Recall: Lec-11

In lec-11, we learned:

- Cryptographic basics
- Symmetric key cryptography
- Public key cryptography
- Hash functions
- Case study: CryptDB
Lecture Roadmap

- Differential Privacy
- Case Study: DJoin
- Midterm Review
Much information is constantly accumulating in databases all around the world ...
Good Uses to this Data

- Recommendation system analysis
- Social networking analysis
- Illness analysis
- ...

Recommendation system analysis
Social networking analysis
Illness analysis
...
Good Uses to this Data

However, in some cases, using this data has potential privacy issue...
Example in Statistic DB

- Bob: Cancer
- Doris: Malaria
- Hank: Malaria
- Greg: HIV

Statistical database only with COUNT query

Adversary
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

<table>
<thead>
<tr>
<th>Bob</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doris</td>
<td>Malaria</td>
</tr>
<tr>
<td>Hank</td>
<td>Malaria</td>
</tr>
<tr>
<td>Greg</td>
<td>HIV</td>
</tr>
</tbody>
</table>

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Statistical database only with COUNT query

A: 2

Q: How many people have Malaria?

Aha! That's Hank
The adversary infers sensitive information of individual with basic queries ...
Differential Privacy

Differential privacy can be used to address this type of problem
Differential Privacy [Dwork, ICALP’06]

- Differential privacy is a property of randomized function that take a dataset as input and return an aggregate result.
- Randomized function is differentially private if arbitrary changes to a single item result in only statistically insignificant changes in the function’s output.
- The presence or absence of any single item has a statistically negligible effect.
Differential Privacy [Dwork, ICALP’06]

In distributed system area, differential privacy is achieved by adding carefully chosen noises to the output of queries.

Differential privacy does not make any assumptions on adversary.
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

<table>
<thead>
<tr>
<th>Bob</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doris</td>
<td>Malaria</td>
</tr>
<tr>
<td>Hank</td>
<td>Malaria</td>
</tr>
<tr>
<td>Greg</td>
<td>HIV</td>
</tr>
</tbody>
</table>

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

Statistical database only with COUNT query
The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Q: How many people have Malaria?

A: 2 ± 1

Statistical database only with COUNT query
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not, but what about Hank?

Statistical database only with COUNT query
Differential privacy is proposed for statistical databases which only support computation operations (e.g., COUNT).
Differential Privacy

• Differential privacy is proposed for statistical databases which only support computation operations (e.g., COUNT).

• Differential privacy adds carefully chosen noises to the output of a query.
Differential Privacy

- Differential privacy is proposed for statistical databases which only support computation operations (e.g., COUNT).

- Differential privacy adds carefully chosen noises to the output of a query.

- More noises you add you will have higher level privacy but less accuracy of your result.
Differential Privacy

- Differential privacy adds carefully chosen noises to the output of a query.
Differential Privacy

- Differential privacy adds carefully chosen noises to the output of a query.
Lecture Roadmap

• Differential Privacy
• Case Study: DJoin
• Midterm Review
Example in Statistic DB

- The adversary knows Doris has Malaria and Bob and Greg do not.

Does differential privacy always work?
Unfortunately, differential privacy-based approaches (e.g., PINQ) assume all the data is from a single database ...
Distributed DB Scenario

**Airlines**
- Doris
- Hank
- Emil
- Bob
- ...

**Doctors**
- Bob
- Doris
- Hank
- Greg
- ...

- Elbonia
- Elbonia
- Vegas
- Paris
- ...

- Cancer
- Malaria
- Malaria
- HIV
- ...

Q: How many people went to Elbonia and had Malaria?
Q: How many people went to Elbonia and had Malaria?
Q: How many people went to Elbonia and had Malaria?
Q: How many people went to Elbonia and had Malaria?
Q: How many people went to Elbonia and had Malaria?
A Join query across multiple databases ...

There is no differentially private query which can handle this case ...
Target Problem

Can we provide differentially private Join queries for distributed databases?
Solution 1

Researchers

Airlines
Doris Elbonia
Hank Elbonia
Emil Vegas
Bob Paris
...

Trusted party

Doctors
Bob Cancer
Doris Malaria
Hank Malaria
Greg HIV
...


