
<tr>
<th>AccountId</th>
<th>BranchId</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0x4C25A8A4AC203DCEEF1DA14022BBD9DA7B3F7FDBD0D5EBF7B...</td>
<td>0x5CEC1E3B6C000E2A0D5EAF7F3BAC0D3B3B1E0B1C48...</td>
</tr>
<tr>
<td>2</td>
<td>0xE9F61B638D12406F4A078F8F4978F2B48DB3D1D31E28...</td>
<td>0x56071C5C021795674B0679F65EF201187D0C59A42BF1C0C7...</td>
</tr>
<tr>
<td>3</td>
<td>0xF6F8954A39825009CAF7718376BE3ECE71771D98161F5618...</td>
<td>0x618E207972512AE7614A8848D1F2014F687F13411ED68...</td>
</tr>
<tr>
<td>4</td>
<td>0x2C4D52F7CD1C4802B4A90E3DEBD75DAE9AE0BDEB13186...</td>
<td>0x1BE1F61BEF2954D3EB19866A9100010000B4499F3A7C59C2E34...</td>
</tr>
<tr>
<td>5</td>
<td>0x0C22D55A00FB389426A31B929B4A3985FCE194D22ED7A...</td>
<td>0x4872C146803B5F87879A1EAD3D1B4D2D2D89E93857D...</td>
</tr>
</tbody>
</table>

**PK:** AccountId: AES-CBC
BranchId: AES-CBC
Balance: AES-CBC
### Operational Security

<table>
<thead>
<tr>
<th>Operation</th>
<th>Adversary Learns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{A=5}(R)$</td>
<td>Unknown predicate $p(A)$ over $R$ tuples</td>
</tr>
<tr>
<td>$R \bowtie_A S$ (hash-based)</td>
<td>The join graph and the equivalence relation over $R(A)$ and $S(A)$ for joining $A$ values</td>
</tr>
<tr>
<td>$\pi_{A+B}(R)$</td>
<td>Nothing</td>
</tr>
<tr>
<td>$\text{Groupby}_{A}^{\text{SUM}(B)}(R)$</td>
<td>The equivalence relation over $R(A)$</td>
</tr>
</tbody>
</table>

Data Security depends on the operations performed
Transaction Processing Performance Challenges

Life of a transaction

TPCC New Order:

x86

FPGA

1M instrs

≈ 10 instrs x 300

Expression evaluation

≈ μsec

Time/progress

Expression evaluation
Summary of Performance Optimizations

- **Cipherbase Client Lib**
- **Modified SQL Server**
- **Batch FPGA work**
  - Amortize communication latency
- **Multiple FPGA cores**
  - Parallelism
  - More FPGA compute
- **Plaintext Data Caches**
  - Minimize network comm.
  - Reduce decryption

**Expression folding**
- Minimize FPGA roundtrips

**Vectorize index comparisons**
- Minimize FPGA roundtrips

**Plaintext Data Caches**
- Core 1
- Core 2
- Core 3
- Core 4
Organization

• Introduction

• Solution Landscape & Design Choices

• Cipherbase Design & Engineering

• Evaluation
Cipherbase Prototype

• SQL Server code
  – Basic functionality
    • \( \approx \) 1000 LoC
    • Localized to expression evaluation module
  – Optimizations
    • \( \approx \) 5000-10000 LoC
    • Localized to FPGA driver, indexing
  – Unchanged: everything else
Performance on TPCC

Encryption schemes:

Customer: Customer PII data strongly encrypted

Strong/Weak: Index columns deterministic, all others strongly encrypted

Strong/Strong: All columns strongly encrypted

Increasing strength of encryption

Transactions per sec (relative to SQL Server)

Plaintext  Customer  Strong/Weak  Strong/Strong

Opt  NoOpt
Cipherbase Summary

• Security:
  – Strong encryption
  – Decoupled from functionality

• Functionality:
  – Industrial Strength Database system (SQL Server)
  – Transaction Processing

• Performance on TPCC
  – 85% of plaintext for typical encryption
  – 40% of plaintext for “worst case” encryption

• Lightweight “trusted module” in secure hardware