Privacy Preserving Medical Data Analytics using Secure Multi Party Computation.
An End-To-End Use Case.

Athanasios Giannopoulos, Dimitris Mouris

M.Sc. thesis

Department of Informatics & Telecommunications
Computing Systems: Software & Hardware

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Supervisors: Yannis Ioannidis, Minos Garofalakis, Omiros Metaxas

September 2018

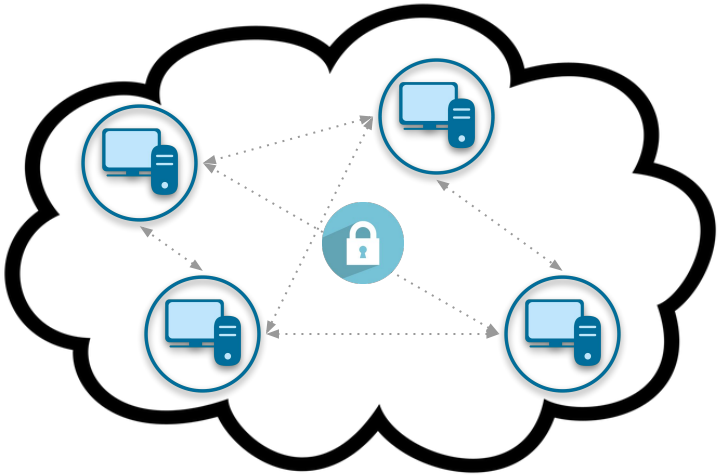
Privacy Preserving Medical Data Analytics using Secure Multi Party Computation. An End-To-End Use Case.

Privacy Preserving Medical Data Analytics using Secure Multi Party Computation. An End-To-End Use Case.

Privacy Preserving Medical Data Analytics using Secure Multi Party Computation. An End-To-End Use Case.

Privacy Preserving Medical Data Analytics using Secure Multi Party Computation.

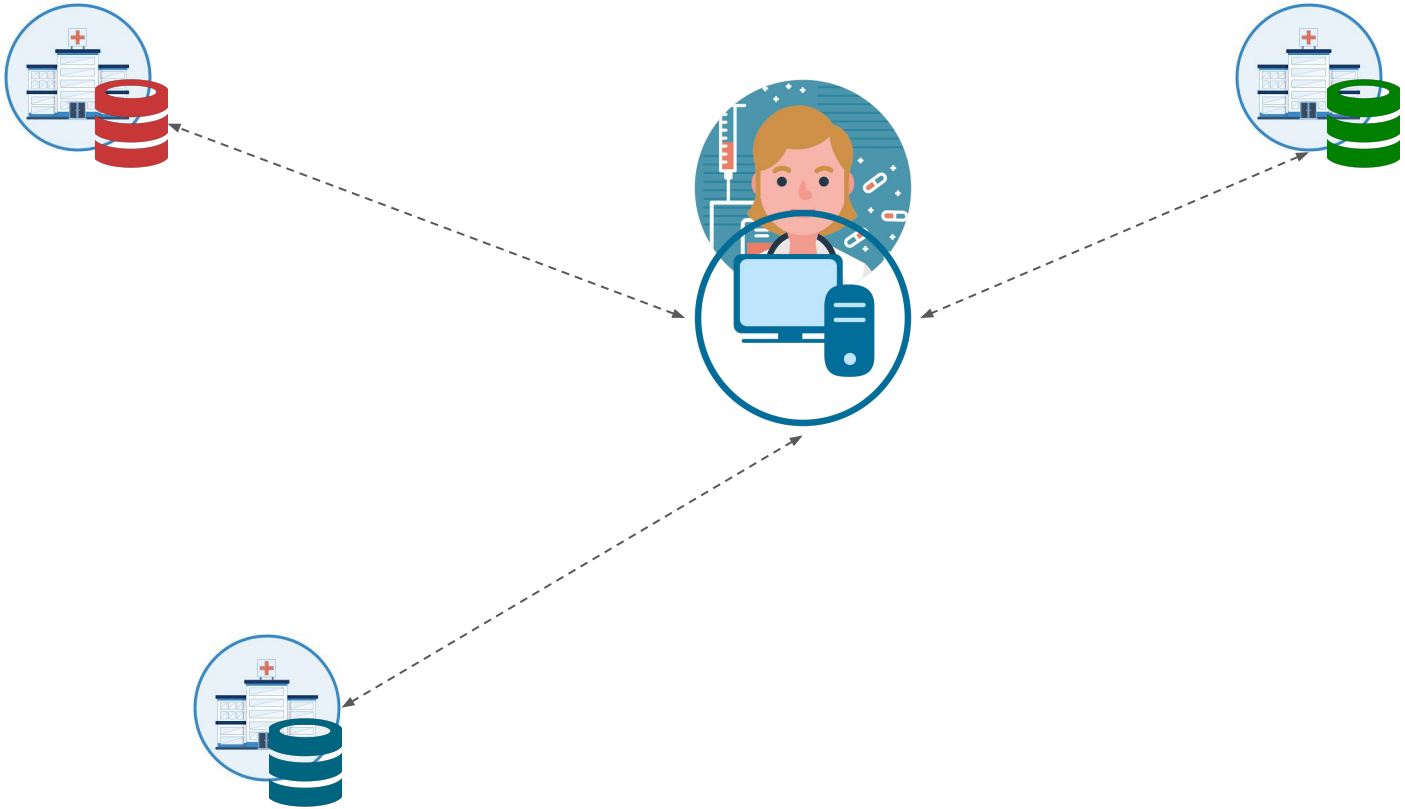
An End-To-End Use Case.

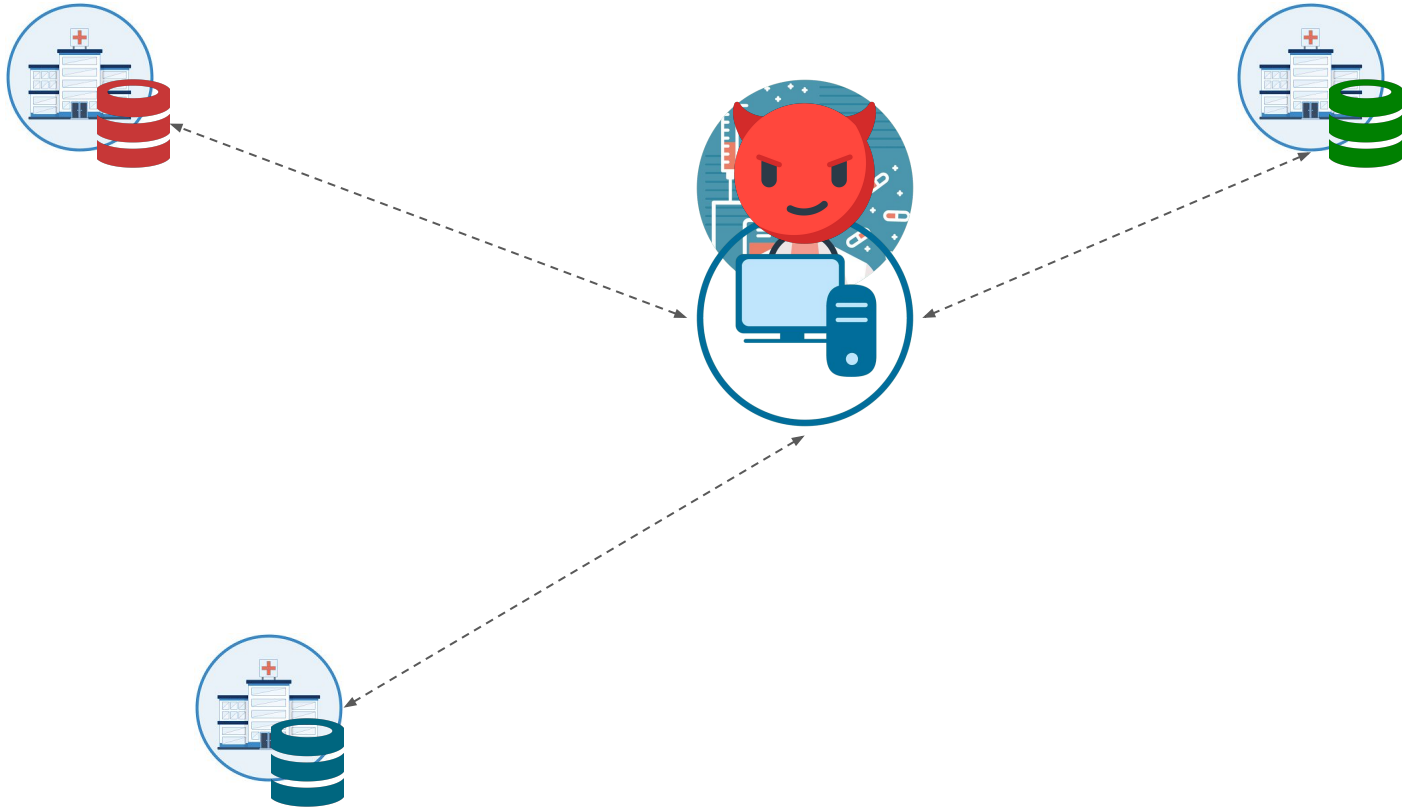


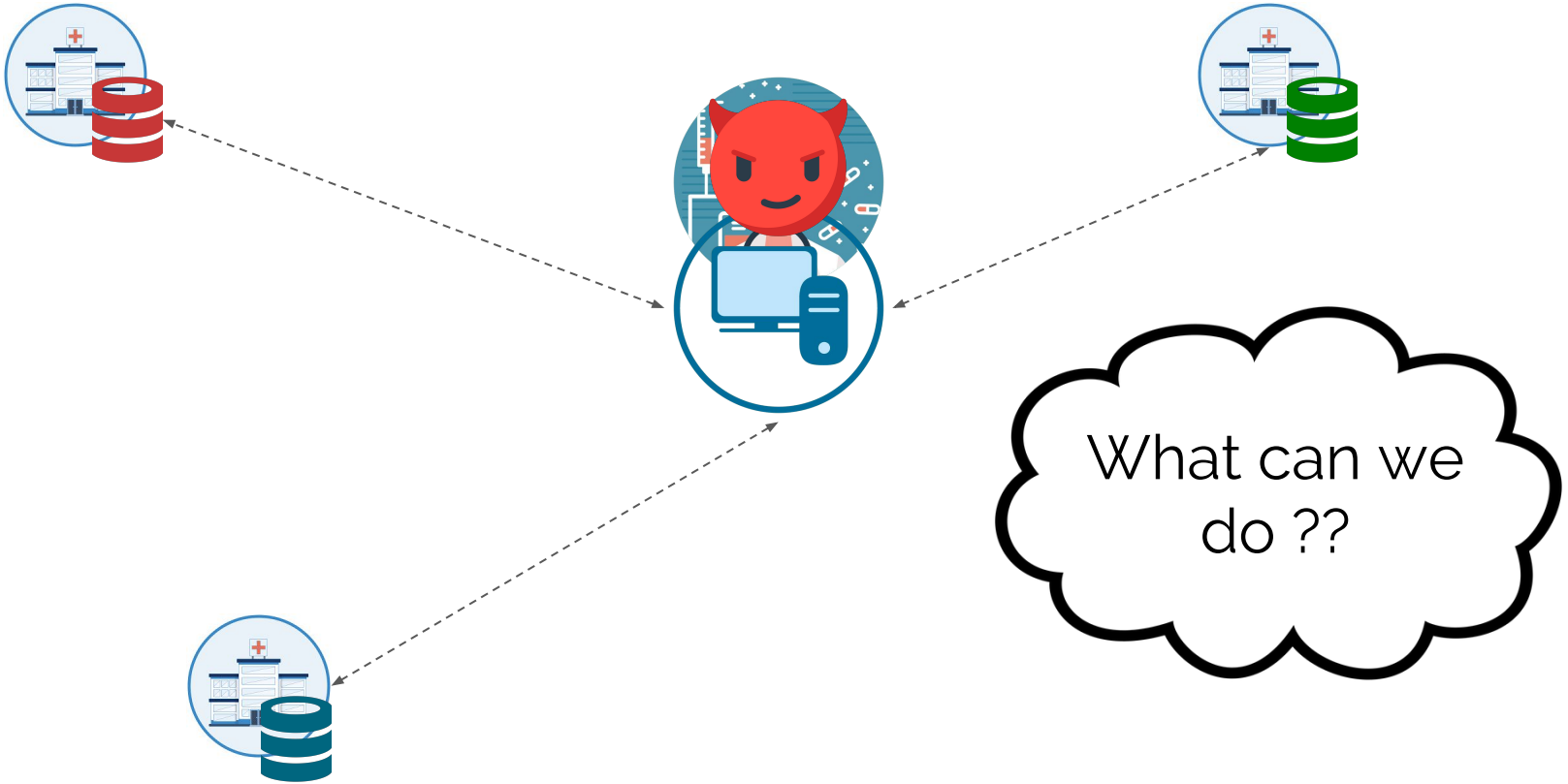
Privacy Preserving Medical Data Analytics using Secure Multi Party Computation.
An End-To-End Use Case.

Problem Formation

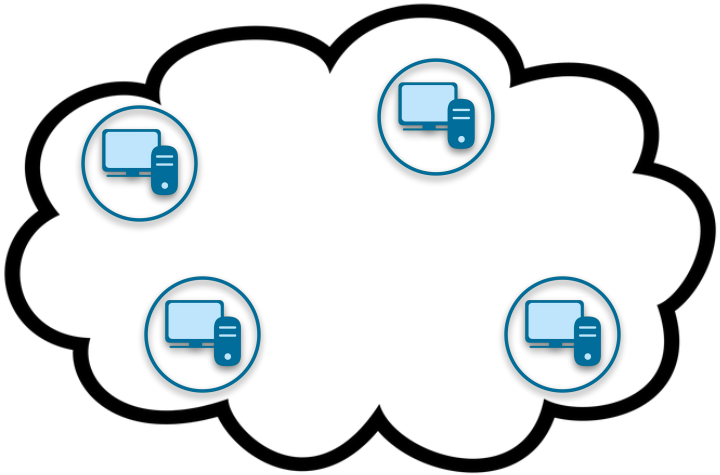

f(, ,)



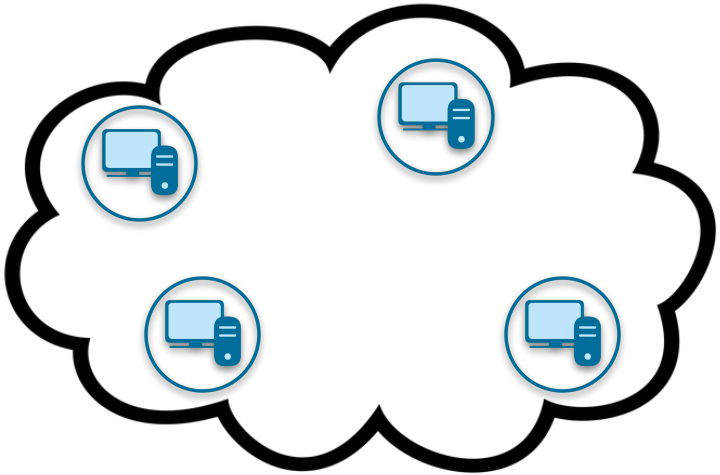



What can we do ??