Solution 1

We do not have such trusted party in practice

Researchers

<table>
<thead>
<tr>
<th>Hank</th>
<th>Elbonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emil</td>
<td>Vegas</td>
</tr>
<tr>
<td>Bob</td>
<td>Paris</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Doris</th>
<th>Malaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hank</td>
<td>Malaria</td>
</tr>
<tr>
<td>Greg</td>
<td>HIV</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Solution 2

Airlines

Researchers

Doctors

SMPC

Doris, Elbonia
Hank, Elbonia
Emil, Vegas
Bob, Paris
...
...

Bob, Cancer
Doris, Malaria
Hank, Malaria
Greg, HIV
...
...
Solution 2

SMPC works but it will take years ...
DJoin

DJoin is a system supporting differentially private JOIN queries across distributed DBs

Q: How many people went to Elbonia and had Malaria?

DJoin’s Differentially Private Query Processor
DJoin

DJoin is a system supporting differentially private **JOIN** queries across distributed DBs.

Q: How many people went to Elbonia and had Malaria?

About 302

DJoin’s Differentially Private Query Processor
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE WHERE Destination="Elbonia" AND Diagnosis="Malaria"

Who went to Elbonia?

<table>
<thead>
<tr>
<th>Doris</th>
<th>Elbonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hank</td>
<td>Elbonia</td>
</tr>
<tr>
<td>Emil</td>
<td>Vegas</td>
</tr>
<tr>
<td>Bob</td>
<td>Paris</td>
</tr>
<tr>
<td>Charlie</td>
<td>Elbonia</td>
</tr>
</tbody>
</table>

Who had Malaria?

<table>
<thead>
<tr>
<th>Bob</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doris</td>
<td>Malaria</td>
</tr>
<tr>
<td>Hank</td>
<td>Malaria</td>
</tr>
<tr>
<td>Greg</td>
<td>HIV</td>
</tr>
<tr>
<td>Alice</td>
<td>Malaria</td>
</tr>
</tbody>
</table>
SELECT COUNT(*) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"

Who went to Elbonia?
Who had Malaria?

Step 2

Doris Elbonia
Hank Elbonia
Emil Vegas
Bob Paris
Charlie Elbonia

Doris Hank
Charlie

Bob Cancer
Doris Malaria
Hank Malaria
Greg HIV
Alice Malaria

\(|= 2 \pm 1|
The main challenge is how to do set intersection cardinality while preserving differential privacy.

SELECT COUNT(*) FROM HOSPITAL 
JOIN AIRLINE 
WHERE Destination="Elbonia" AND Diagnosis="Malaria"
Solution: BN-PSI-CA

❖ PSI-CA is a protocol named private set-intersection cardinality.
❖ PSI-CA allows a group of k parties with multisets $S_1...S_k$ to privately compute $|\bigcap S_i|$, i.e., the exact number of elements they have in common.
Solution: BN-PSI-CA

❖ PSI-CA is a protocol named private set-intersection cardinality.
❖ PSI-CA allows a group of k parties with multisets $S_1\ldots S_k$ to privately compute $|\bigcap S_i|$, i.e., the exact number of elements they have in common.

BN-PSI-CA adds noise to the set intersection cardinality
PSI-CA [EuroCrypt’04]

- PSI-CA is a protocol named private set-intersection cardinality.
- PSI-CA allows a group of $k$ parties with multisets $S_1 \ldots S_k$ to privately compute $|\bigcap S_i|$, i.e., the exact number of elements they have in common.
PSI-CA [EuroCrypt’04]

- PSI-CA is a protocol named private set-intersection cardinality.
- PSI-CA allows a group of $k$ parties with multisets $S_1 \ldots S_k$ to privately compute $|\cap S_i|$, i.e., the exact number of elements they have in common.
PSI-CA [EuroCrypt’04]

- PSI-CA is a protocol named private set-intersection cardinality.
- PSI-CA allows a group of k parties with multisets $S_1...S_k$ to privately compute $|\bigcap S_i|$, i.e., the **exact** number of elements they have in common.
PSI-CA [EuroCrypt’04]

- PSI-CA is a protocol named private set-intersection cardinality.
- PSI-CA allows a group of \( k \) parties with multisets \( S_1 \ldots S_k \) to privately compute \( |\cap S_i| \), i.e., the exact number of elements they have in common.
PSI-CA [EuroCrypt’04]

- PSI-CA is a protocol named private set-intersection cardinality.
- PSI-CA allows a group of k parties with multisets $S_1...S_k$ to privately compute $|\bigcap S_i|$, i.e., the exact number of elements they have in common.
• The airline and hospital have set A and B respectively and the airline wants to jointly compute \(|A \cap B|\).
PSI-CA [EuroCrypt’04]

\[ P = (X-12)(X-5)(X-4) = x^3 - 21x^2 + 128x - 240 \]

- The airline and hospital has set A and B respectively and airline wants to jointly compute \( |A \cap B| \).
- The airline makes a polynomial P whose roots are the elements of A.
PSI-CA [EuroCrypt’04]

- The airline and hospital have set A and B respectively, and the airline wants to jointly compute $|A \cap B|$.
- The airline makes a polynomial $P$ whose roots are the elements of A.
- The airline encrypts the coefficients of $P$ and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.

$$P = (X-12)(X-5)(X-4) = x^3 - 21x^2 + 128x - 240$$

$\{E(1), E(-21), E(128), E(-240)\}$
The airline and hospital has set A and B respectively and airline wants to jointly compute $|A \cap B|$.

The airline makes a polynomial $P$ whose roots are the elements of A.

The airline encrypts the coefficients of $P$ and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.

$$P = (X-12)(X-5)(X-4) = x^3 -21x^2 + 128x - 240$$

Set A: \{12, 5, 4\}

Set B: \{13, 4, 2, 6\}

PSI-CA [EuroCrypt’04]
The airline and hospital has set A and B respectively and airline wants to jointly compute $|A \cap B|$.

The airline makes a polynomial $P$ whose roots are the elements of A.

The airline encrypts the coefficients of $P$ and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.

The doctor evaluates $P(B_i)$ for each element in B.
The airline and hospital has set A and B respectively and airline wants to jointly compute $|A \cap B|$.

The airline makes a polynomial $P$ whose roots are the elements of A.

The airline encrypts the coefficients of $P$ and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.

The doctor evaluates $P(B_i)$ for each element in B.
The airline and hospital has set A and B respectively and airline wants to jointly compute $|A \cap B|$. The airline makes a polynomial $P$ whose roots are the elements of A. The airline encrypts the coefficients of $P$ and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor. The doctor evaluates $P(B_i)$ for each element in B.
The airline and hospital has set A and B respectively and airline wants to jointly compute \(|A \cap B|\).

- The airline makes a polynomial \(P\) whose roots are the elements of A.
- The airline encrypts the coefficients of \(P\) and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.
- The doctor evaluates \(P(B_i)\) for each element in B.
- The doctor returns the encrypted evaluations to the airline.
The airline and hospital has set A and B respectively and airline wants to jointly compute \(|A \cap B|\).

- The airline makes a polynomial \(P\) whose roots are the elements of A.
- The airline encrypts the coefficients of \(P\) and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.
- The doctor evaluates \(P(B_i)\) for each element in B.
- The doctor returns the encrypted evaluations to the airline.
- The airline decrypts it and counts the number of zeroes.

\[(X-12)(X-5)(X-4) = x^3 - 21x^2 + 128x - 240\]

\[\{E(1), E(-21), E(128), E(-240)\}\]

\[\{E(152), E(0), E(6612), E(152)\}\]
The airline and hospital has set A and B respectively and airline wants to jointly compute \(|A \cap B|\).

- The airline makes a polynomial \(P\) whose roots are the elements of A.
- The airline encrypts the coefficients of \(P\) and sends them to the doctor. Note that the airline sends homomorphic encryptions of the coefficients to the doctor.
- The doctor evaluates \(P(B_i)\) for each element in B.
- The doctor returns the encrypted evaluations to the airline.
- The airline decrypts it and counts the number of zeroes.
**PSI-CA is not differentially private**

\[(X-12)(X-5)(X-4) = x^3 - 21x^2 + 128x - 240\]

\[\{E(1), E(-21), E(128), E(-240)\}\]

\[\{E(152), E(0), E(6612), E(152)\}\]

\[\{152, 0, 6612, 152\}\]

- This protocol is **not differentially private** since:
  - The first party can learn the exact size of the intersection.
  - Both parties can learn the exact size of the other database.
PSI-CA is not differentially private

\[(X-12)(X-5)(X-4) = x^3 - 21x^2 + 128x - 240\]

BN-PSI-CA is built by extending PSI-CA to make it differentially private ...