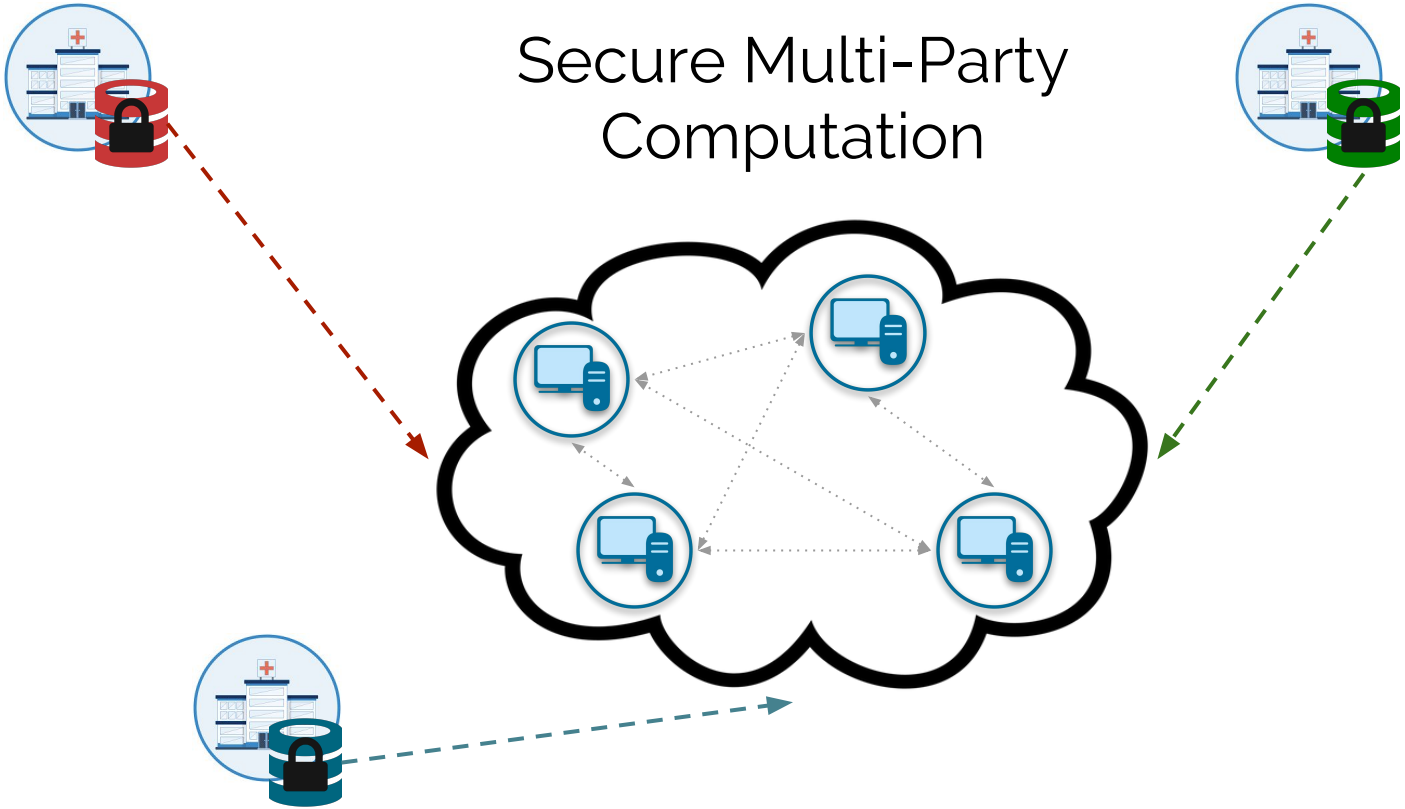
Secure Multi-Party Computation



Secure Multi-Party Computation

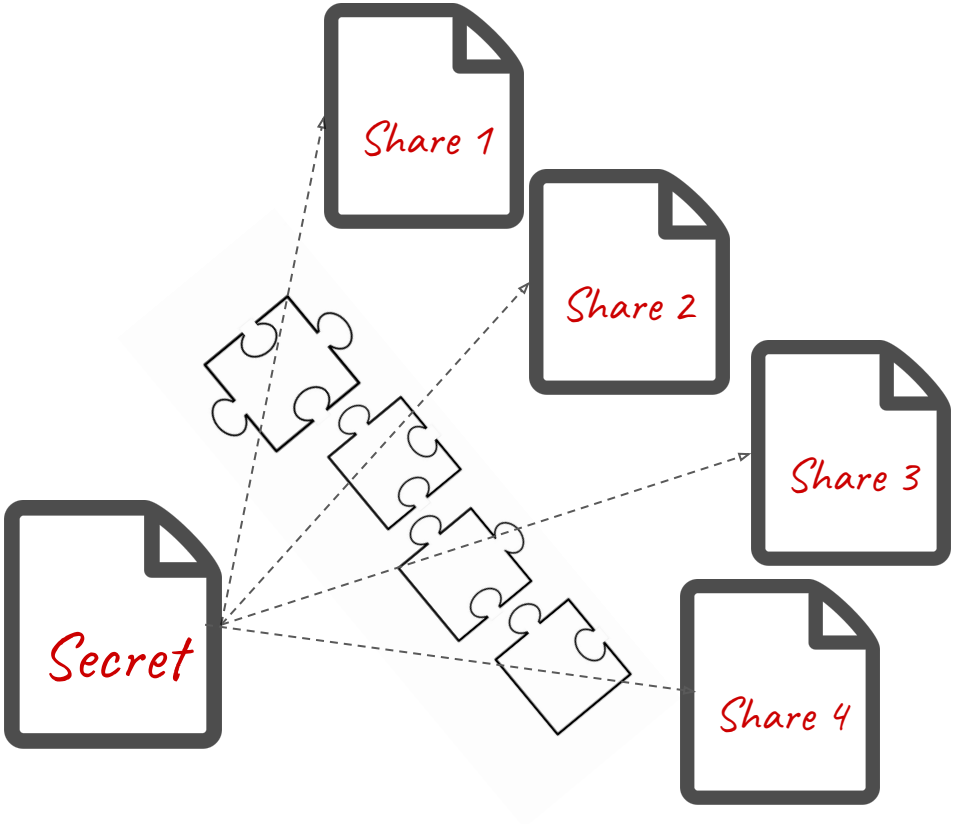


Secure Multi-Party Computation

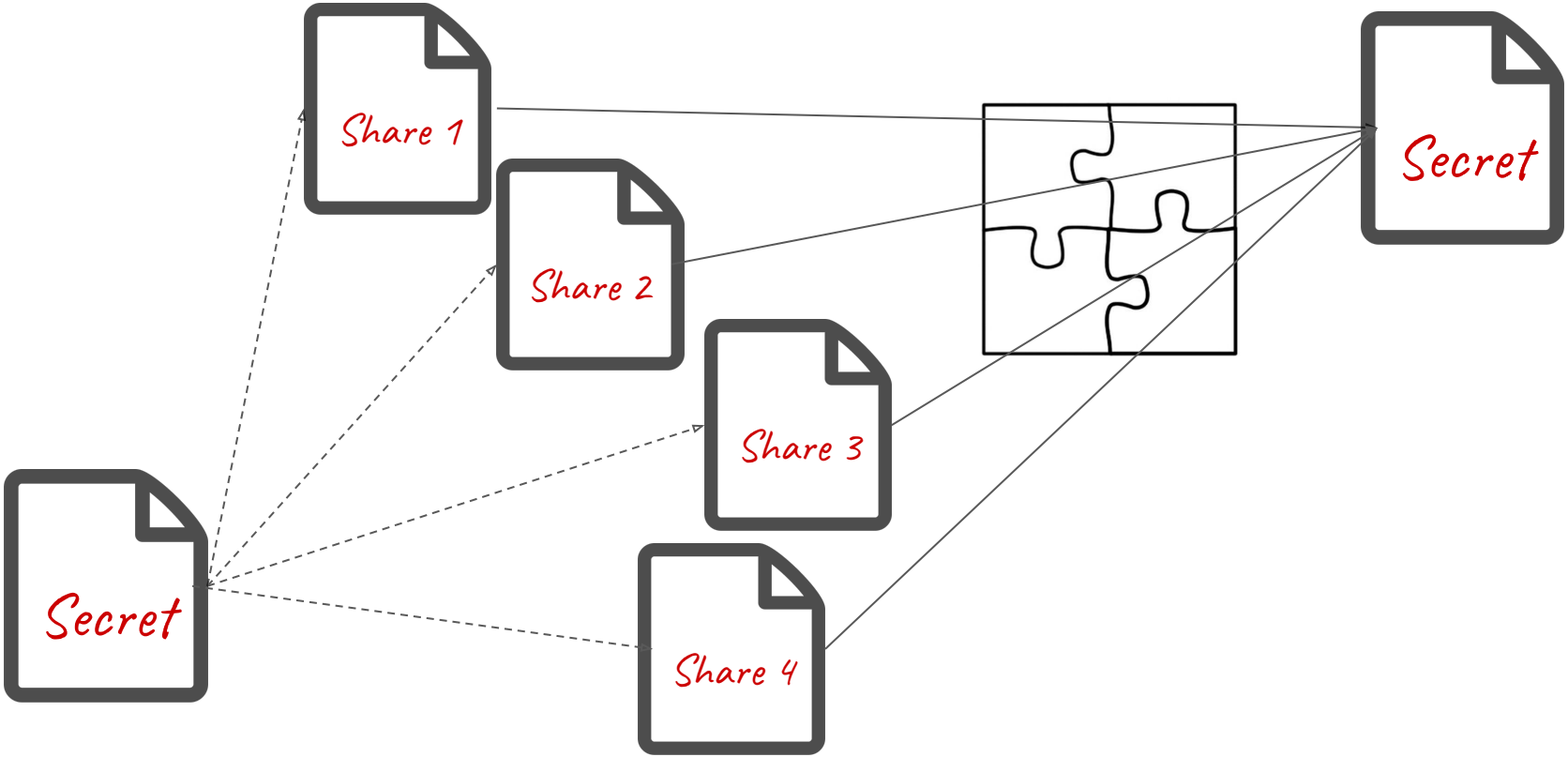


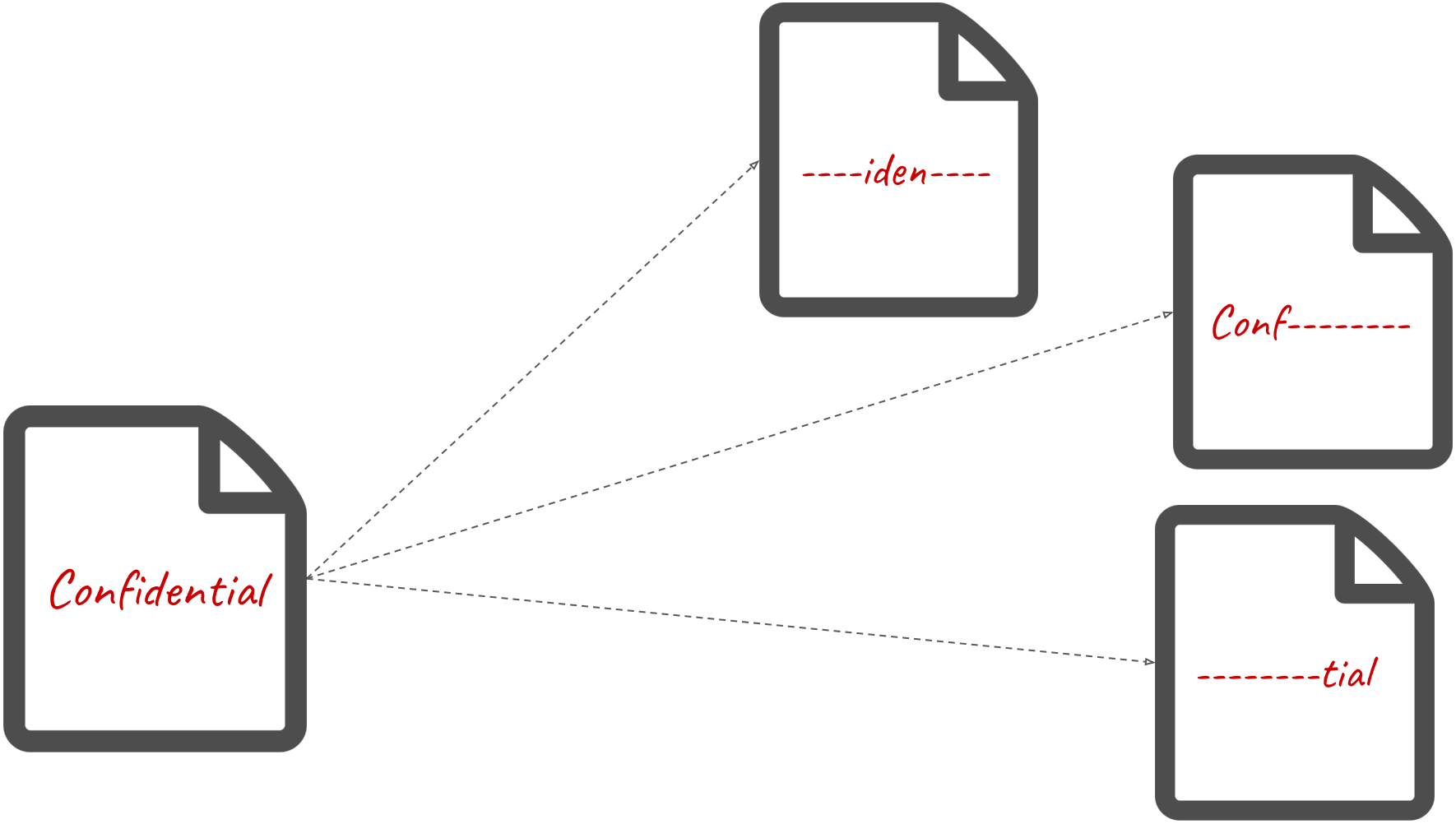
Preliminaries

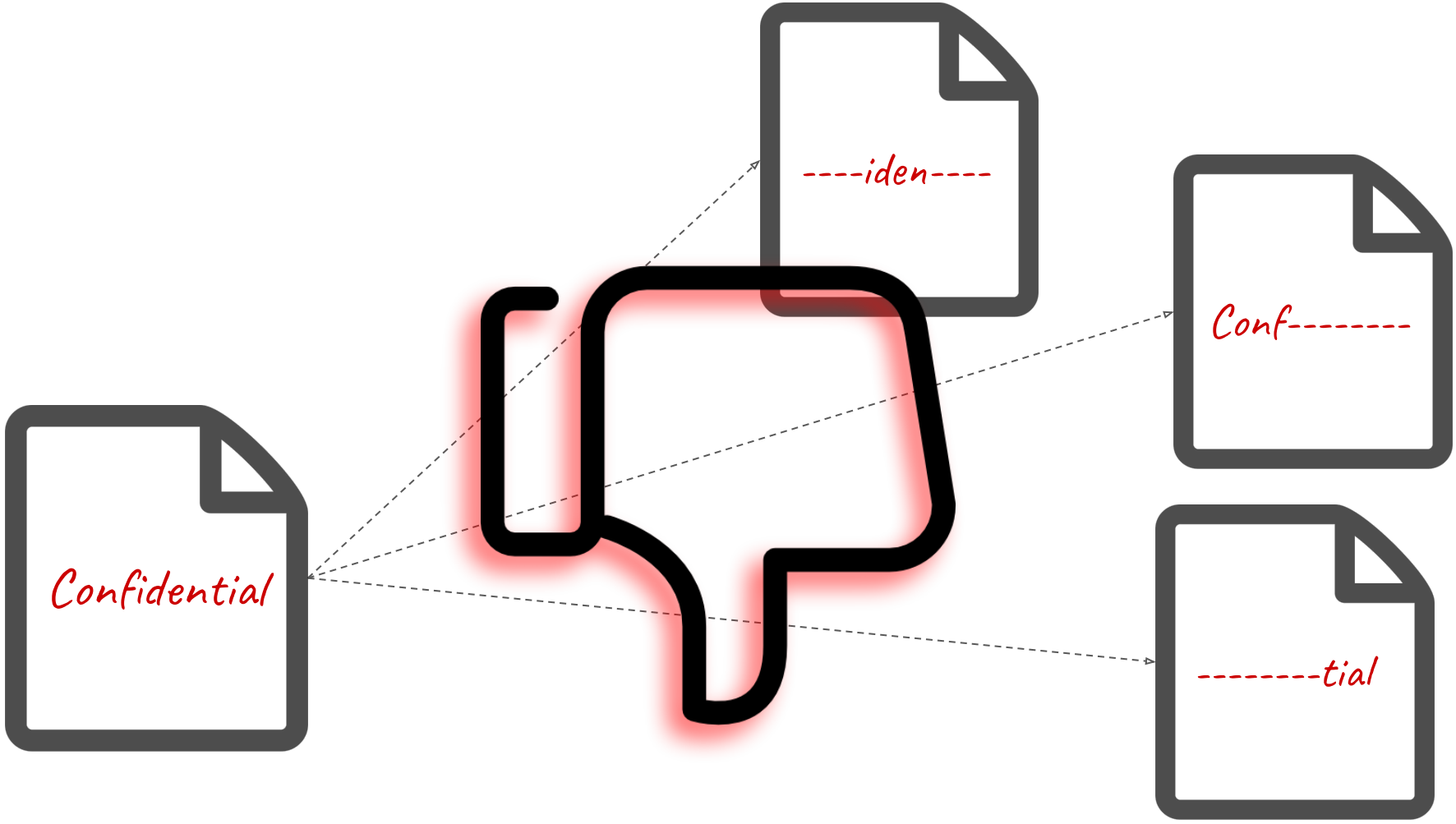
Secret Sharing

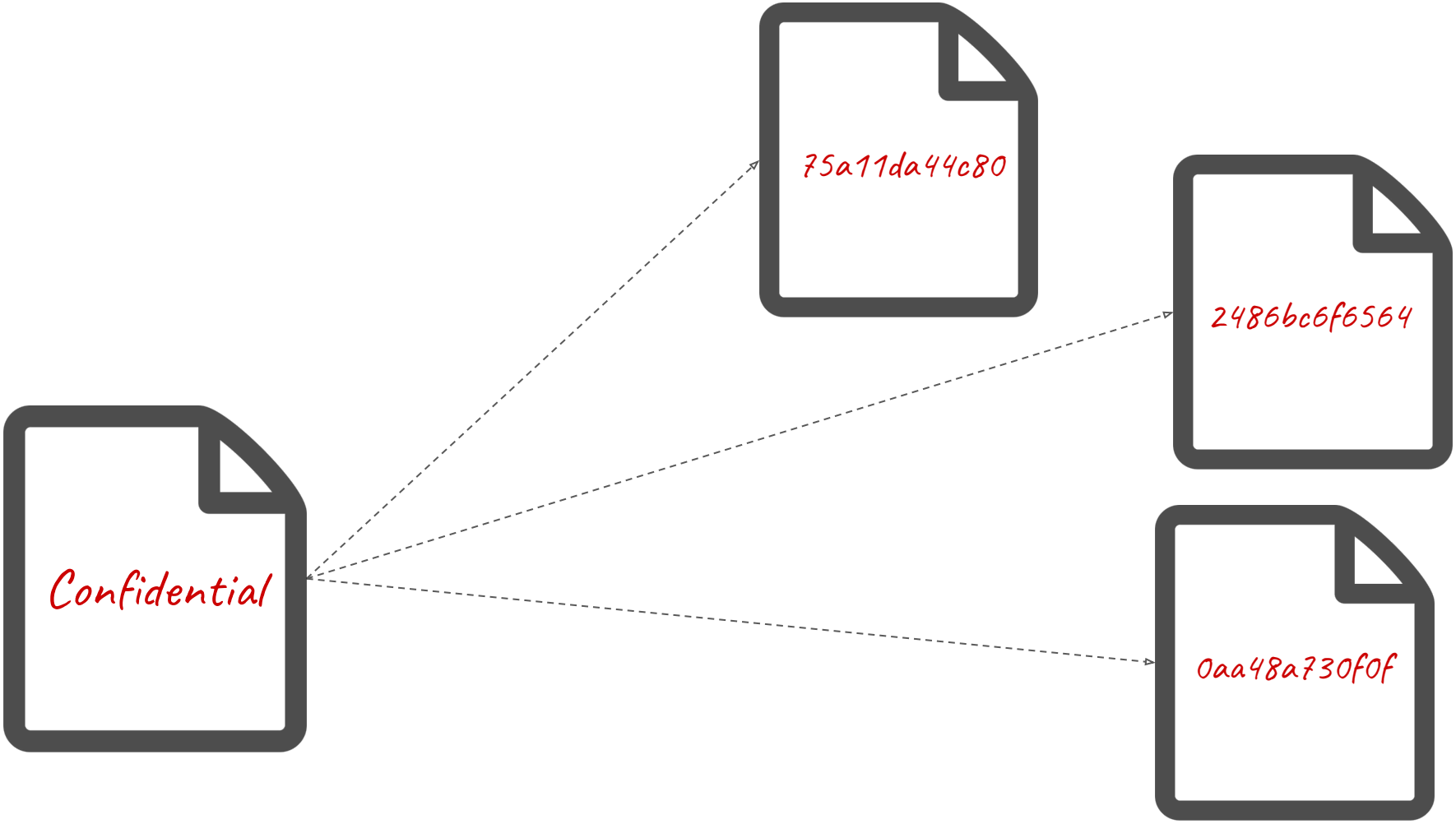


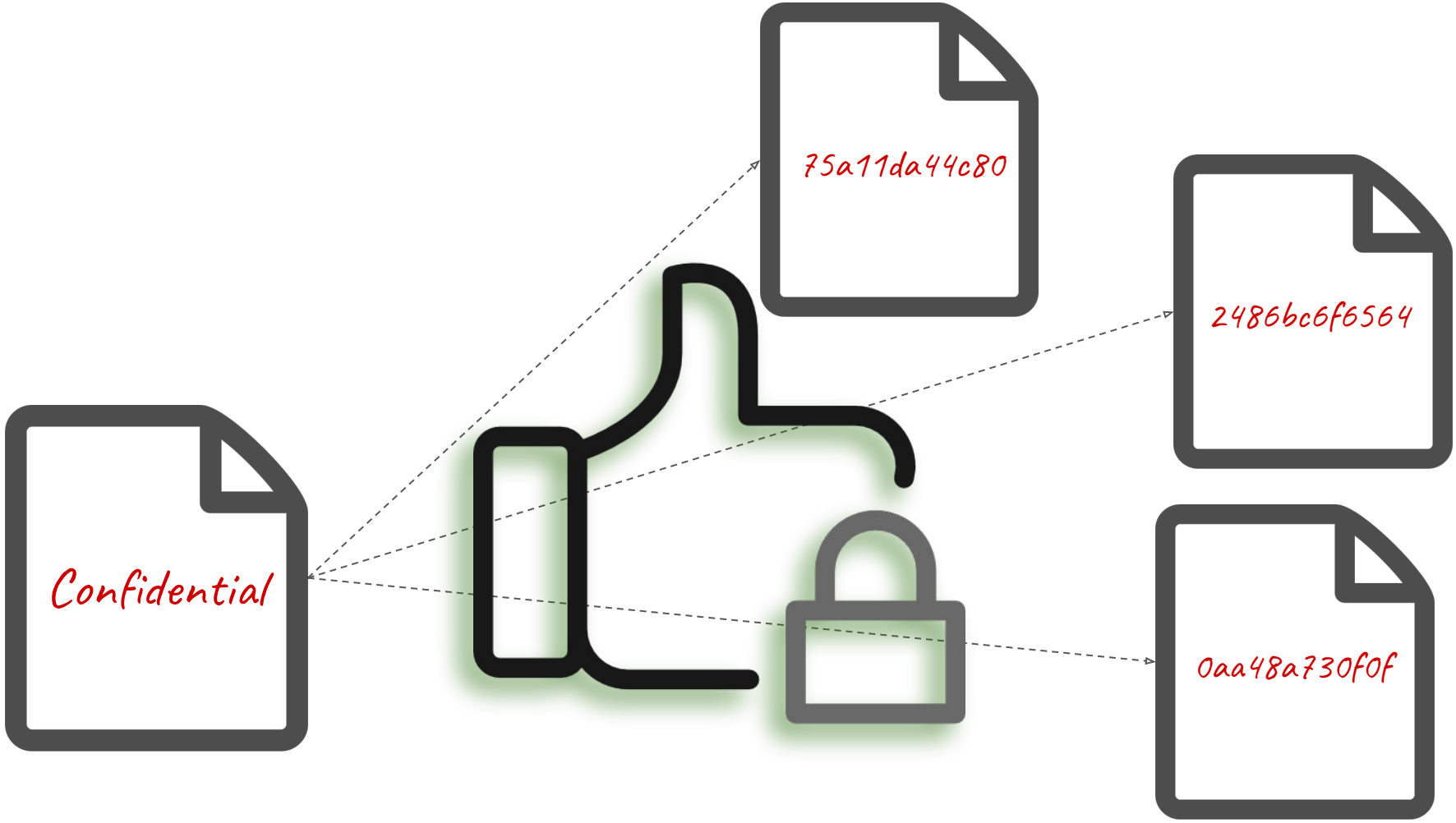
Secret Sharing




Confidential
----iden----
Conf--------
--------tial


Confidential
----iden----
Conf--------
--------tial


75a11da44c80
2486bc6f6564