• This protocol is not differentially private since:
  • The first party can learn the exact size of the intersection.
  • Both parties can learn the exact size of the other database.
BN-PSI-CA
BN-PSI-CA

\[(X-12)(X-5)(X-4)(X-975)(X-7255) = x^5 - 36x^4 + 497x^3 + 3294x^2 + 10512x - 12960\]
$\{E(1), E(36), E(497), E(3294), E(10512), E(-12960)\}$

$$(X-12)(X-5)(X-4)(X-9125)(X-7255) = x^5-36x^4+497x^3+3294x^2+10512x-12960$$
BN-PSI-CA

\[(X-12)(X-5)(X-4)(X-9125)(X-7255) = x^5 - 36x^4 + 497x^3 + 3294x^2 + 10512x - 12960\]

\{E(1), E(36), E(497), E(3294), E(10512), E(-12960)\}
\[(X-12)(X-5)(X-4)(X-9125)(X-7255) = x^5 - 36x^4 + 497x^3 + 3294x^2 + 10512x - 12960\]

\[\{E(1), E(36), E(497), E(3294), E(10512), E(-12960)\}\]

\[\{E(6752), E(0), E(4612), E(7452), E(0), E(0), E(242), E(125), E(525)\}\]

\[C=5, N=2\]
BN-PSI-CA

\[(X-12)(X-5)(X-4)(X-9125)(X-7255) = x^5 - 36x^4 + 497x^3 + 3294x^2 + 10512x - 12960\]

\[\{E(1), E(36), E(497), E(3294), E(10512), E(-12960)\}\]

\[\{E(6752), E(0), E(4612), E(7452), E(0), E(0), E(242), E(125), E(525)\}\]

C=5, N=2
(X-12)(X-5)(X-4)(X-9125)(X-7255) = x^5-36x^4+497x^3+3294x^2+10512x-12960

\{E(1), E(36), E(497), E(3294), E(10512), E(-12960)\}

\{E(6752), E(0), E(4612), E(7452), E(0), E(0), E(242), E(125), E(525)\}

Blinded result is: 3

C=5, N=2

12
5
4
= \{6752, 0, 4612, 7452, 0, 0, 242, 125, 525\}
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"

Who went to Elbonia?

Who had Malaria?

Step 2

Doris  Elbonia
Hank  Elbonia
Emil  Vegas
Bob  Paris
Charlie  Elbonia
Doris  Elbonia
Hank
Charlie

Bob  Cancer
Doris  Malaria
Hank  Malaria
Greg  HIV
Alice  Malaria

| Doris  |
| Hank  |
| Charlie |

∩

| Doris  |
| Hank  |
| Alice |

= 2
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"

Who went to Elbonia?
Who had Malaria?

Step 2

<table>
<thead>
<tr>
<th>Doris</th>
<th>Elbonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hank</td>
<td></td>
</tr>
<tr>
<td>Emil</td>
<td>Vegas</td>
</tr>
<tr>
<td>Bob</td>
<td>Paris</td>
</tr>
<tr>
<td>Charlie</td>
<td>Elbonia</td>
</tr>
<tr>
<td>Doris</td>
<td></td>
</tr>
<tr>
<td>Hank</td>
<td></td>
</tr>
<tr>
<td>Charlie</td>
<td></td>
</tr>
<tr>
<td>Bob</td>
<td>Cancer</td>
</tr>
<tr>
<td>Doris</td>
<td>Malaria</td>
</tr>
<tr>
<td>Hank</td>
<td>Malaria</td>
</tr>
<tr>
<td>Greg</td>
<td>HIV</td>
</tr>
<tr>
<td>Alice</td>
<td>Malaria</td>
</tr>
</tbody>
</table>

\[
\text{Step 2} \quad \left| \begin{array}{c}
Doris \\
Hank \\
Charlie
\end{array} \right| \cap \left| \begin{array}{c}
Doris \\
Hank \\
Alice
\end{array} \right| = 2 \pm 1
\]
SELECT COUNT(X) FROM HOSPITAL JOIN AIRLINE
WHERE Destination="Elbonia" AND Diagnosis="Malaria"

Step 2

BN-PSI-CA is sufficient to answer queries that require only one intersection operation

<table>
<thead>
<tr>
<th>Who went to Elbonia?</th>
</tr>
</thead>
</table>
| Doris
| Hank
| Charlie |

<table>
<thead>
<tr>
<th>Who had Malaria?</th>
</tr>
</thead>
</table>
| Doris
| Bob
| Emil
| Charlie |

\[ = 2 \pm 1 \]
Some queries need more than one BN-PSI-CA, e.g.,

```
SELECT |X.a| FROM X,Y WHERE X.a=Y.a OR X.b=Y.b
```
Some queries need more than one BN-PSI-CA, e.g.,

```
SELECT |X.a| FROM X, Y WHERE X.a = Y.a OR X.b = Y.b
```

Need to compute $|X.a \cap Y.a| + |X.b \cap Y.b| - |X.ab \cap Y.ab|$
Some queries need more than one BN-PSI-CA, e.g.,

```
SELECT |X.a| FROM X,Y WHERE X.a=Y.a OR X.b=Y.b
```

Need to compute

\[ |X.a \cap Y.a| + |X.b \cap Y.b| - |X.ab \cap Y.ab| \]

\[ |X.a \cap Y.a| + \underline{+} + |X.b \cap Y.b| + \underline{+} - |X.ab \cap Y.ab| + \underline{+} \]

\[ = |X.a \cap Y.a| + |X.b \cap Y.b| - |X.ab \cap Y.ab| + \underline{+} \]
Some queries need more than one BN-PSI-CA, e.g.,

```
SELECT |X.a| FROM X,Y WHERE X.a=Y.a OR X.b=Y.b
```

Need to compute

```
|X.a\cap Y.a| + |X.b\cap Y.b| - |X.ab\cap Y.ab|
```

Denoise-Combine-Renoise
Some queries need more than one BN-PSI-CA, e.g.,

\[
\text{SELECT } |X.a| \text{ FROM } X,Y \text{ WHERE } X.a=Y.a \text{ OR } X.b=Y.b
\]

Need to compute \(|X.a \cap Y.a| + |X.b \cap Y.b| - |X.ab \cap Y.ab|\)

\[
|X.a \cap Y.a| + \_ + |X.b \cap Y.b| + \_ - |X.ab \cap Y.ab| + \_
\]

\[
= |X.a \cap Y.a| + |X.b \cap Y.b| - |X.ab \cap Y.ab| + \_\]
A Case Study
SELECT NOISY COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis='malaria'
SELECT NOISY COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis= 'malaria'

Query execution with a centralized database.
SELECT NOisy COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis = 'malaria'

Query execution with a centralized database.

Differentially private query execution: with only local operations, set intersections and DCR.
SELECT NOISY COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis='malaria'

Query execution with a centralized database.

We can’t do this!

Differentially private query execution: with only local operations, set intersections and DCR.
SELECT NOISY COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis= 'malaria'

Query execution with a centralized database.

Differentially private query execution: with only local operations, set intersections and DCR.
SELECT NOISY COUNT(A.ssn) FROM A,B WHERE (A.ssn=B.ssn OR A.id=B.id) AND A.diagnosis='malaria'

Query execution with a centralized database.

Differentially private query execution: with only local operations, set intersections and DCR.
Lecture Roadmap

• Differential Privacy
• Case Study: DJoin
• Midterm Review
Midterm Exam

- Location: WTS A51 (the same room as our class)
- Time: 1-2:15pm Oct 16
- Closed-book
- Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
Midterm Exam

- Location: WTS A51 (the same room as our class)
- Time: 1-2:15pm Oct 16
- Closed-book
- Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
UseNet and Gossip

- Understanding basic UseNet format (RFC 1036)
- Gossip protocol
  - Rumor-mongering
  - Anti-entropy
  - Security problem
  - How to create unique ID and why?
UseNet Format

From: jerry@eagle.ATT.COM (Jerry Schwarz)
Path: cbosgd!mhuxj!mhuxt!eagle!jerry
Newsgroups: news.announce
Subject: Usenet Etiquette -- Please Read
Message-ID: <642@eagle.ATT.COM>
Date: Fri, 19 Nov 82 16:14:55 GMT
Followup-To: news.misc
Expires: Sat, 1 Jan 83 00:00:00 -0500
Organization: AT&T Bell Laboratories, Murray Hill

The body of the message comes here, after a blank line.