Confidential
0aa48a730f0f


75a11da44c80
2486bc6f6564
Confidential
0aa48a730f0f


25

57

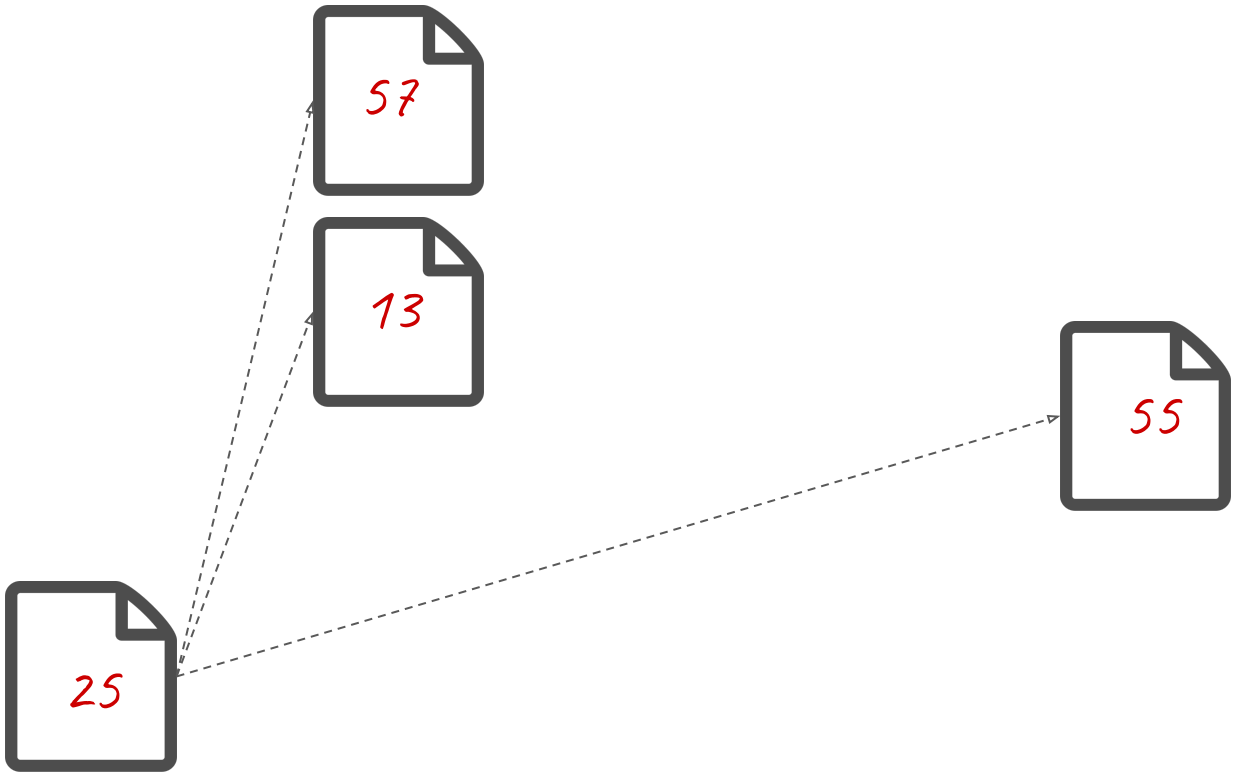
13

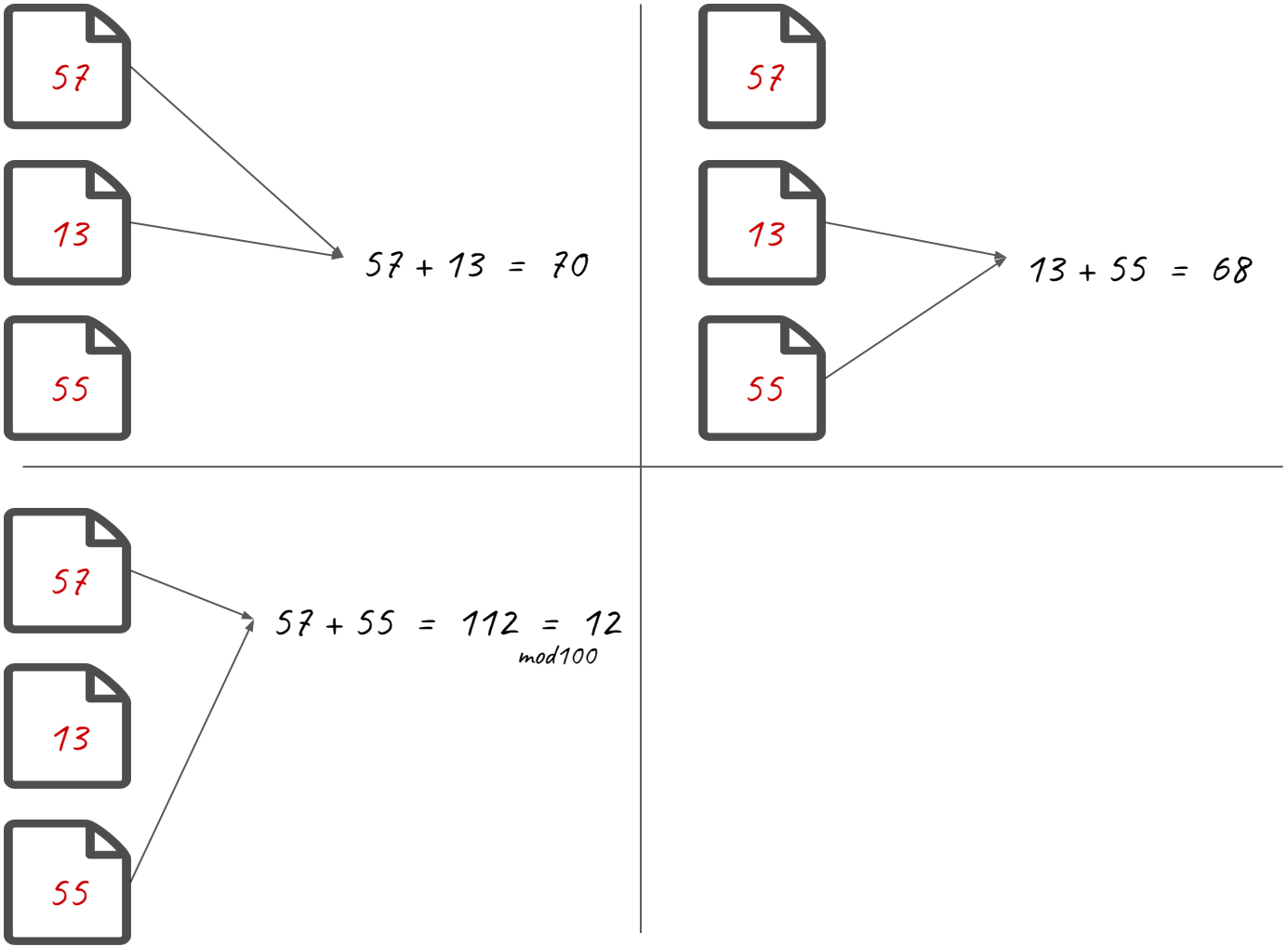
57

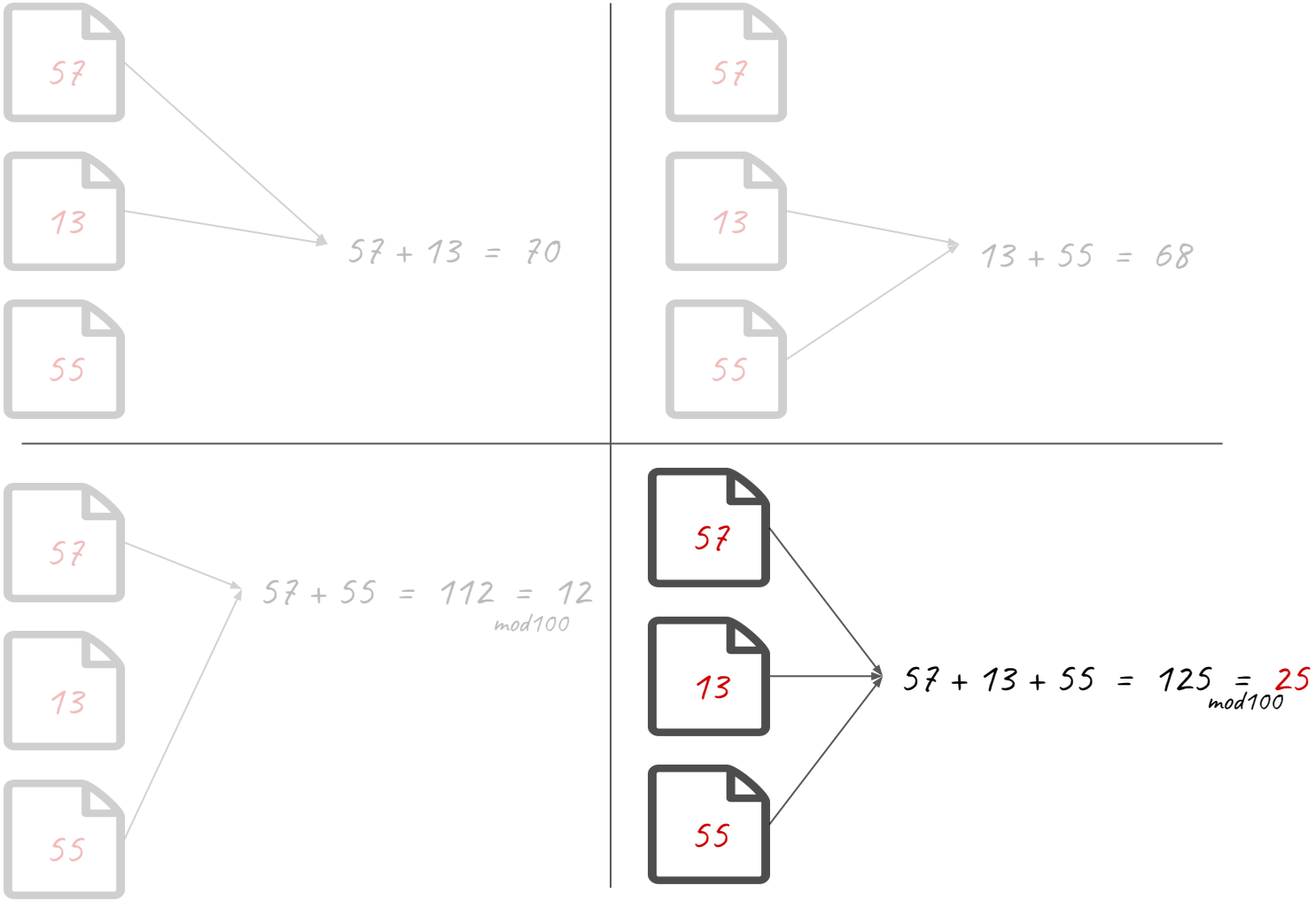
13

$$25 - 57 - 13 = -45 \underset{mod100}{=} 55$$

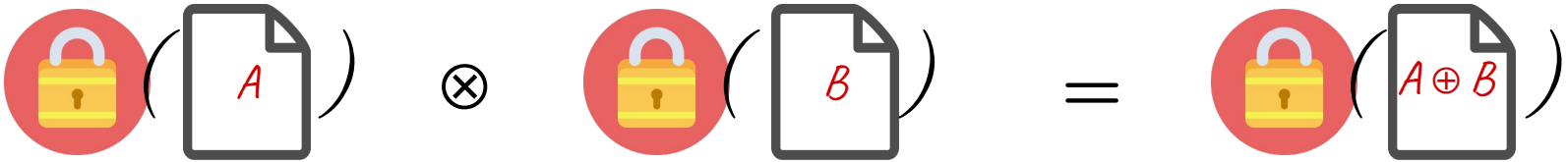
25

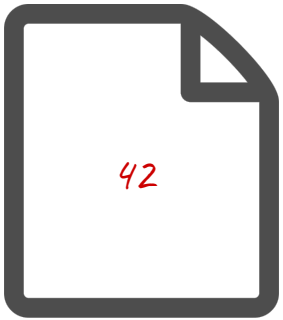

57
13
55
25

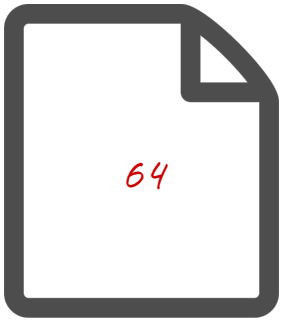

57
13
55
57 + 13 = 70
57
13
55
13 + 55 = 68
57
13
55
57 + 55 = 112 = 12
mod100

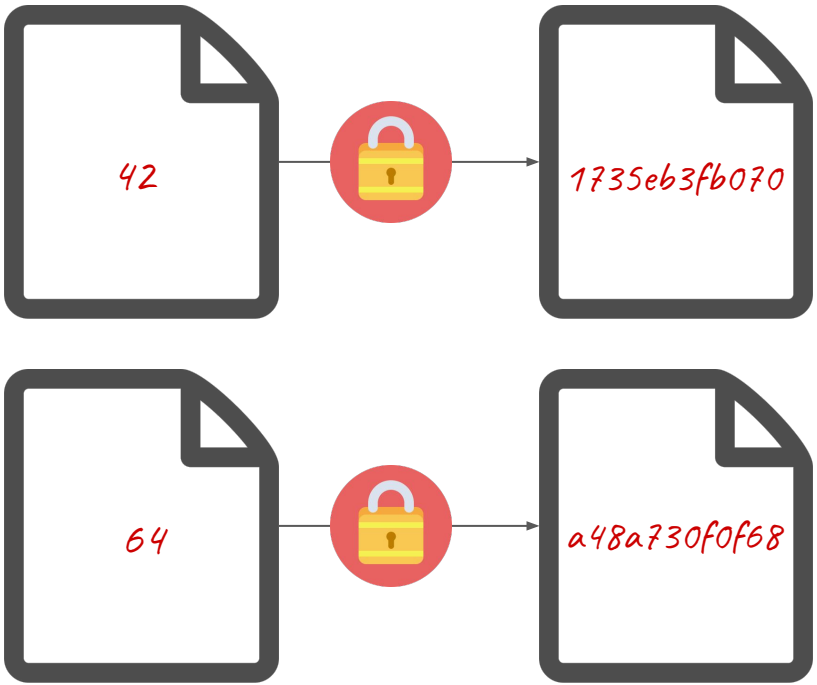

57
13
55
57 + 13 = 70
57
13
55