- If Alice reads a message locally on her machine containing the above headers, what can you infer is the name of Alice’s machine?
### UseNet Format

<table>
<thead>
<tr>
<th>From:</th>
<th><a href="mailto:jerry@eagle.ATT.COM">jerry@eagle.ATT.COM</a> (Jerry Schwarz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path:</td>
<td>cbosgd!mhuxj!mhuxt!eagle!jerry</td>
</tr>
<tr>
<td>Newsgroups:</td>
<td>news.announce</td>
</tr>
<tr>
<td>Subject:</td>
<td>Usenet Etiquette -- Please Read</td>
</tr>
<tr>
<td>Message-ID:</td>
<td><a href="mailto:642@eagle.ATT.COM">642@eagle.ATT.COM</a></td>
</tr>
<tr>
<td>Date:</td>
<td>Fri, 19 Nov 82 16:14:55 GMT</td>
</tr>
<tr>
<td>Followup-To:</td>
<td>news.misc</td>
</tr>
<tr>
<td>Expires:</td>
<td>Sat, 1 Jan 83 00:00:00 -0500</td>
</tr>
<tr>
<td>Organization:</td>
<td>AT&amp;T Bell Laboratories, Murray Hill</td>
</tr>
</tbody>
</table>

The body of the message comes here, after a blank line.

- If Alice reads a message locally on her machine containing the above headers, what can you infer is the name of Alice’s machine?
  
  **cbosgd**
UseNet Format

From: jerry@eagle.ATT.COM (Jerry Schwarz)
Path: cbosgd!mhuxj!mhuxt!eagle!jerry
Newsgroups: news.announce
Subject: Usenet Etiquette -- Please Read
Message-ID: <642@eagle.ATT.COM>
Date: Fri, 19 Nov 82 16:14:55 GMT
Followup-To: news.misc
Expires: Sat, 1 Jan 83 00:00:00 -0500
Organization: AT&T Bell Laboratories, Murray Hill

The body of the message comes here, after a blank line.

- Alice sends a private message to Jerry’s UseNet post, but Jerry did not receive it. After several days, Alice sees some new posts from Jerry with a Path: header line of ‘cbosgd!mhuxj!ucbvax!eagle!jerry’
- What do you expect happened to Alice’s original message to Jerry?
UseNet Format

From: jerry@eagle.ATT.COM (Jerry Schwarz)
Path: cbosgd!mhuxj!mhuxt!eagle!jerry
Newsgroups: news.announce
Subject: Usenet Etiquette -- Please Read
Message-ID: <642@eagle.ATT.COM>
Date: Fri, 19 Nov 82 16:14:55 GMT
Followup-To: news.misc
Expires: Sat, 1 Jan 83 00:00:00 -0500
Organization: AT&T Bell Laboratories, Murray Hill

The body of the message comes here, after a blank line.

• Alice sends a private message to Jerry’s UseNet post, but Jerry did not receive it. After several days, Alice sees some new posts from Jerry with a Path: header line of `cbosgd!mhuxj!ucbvax!eagle!jerry’
• What do you expect happened to Alice’s original message to Jerry?
Midterm Exam

- Location: WTS A51 (the same room as our class)
- Time: 1-2:15pm Oct 16
- Closed-book
- Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
Unstructured Search

- Flooding protocol
- Random walk

Running algorithm A. What's the traffic?
DHT and Chord

- Understand how Chord works
- The complexity of Chord (e.g., space complexity)
- Potential practical issues of Chord
DHT and Chord

- Understand how Chord works
- The complexity of Chord (e.g., space complexity)
- Potential practical issues of Chord
DHT and Chord