13 + 55 = 68
57
13
55
57 + 55 = 112 = 12
mod100
57
13
55
57 + 13 + 55 = 125 = 25
mod100

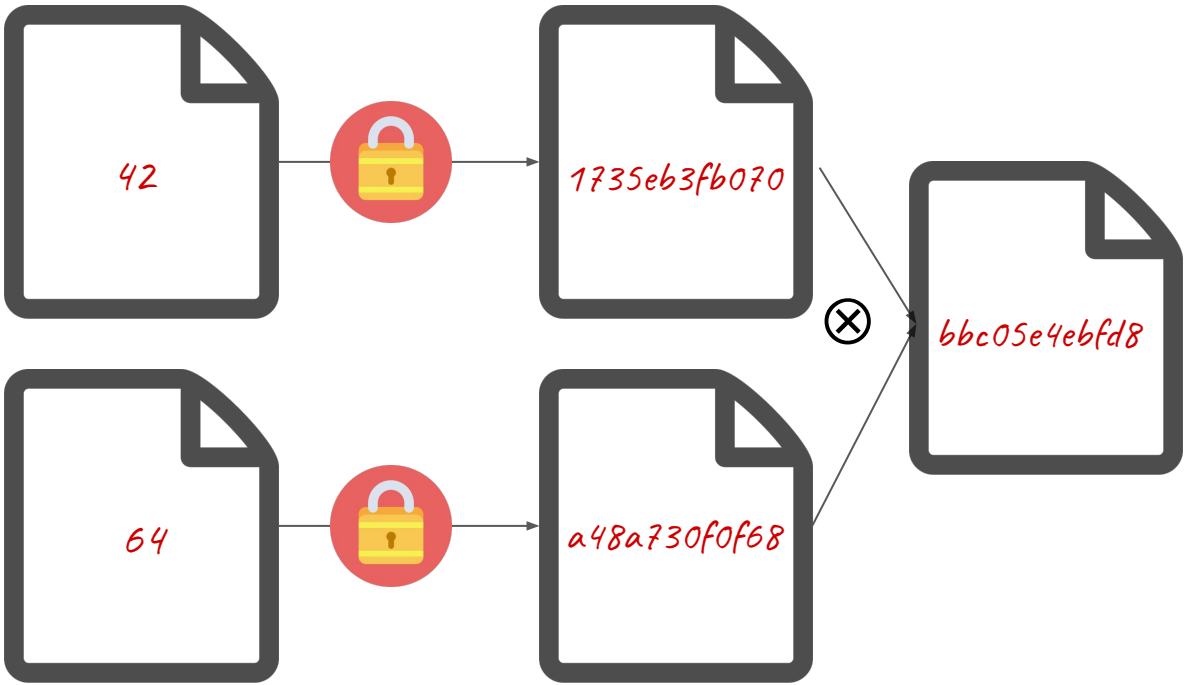
Homomorphic Encryption




42


64

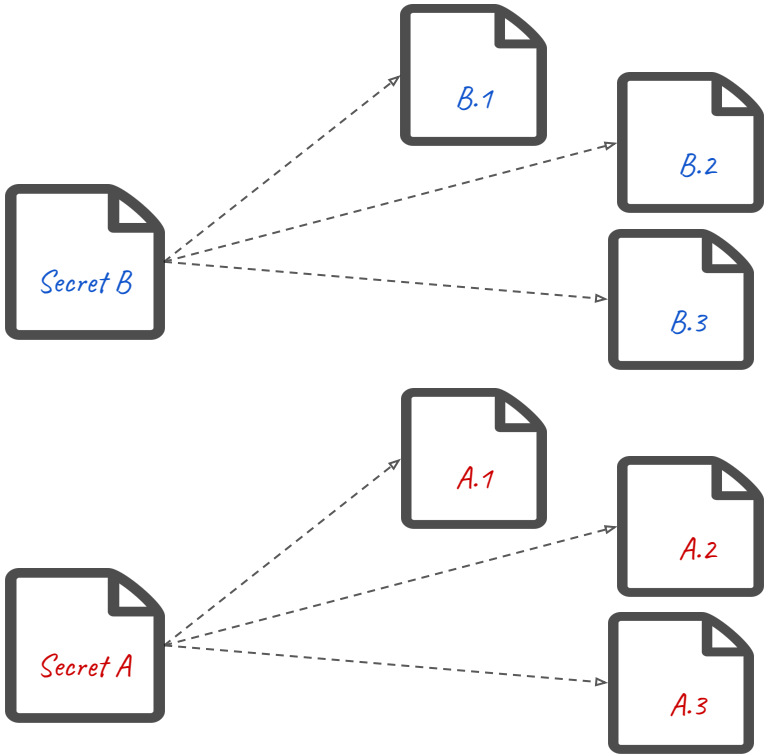

42
1735eb3fb070
64
a48a730f0f68


42
64
1735eb3fb070
a48a730f0f68
bbc05e4ebfd8

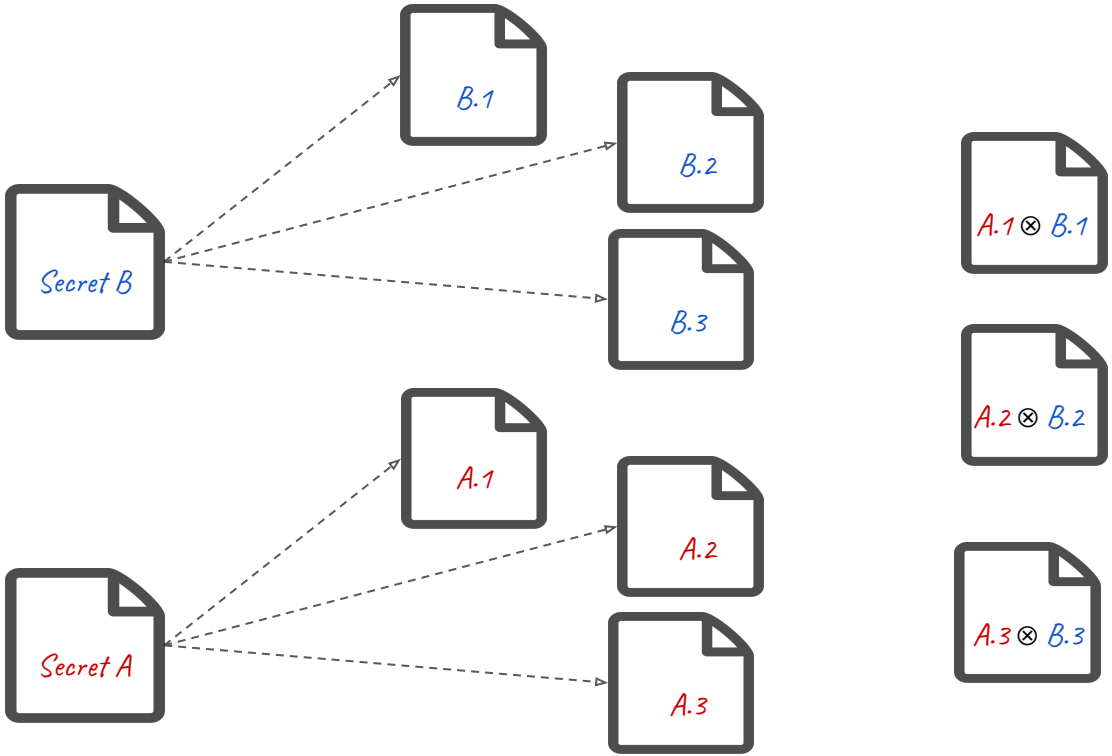

42
1735eb3fb070
64
a48a730f0f68
bbc05e4ebfd8
106

Secret Sharing Homomorphism (Example)

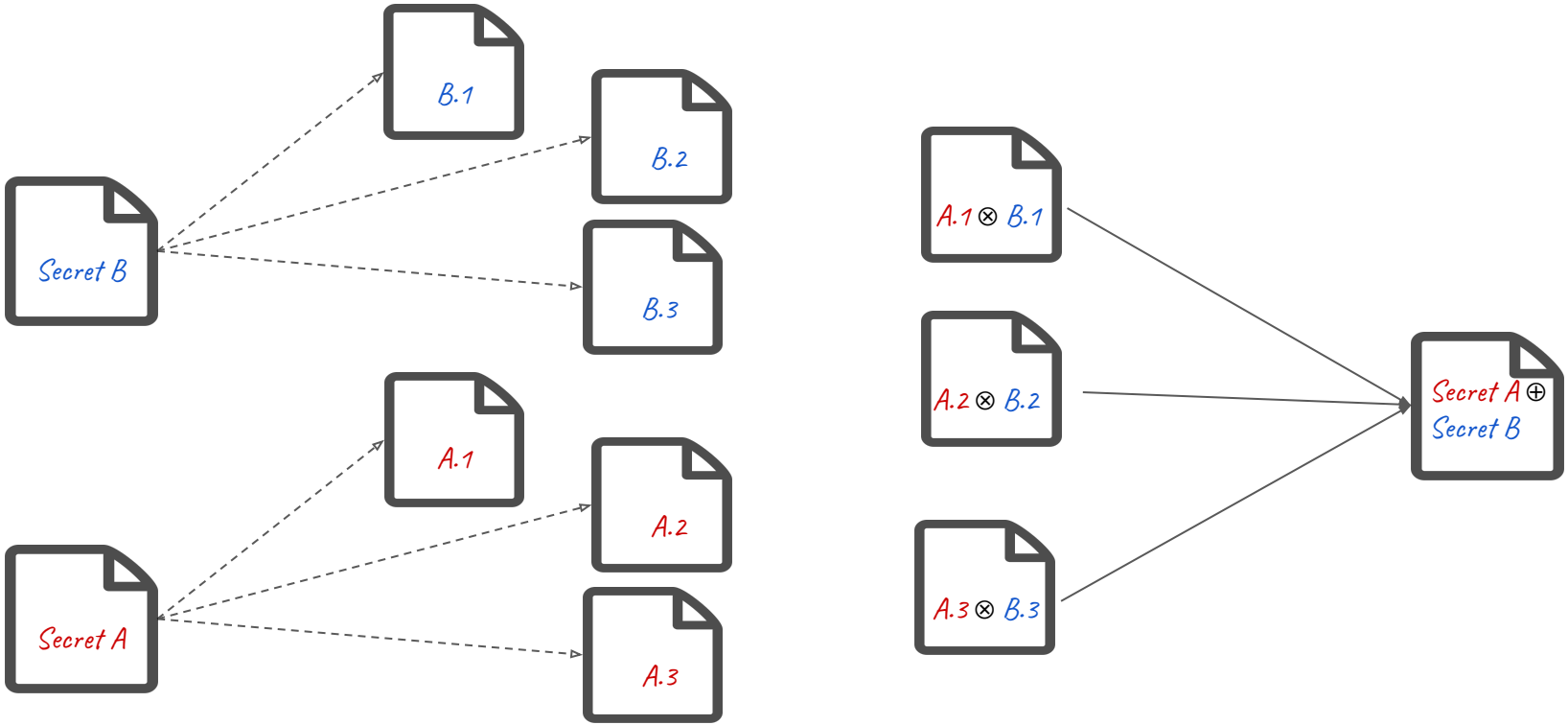
Secret Sharing Homomorphism (Example)



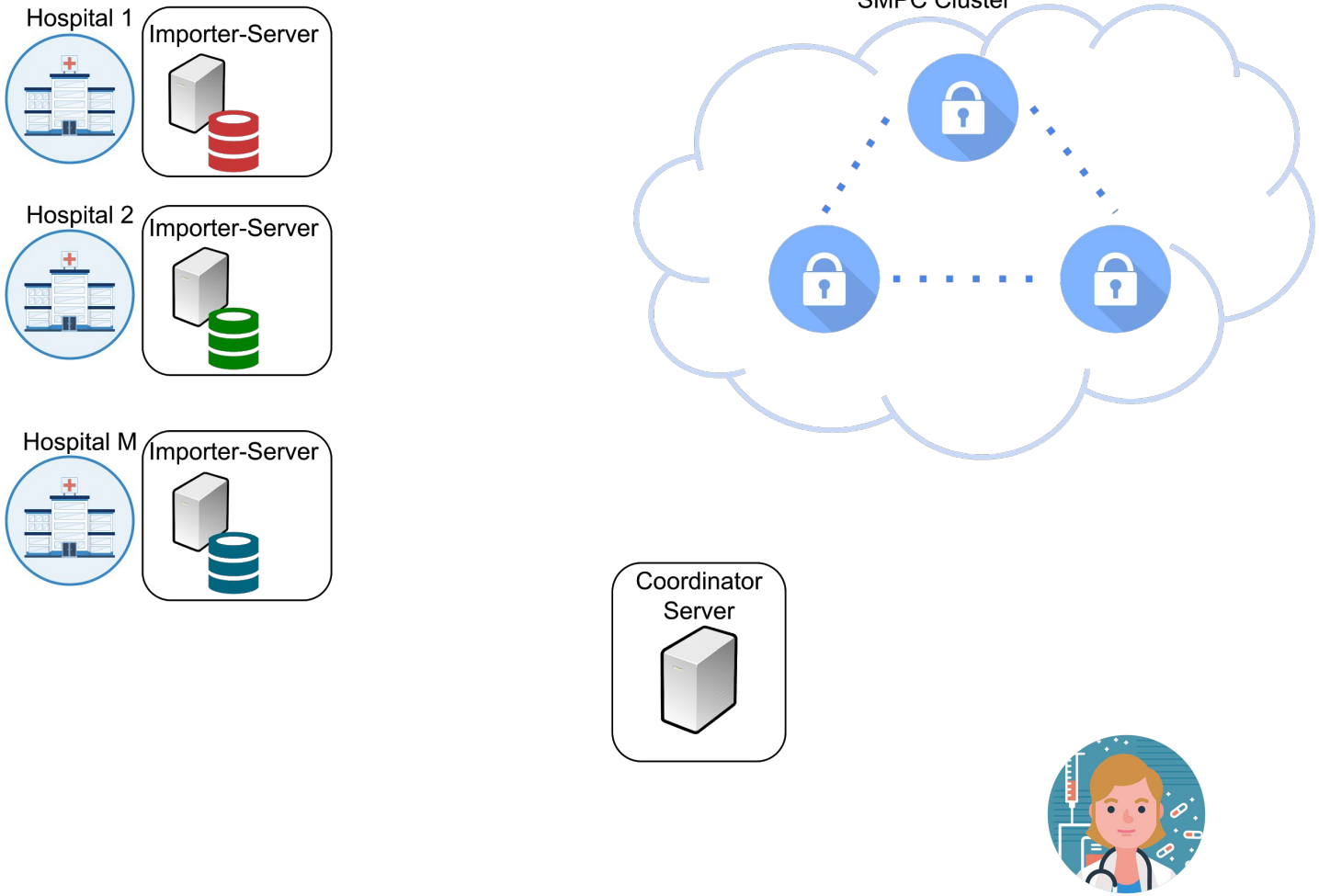
Secret Sharing Homomorphism (Example)

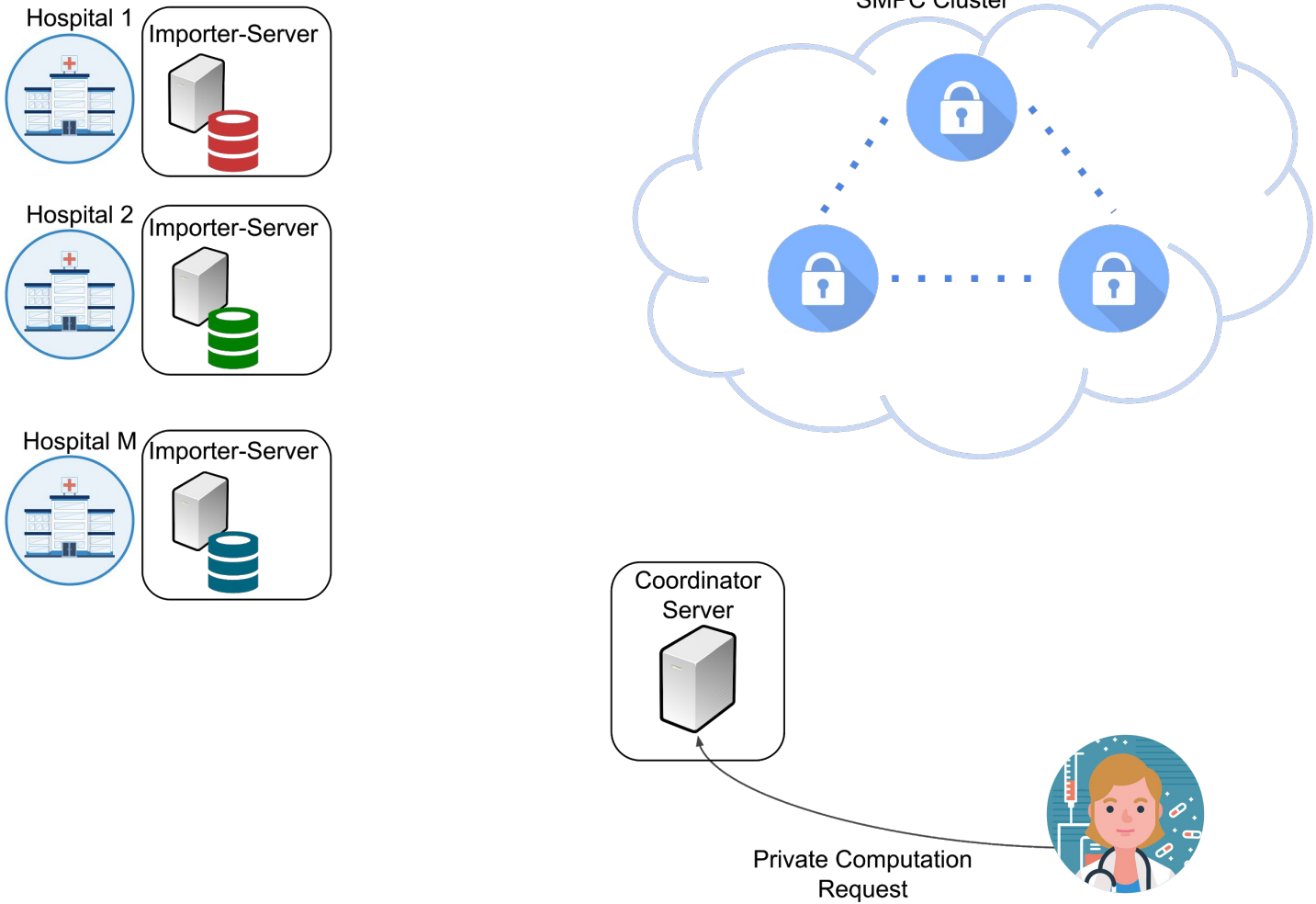


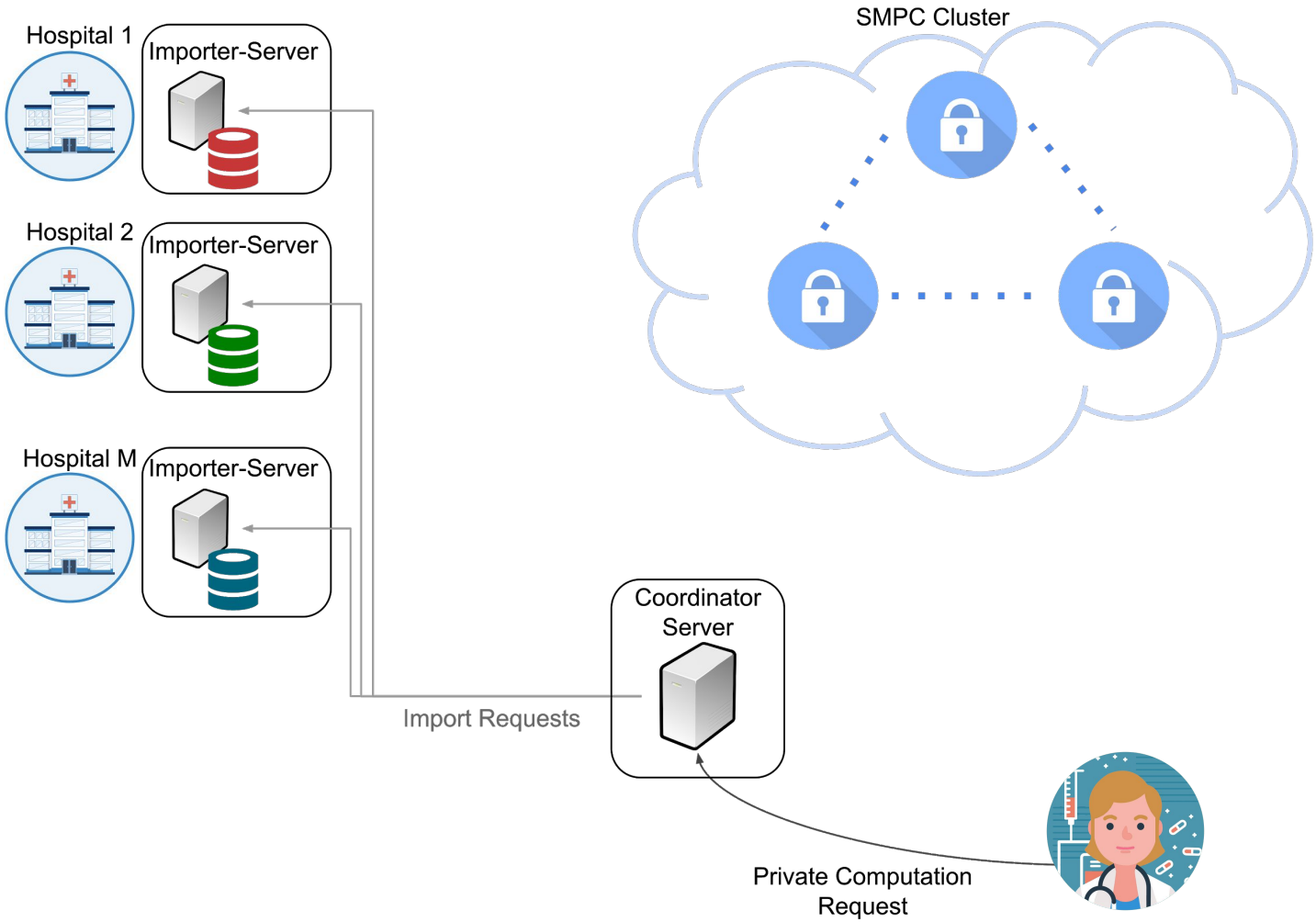
Secret Sharing Homomorphism (Example)

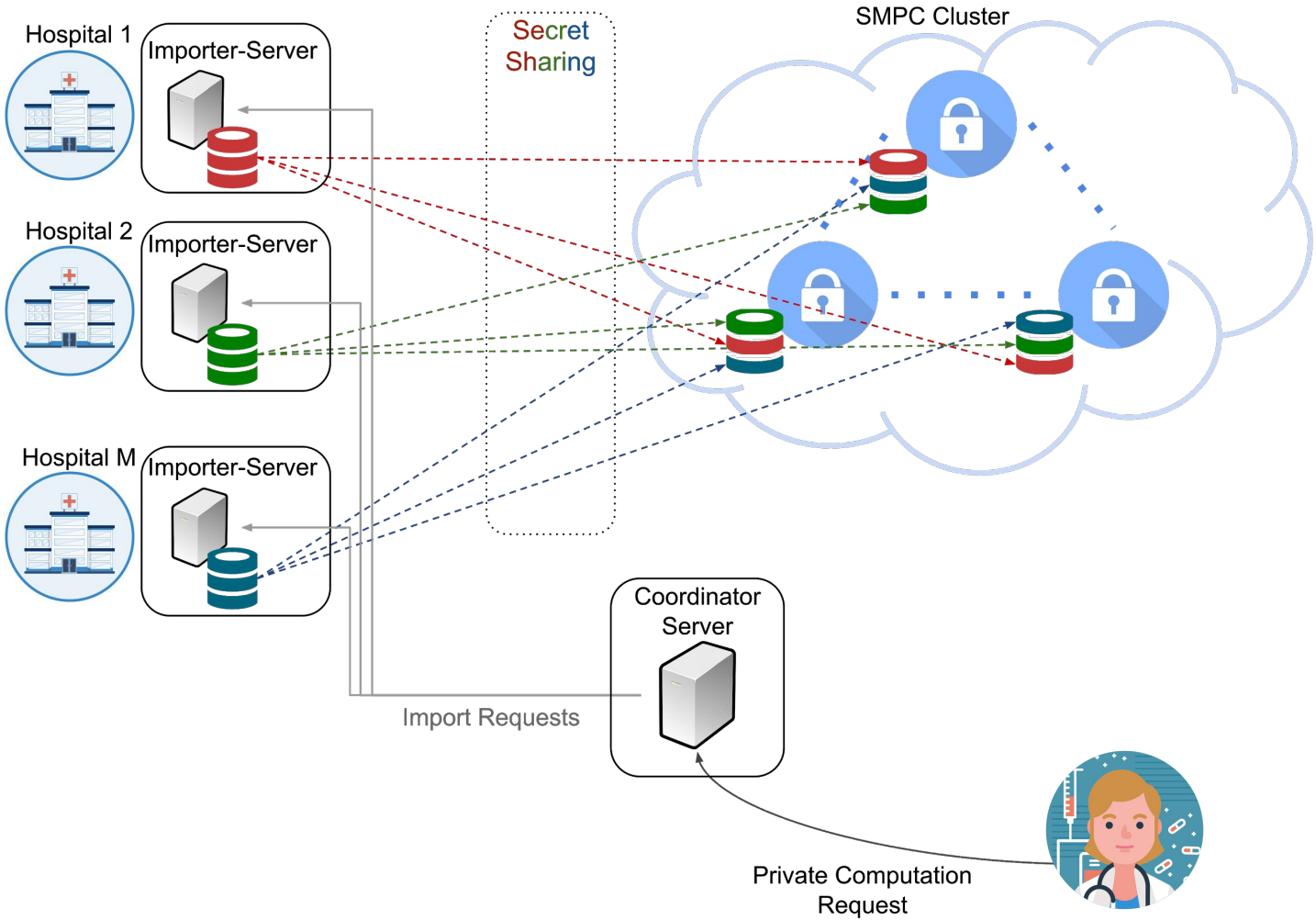


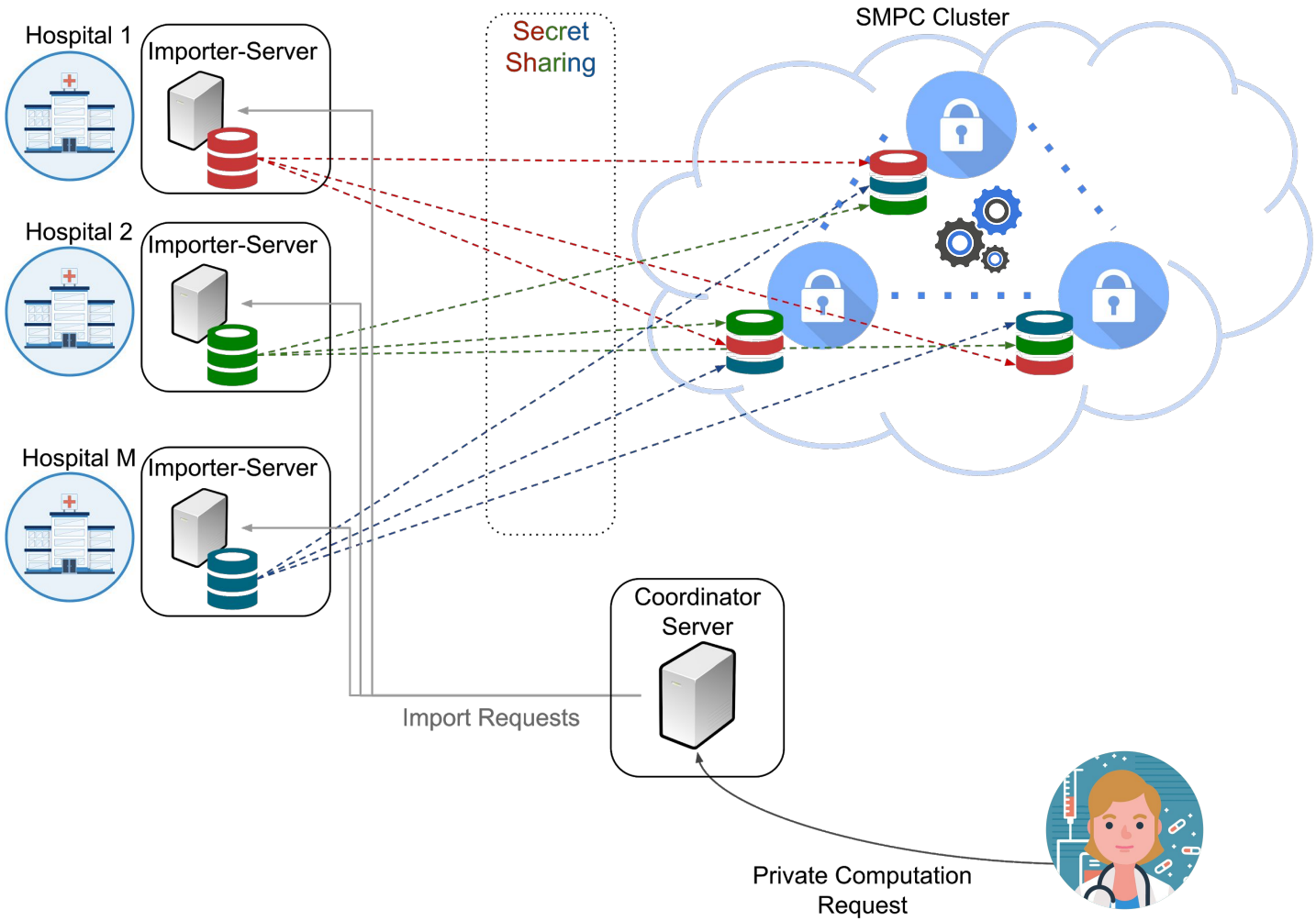
Our Architecture

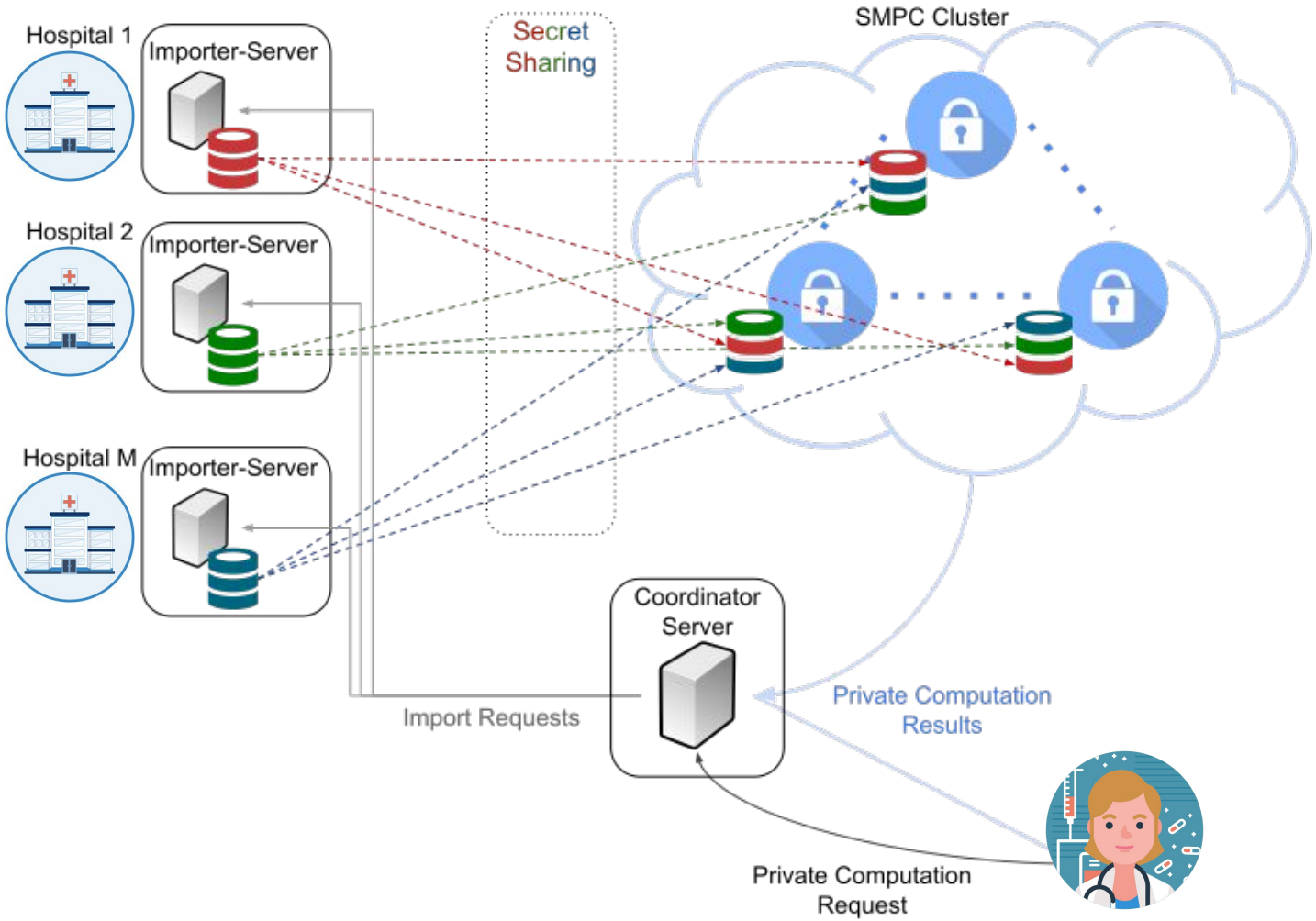

Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server
Private Computation Request