- Understand how Chord works
- The complexity of Chord (e.g., space complexity)
- Potential practical issues of Chord

Why I only need at most 160 steps to find an object in Chord?
Midterm Exam

- Location: WTS A51 (the same room as our class)
- Time: 1-2:15pm Oct 16
- Closed-book
- Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
Content Sharing and Reputation

• Global trust model: PageRank
• Personalized reputation model
  - Peer-based reputation systems, e.g., EigenTrust
  - Object-based reputation systems, e.g., Credence

We will provide equation for reputation computation
Content Sharing and Reputation

A: -1
B: -1
C: +1
D: -1

A: +1
B: +1
C: -1

A: +1
B: +1
C: -1

A: -1
B: -1
C: +1
D: -1

A: +1
D: -1

C: +1
G: -1
Midterm Exam

• Location: WTS A51 (the same room as our class)
• Time: 1-2:15pm Oct 16
• Closed-book
• Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
Attacks and Defenses

• We do not test DSybil
• Understand sybil attacks and defense
  - How to launch sybil attacks to out-vote reputation system
  - How to use SybilGuard to detect sybil identities
 Attacks and Defenses

- We do not test DSybil
- Understand sybil attacks and defense
  - How to launch sybil attacks to out-vote reputation system
  - How to use SybilGuard to detect sybil identities
Midterm Exam

• Location: WTS A51 (the same room as our class)
• Time: 1-2:15pm Oct 16
• Closed-book

• Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
Firewalls

• Which machine is my internal machine? Why?

- set block src-ip [ip-address] dst-ip [ip-address] protocol [name]
- I do not want to receive ICMP packet from 61.172.201.180
- set block src-ip 61.172.201.180 dst-ip 192.168.2.2 protocol icmp
Firewalls

- Which machine is my internal machine? Why?
- set block src-ip [ip-address] dst-ip [ip-address] protocol [name]
  - I do not want to receive ICMP packet from 61.172.201.180
Firewalls

- Which machine is my internal machine? Why?
- `set block src-ip [ip-address] dst-ip [ip-address] protocol [name]`
  - I do not want to receive ICMP packet from 61.172.201.180
  - `set block src-ip 61.172.201.180 dst-ip 192.168.2.2 protocol icmp`
NATs

- NATs have four different categories:
  - Full cone NAT
  - A restricted cone NAT
  - A port restricted cone NAT
  - A symmetric NAT
192.168.2.2

Host A

S=192.168.2.2:4445
D=1.1.1.5:7777

1.1.1.4

Restricted cone
NAT

1.1.1.5

Host B

1.1.1.6

Host C
Mapping:
192.168.2.2:4445  1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.5:7777
Mapping:
192.168.2.2:4445 ↔ 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

192.168.2.2

S=192.168.2.2:4445
D=1.1.1.5:7777

S=1.1.1.4:10100
D=1.1.1.5:7777

Restricted Cone
NAT

Host A

Host B

Host C

1.1.1.4

1.1.1.5

1.1.1.6
Restricted Cone

Mapping:
192.168.2.2:4445 ↔ 1.1.1.4:10100
Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

Host A

192.168.2.2

1.1.1.4

Restricted cone NAT

Host B

1.1.1.5

1.1.1.6

Host C

S=192.168.2.2:4445
D=1.1.1.5:7777

S=1.1.1.4:10100
D=1.1.1.5:7777

S=1.1.1.5:4321
D=1.1.1.4:10100
Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

Host A
192.168.2.2

Host B
1.1.1.4

Host C
1.1.1.5
1.1.1.6

Restricted cone
NAT

S=192.168.2.2:4445
D=1.1.1.5:7777
S=1.1.1.5:4321
D=192.168.2.2:4445

S=1.1.1.4:10100
D=1.1.1.5:7777
S=1.1.1.5:4321
D=1.1.1.4:10100
Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100
Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

Restricted Cone

192.168.2.2

Host A

1.1.1.4

Restricted cone
NAT

1.1.1.5

Host B

1.1.1.6

Host C

S=192.168.2.2:4445
D=1.1.1.5:7777
S=1.1.1.4:10100
D=1.1.1.5:7777
S=1.1.1.5:4321
D=1.1.1.4:10100
S=1.1.1.4:10100
D=1.1.1.4:10100
S=192.168.2.2:4445
D=1.1.1.6:1234
S=1.1.1.6:1234
D=1.1.1.4:10100
Mapping: 192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100
Mapping:
192.168.2.2:4445 ➔ 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

192.168.2.2       1.1.1.4       1.1.1.5       1.1.1.6

Host A       NAT       Host B       Host C

S=192.168.2.2:4445
D=1.1.1.5:7777
S=1.1.1.5:4321
D=192.168.2.2:4445
S=1.1.1.4:10100
D=1.1.1.5:7777
S=1.1.1.5:4321
D=1.1.1.4:10100
S=1.1.1.6:1234
D=1.1.1.4:10100
Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100

Restricted Cone
Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100
Allow 1.1.1.6 to 1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.5:7777
S=1.1.1.5:4321
D=192.168.2.2:4445
S=1.1.1.5:4321
D=1.1.1.4:10100
S=1.1.1.6:1234
D=1.1.1.4:10100
S=192.168.2.2:4445
D=1.1.1.6:7777
S=1.1.1.6:1234
D=1.1.1.4:10100
Restricted Cone

Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100
Allow 1.1.1.6 to 1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.4:10100

S=1.1.1.4:10100
D=1.1.1.5:7777

S=1.1.1.5:4321
D=192.168.2.2:4445

S=1.1.1.5:7777
D=1.1.1.5:4321

S=1.1.1.5:4321
D=1.1.1.4:10100

S=1.1.1.6:1234
D=1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.6:7777

S=1.1.1.4:10100
D=1.1.1.6:7777

S=1.1.1.4:10100
D=1.1.1.6:7777

S=192.168.2.2:4445
D=1.1.1.4:10100
Mapping:
192.168.2.2:4445 <-> 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100
Allow 1.1.1.6 to 1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.5:7777

S=1.1.1.5:4321
D=1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.6:7777

S=1.1.1.6:4321
D=1.1.1.4:10100

S=192.168.2.2:4445
D=1.1.1.4:10100

S=1.1.1.6:1234
D=1.1.1.4:10100

S=1.1.1.6:4321
D=1.1.1.4:10100
Mapping:
192.168.2.2:4445 ➞ 1.1.1.4:10100

Policy:
Allow 1.1.1.5 to 1.1.1.4:10100
Allow 1.1.1.6 to 1.1.1.4:10100
1. Host A:
   Public (155.9.3.1:1900)
   Private (192.168.2.2:4321)
2. Host B:
   Public (132.76.29.7:7777)
   Private (192.168.2.1:9999)

1. Allow 18.181.3.3 to me
2. Allow 132.76.29.7 to me

1. Allow 18.181.3.3 to me
2. Allow 155.9.3.1 to me
Midterm Exam

• Location: WTS A51 (the same room as our class)
• Time: 1-2:15pm Oct 16
• Closed-book

• Lec 2 - 9
  - Lec 2: UseNet and Gossip
  - Lec 3 and 4: P2P and Chord
  - Lec 5: Content sharing and reputation
  - Lec 6: Attacks and Defenses
  - Lec 7: Firewalls and NATs
  - Lec 8 and 9: GFS, MapReduce and BigTable
GFS, MapReduce and BigTable

- Understanding file system basics
- How does MapReduce work
  - Map and Reduce programs if a target is given
  - The input/output of each phase

- I do not want to test BigTable :)}
File Systems

Suppose we have a HDFS. For simplicity, we use a fixed 8KB block size for each block. For each file $f$, we use a blocklist metafile (used to find each block), which is a 8KB file containing hash values as the identifier of each block of $f$. Assume the size of a hash value is 32 byte. What is the maximum size of a file in this HDFS?
File Systems

Suppose we have a HDFS. For simplicity, we use a fixed 8KB block size for each block. For each file \( f \), we use a blocklist metafile (used to find each block), which is a 8KB file containing hash values as the identifier of each block of \( f \). Assume the size of a hash value is 32 byte. What is the maximum size of a file in this HDFS?

\[
(8\text{KB}/32)*8\text{KB} = 2\text{MB}
\]
What are Map and Reducer programs?

MP: 75
CG: 72
OR: 72
Good luck!