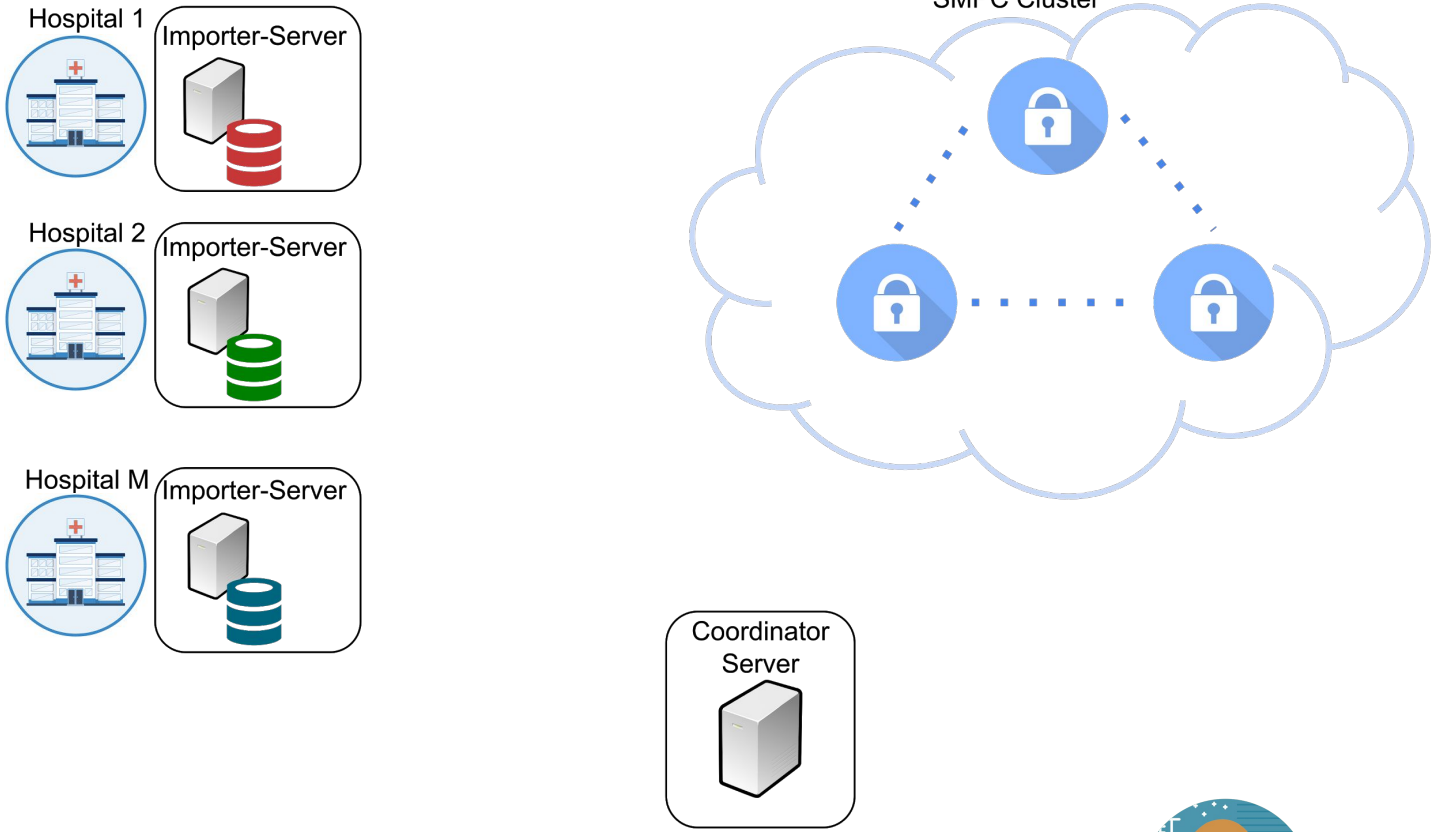

Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server
Import Requests
Private Computation Request


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
Secret Sharing
SMPC Cluster
Coordinator Server
Import Requests
Private Computation Request


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
Secret Sharing
SMPC Cluster
Coordinator Server
Import Requests
Private Computation Request

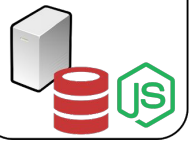

Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
Secret Sharing
SMPC Cluster
Coordinator Server
Import Requests
Private Computation Results
Private Computation Request

Implementation


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server

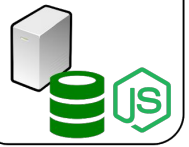
Hospital 1

Importer-Server



Hospital 2

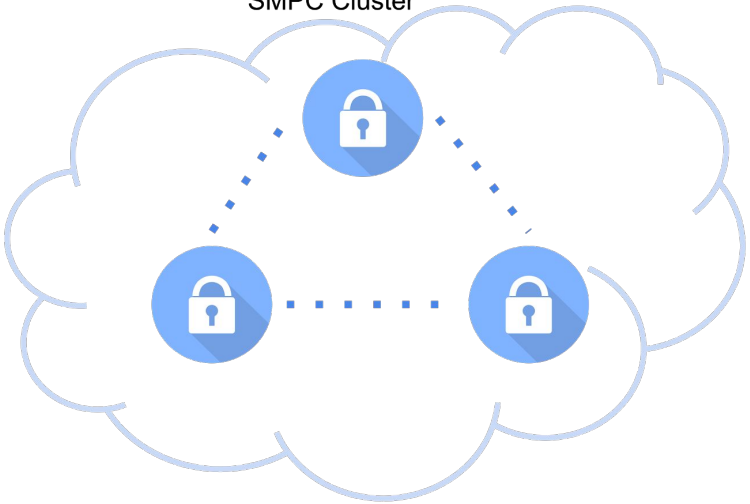
Importer-Server



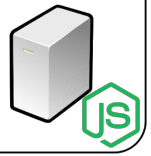
Hospital M

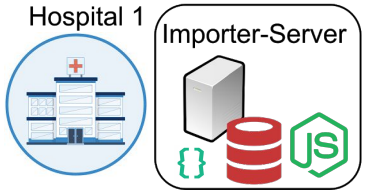
Importer-Server

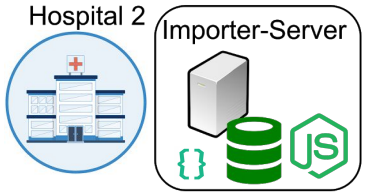
SMPC Cluster

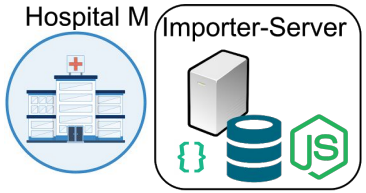


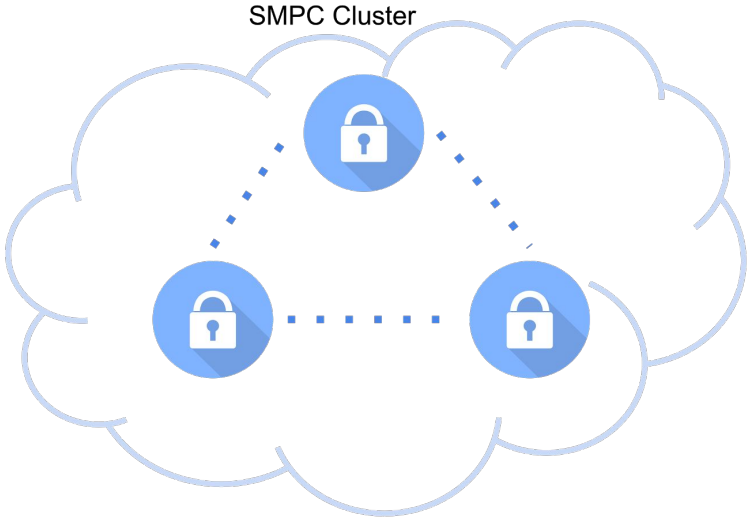
Coordinator Server

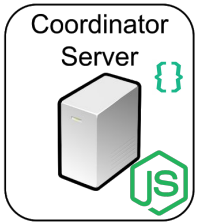



Hospital 1
Importer-Server

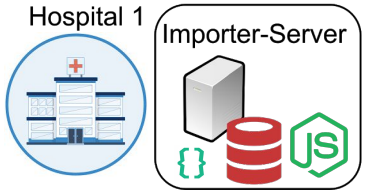

Hospital 2
Importer-Server

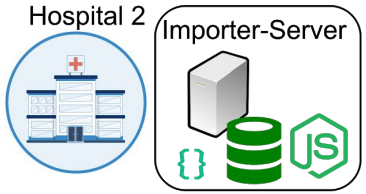

Hospital M
Importer-Server

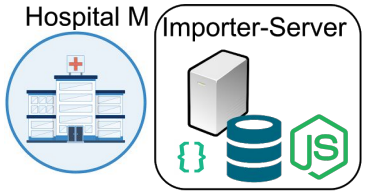

SMPC Cluster

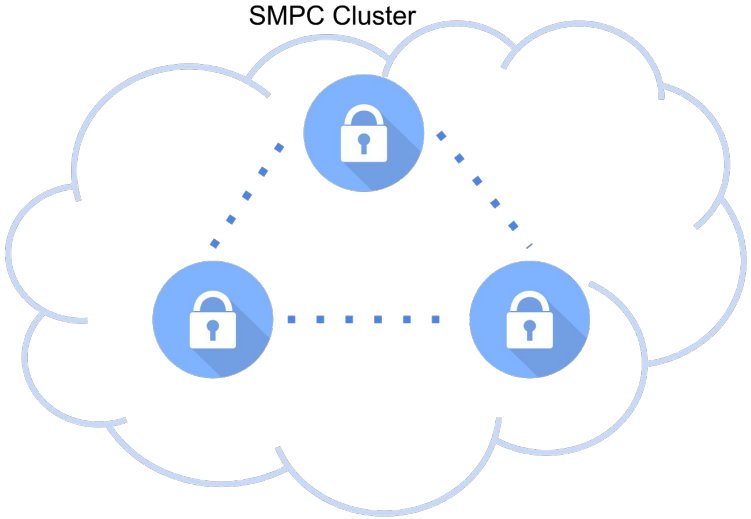

Coordinator
Server

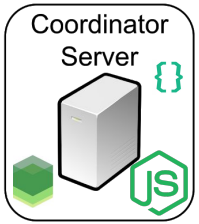
node{ REST }



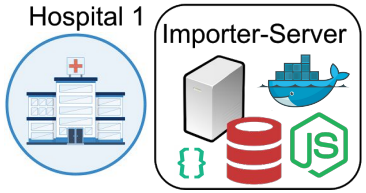


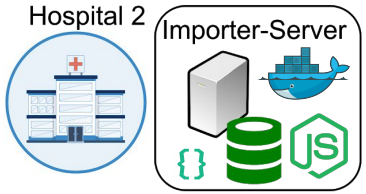


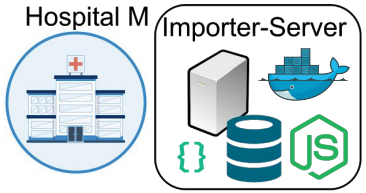


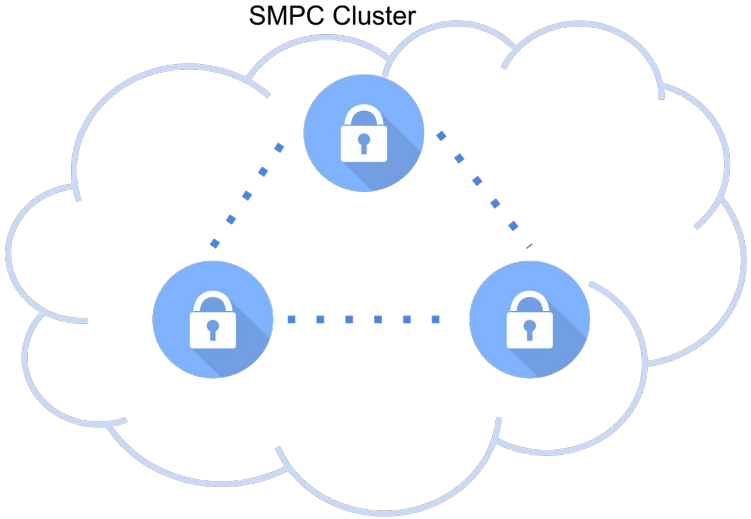


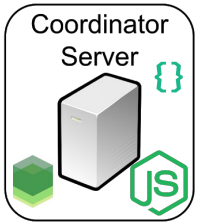
node { REST } LEVELDB



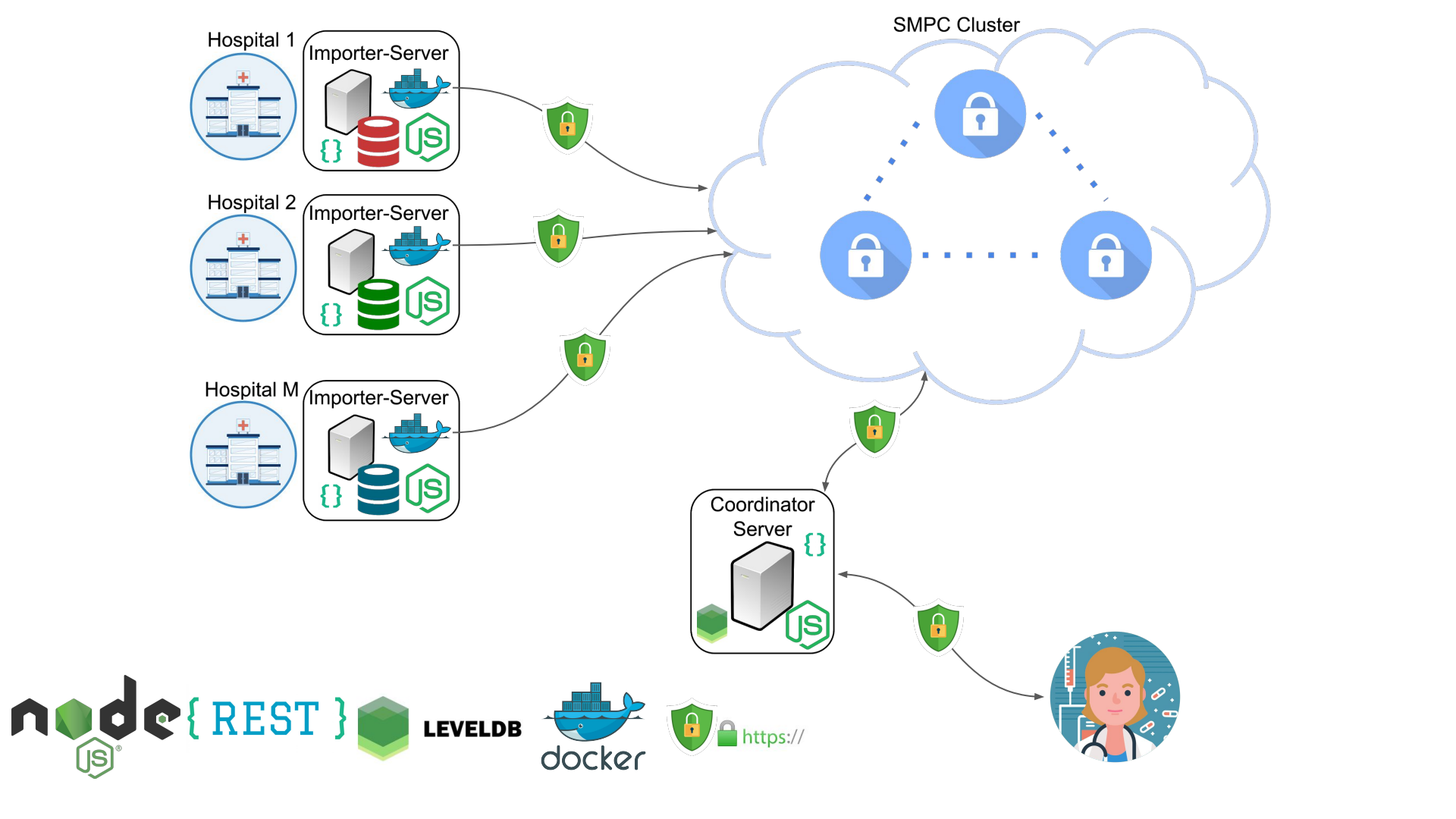


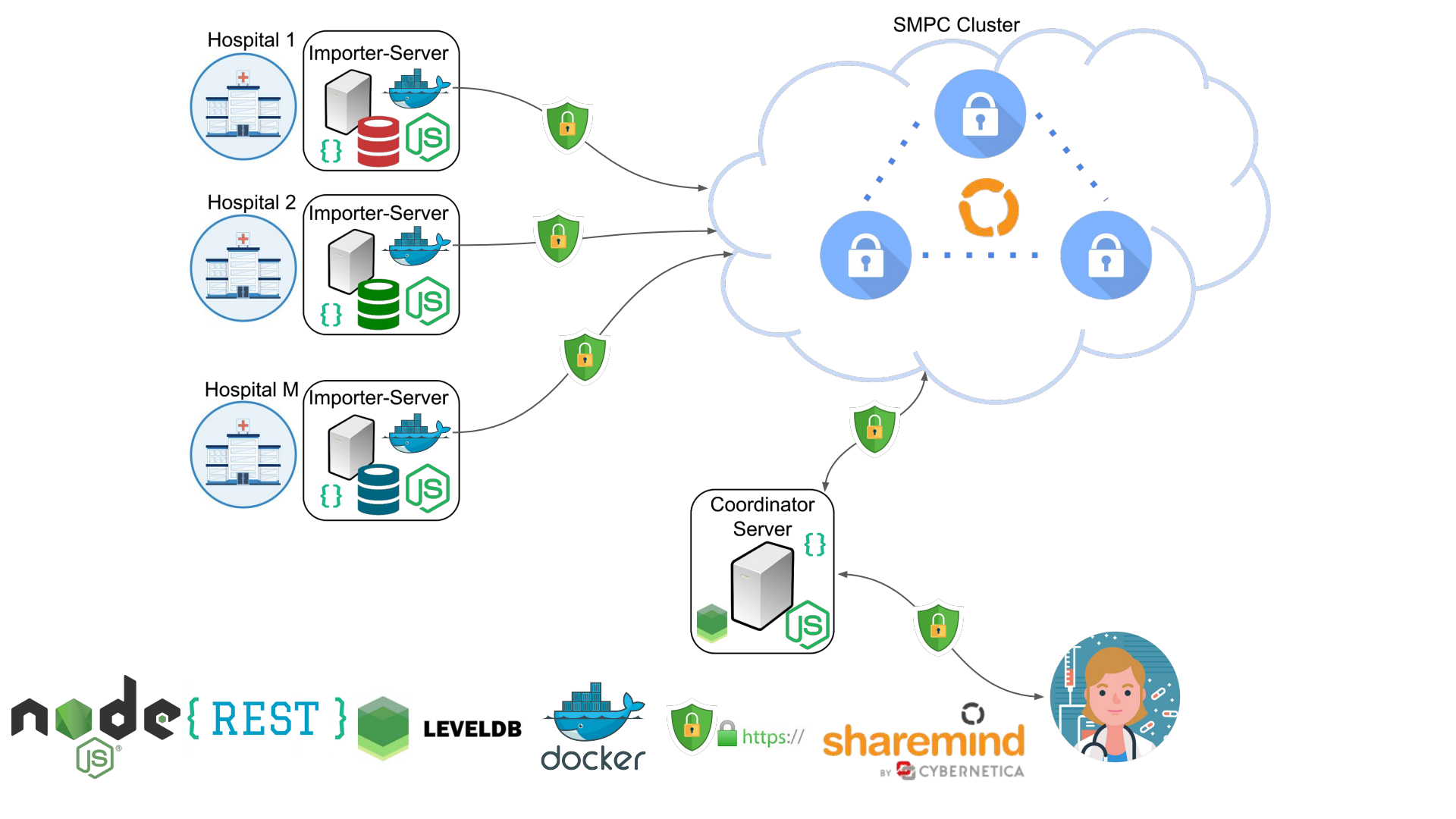






node { REST } LEVELDB docker


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server
node
{ REST }
LEVELDB
docker
https://


Hospital 1
Importer-Server
Hospital 2
Importer-Server
Hospital M
Importer-Server
SMPC Cluster
Coordinator Server
node
{ REST }
LEVELDB
docker
https://
sharemind
BY CYBERNETICA

Sharemind™ is a database and runtime system that works on encrypted data without decrypting it.

Provides SecreC programming language for writing privacy-preserving algorithms.

Privacy-Preserving Algorithms

Threat Model

Program evaluation is outsourced to **honest-but-curious** (semi-honest) third parties

Individual with incentives to eavesdrop sensitive user data

Threat Model

Program evaluation is outsourced to **honest-but-curious** (semi-honest) third parties

Individual with incentives to eavesdrop sensitive user data

- Access to volatile memories and/or storage of computational servers
- Intentional/Unintentional data breaches
- Cyberattacks (*buffer-overflows*, *return-oriented-programming*, *Heartbleed*, *Spectre/Meltdown*, *Hardware Trojans*, *etc.*)
- ...

Pseudocode:

```
if (x = 42) then
    flag ← true
else
    flag ← false
```

Pseudocode:

```
if (x = 42) then
    flag ← true
else
    flag ← false
```

Another way:

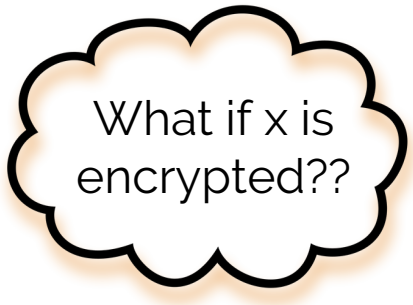
```
flag ← (x = 42)
```

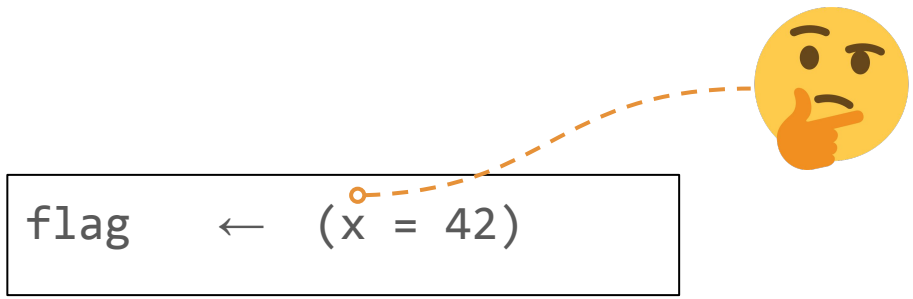
Pseudocode:

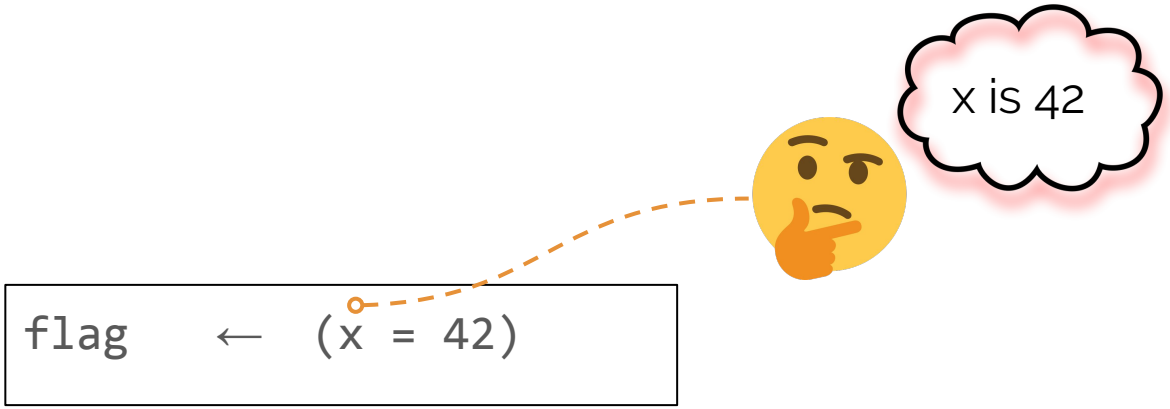
```
if (x = 42) then
    flag ← true
else
    flag ← false
```

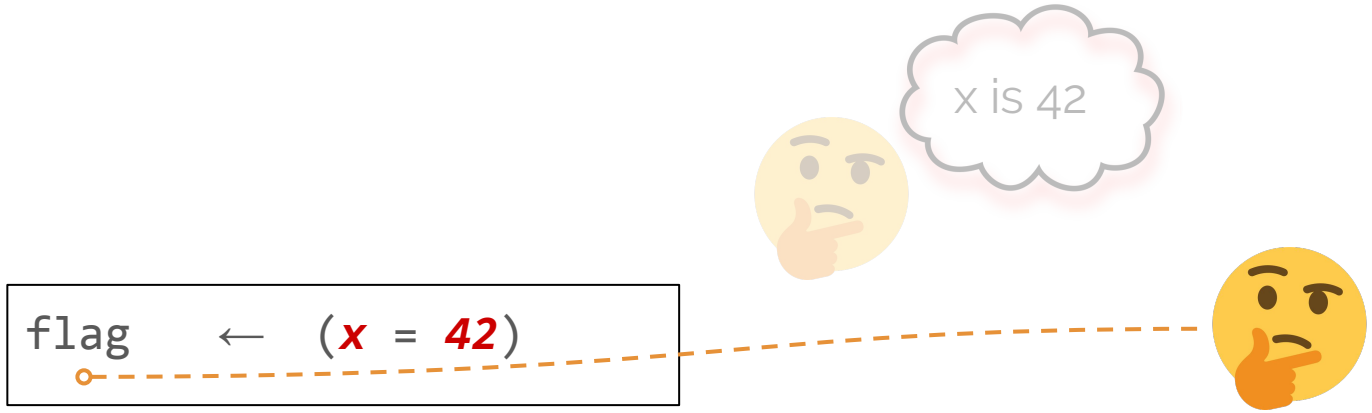
Another way:

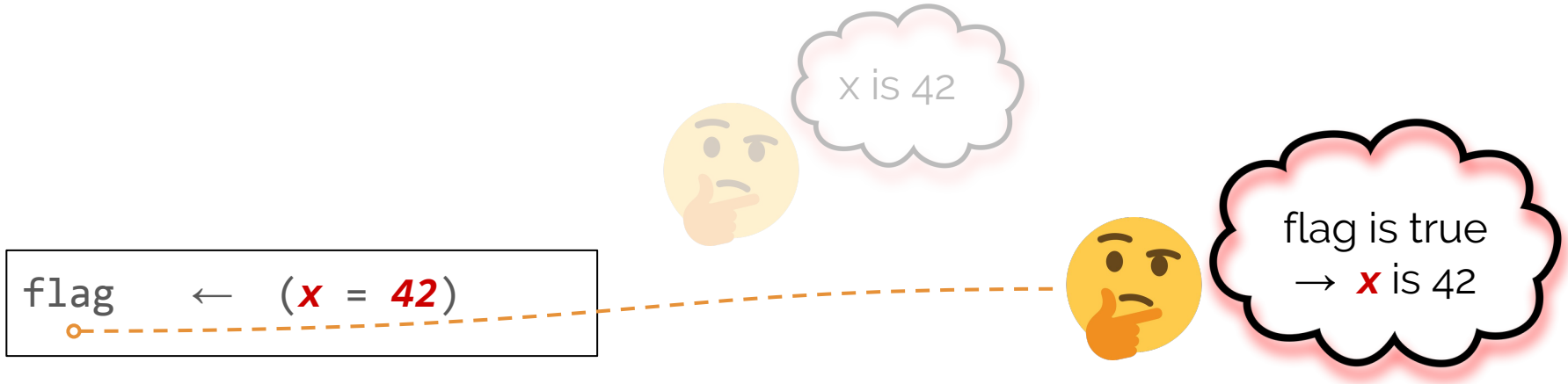
```
flag ← (x = 42)
```

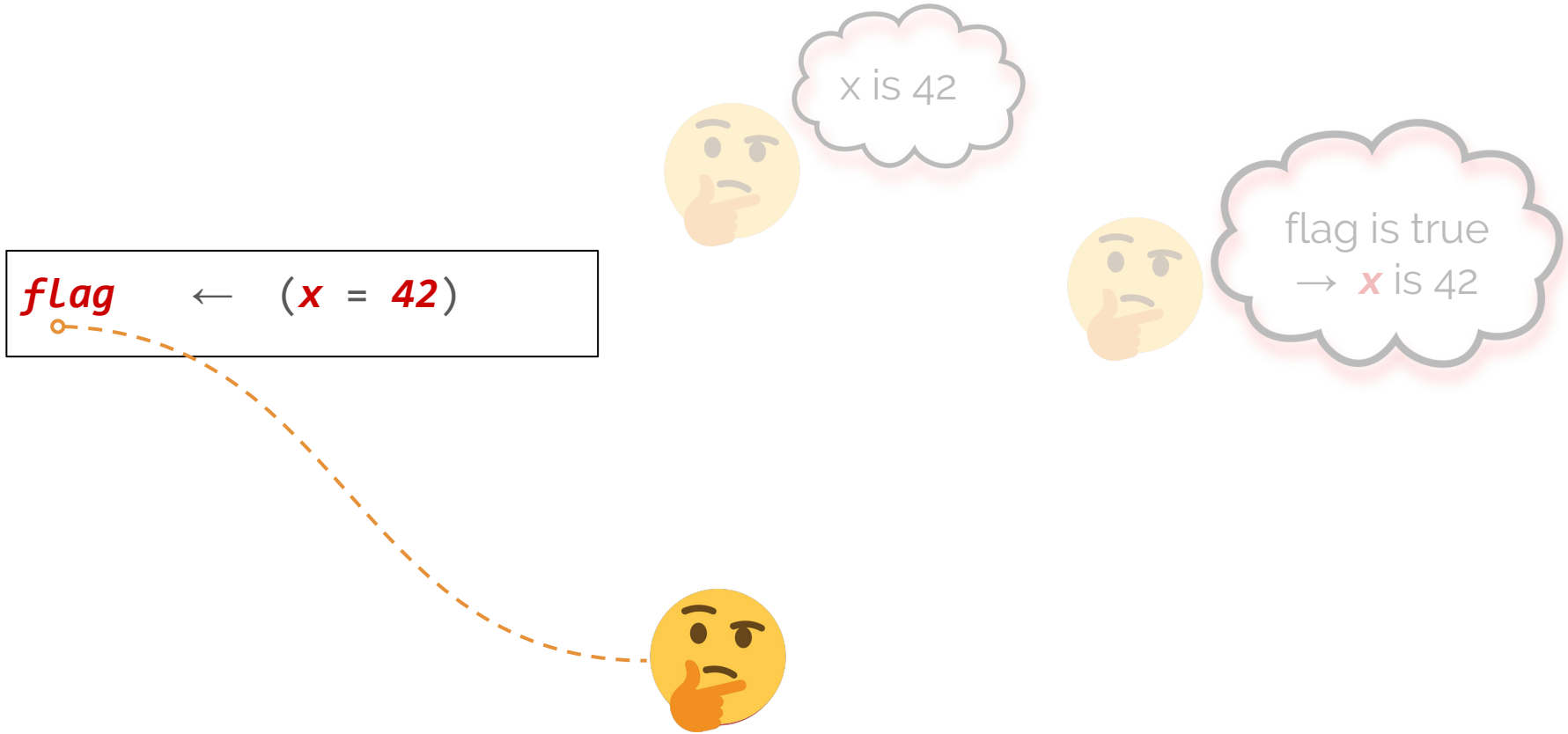


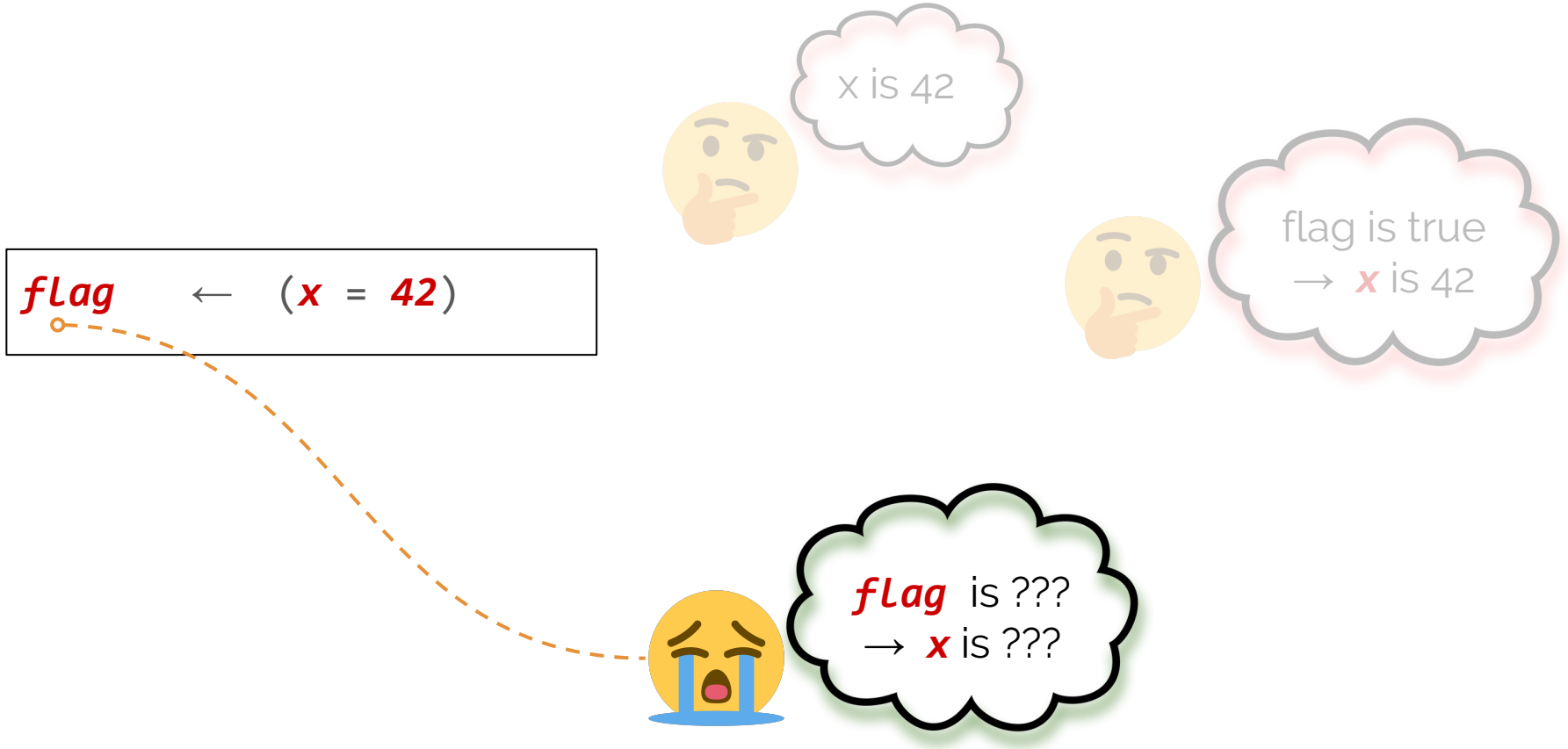

flag ← (x = 42)

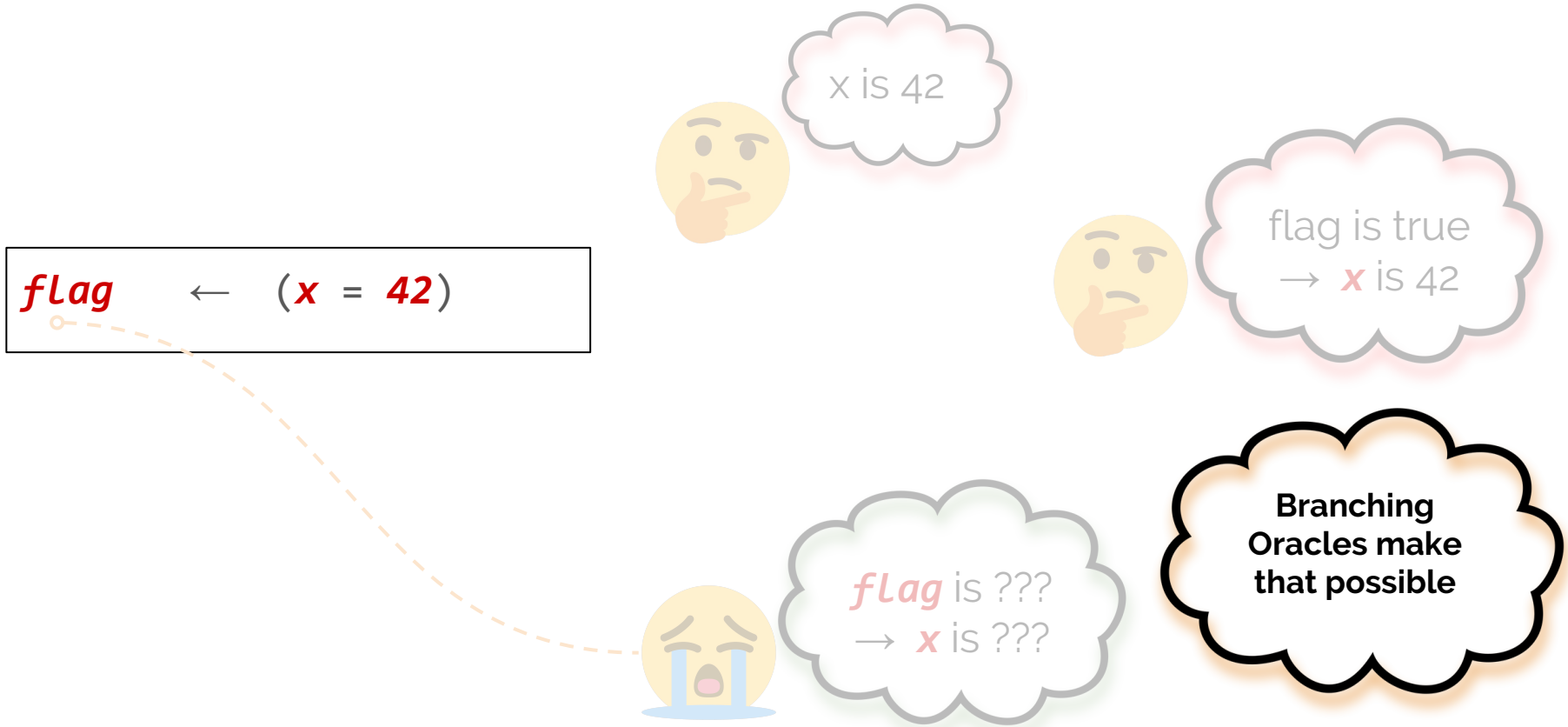

x is 42
flag ← (x = 42)


x is 42
flag ← (x = 42)


x is 42
flag ← (x = 42)
flag is true
→ x is 42


x is 42
flag is true
→ x is 42
flag ← (x = 42)


x is 42
flag is true
→ x is 42
flag ← (x = 42)
flag is ???
→ x is ???


flag ← (x = 42)
x is 42
flag is true
→ x is 42
flag is ???
→ x is ???
Branching
Oracles make
that possible

Computing Minimum of List

	Textbook
1	`min(array[N]) :`
2	`    min ← +∞`
3	`    for i in range(N):`
4	`        if array[i] < min`
5	`            min ← array[i]`
6	`    return min`

Computing Minimum of List

	Textbook	Privacy-Preserving
1	`min(array[N]) :`	`pp-min(array[N]) :`
2	`    min ← +∞`	`    min ← +∞`
3	`    for i in range(N):`	`    for i in range(N):`
4	`        if array[i] < min`	`        lessthan ← array[i] < min`
5	`            min ← array[i]`	`        min ← lessthan * array[i] + (1 - lessthan) * min`
6	`    return min`	`    return min`

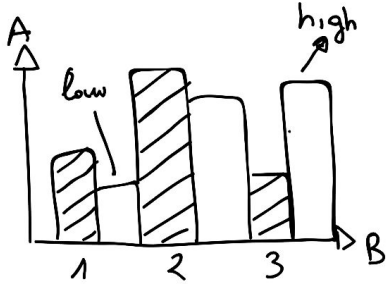
Computing Minimum of List

	Textbook	Privacy-Preserving
1	`min(array[N]) :`	`pp-min(array[N]) :`
2	`    min ← +∞`	`    min ← +∞`
3	`    for i in range(N):`	`    for i in range(N):`
4	`        if array[i] < min`	`        lessthan ← array[i] < min`
5	`            min ← array[i]`	`        min ← lessthan * array[i] + (1 - lessthan) * min`
6	`    return min`	`    return min`

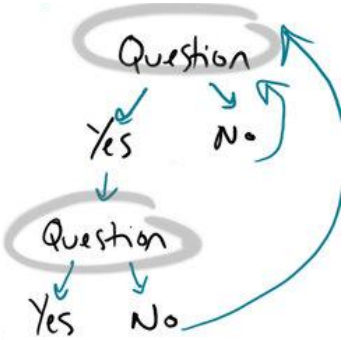
Branching Oracle returns Enc(1) if Dec(array[i]) < min, otherwise Enc(0)

Developed Algorithms

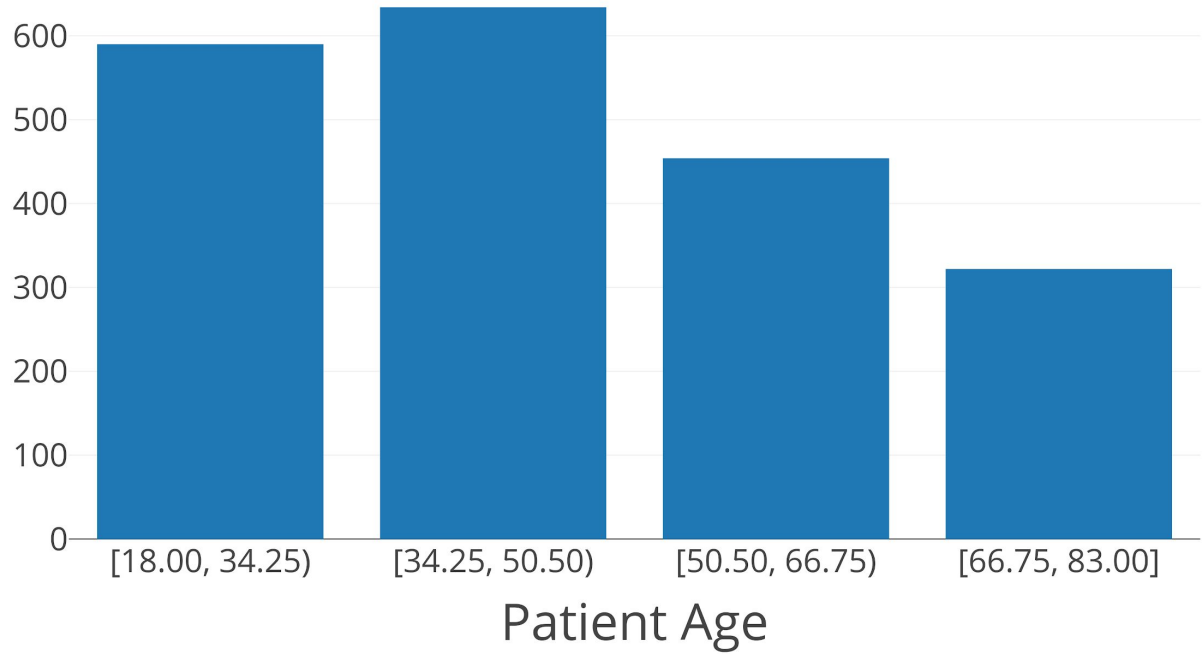
1) Aggregators



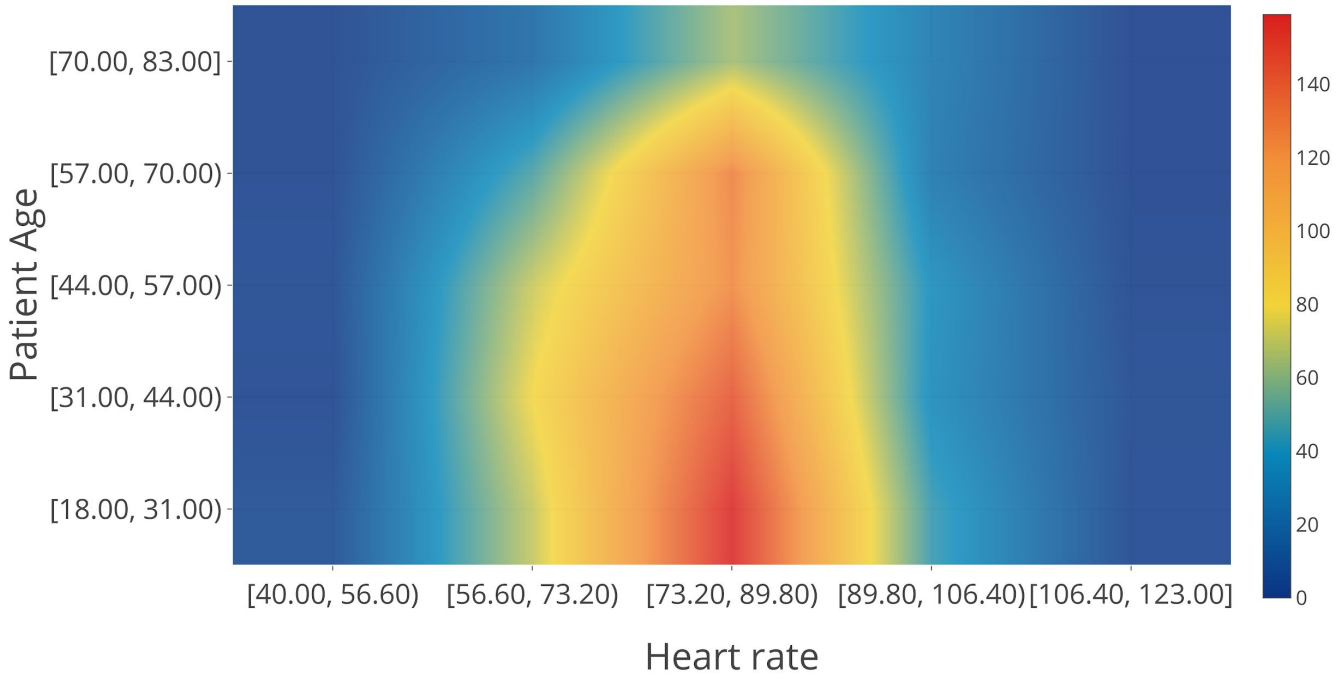
2) ML models/Decision Trees



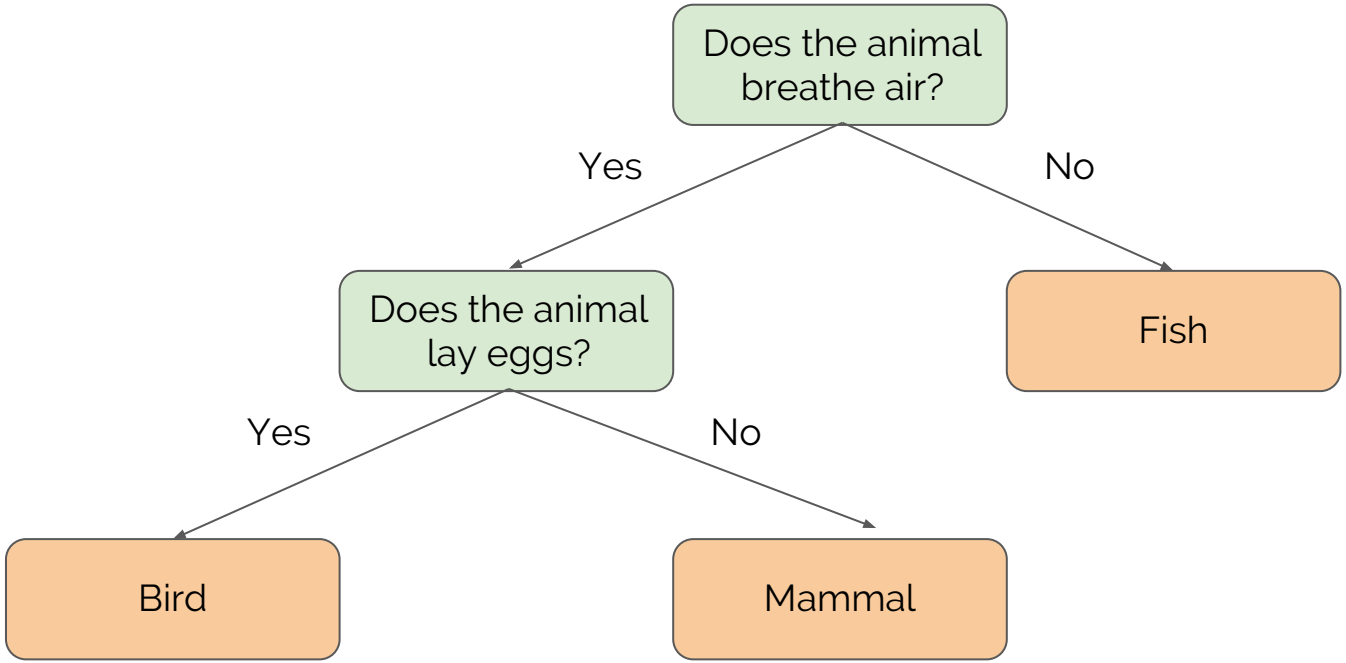
1) Aggregators (1D)



1) Aggregators (2D)

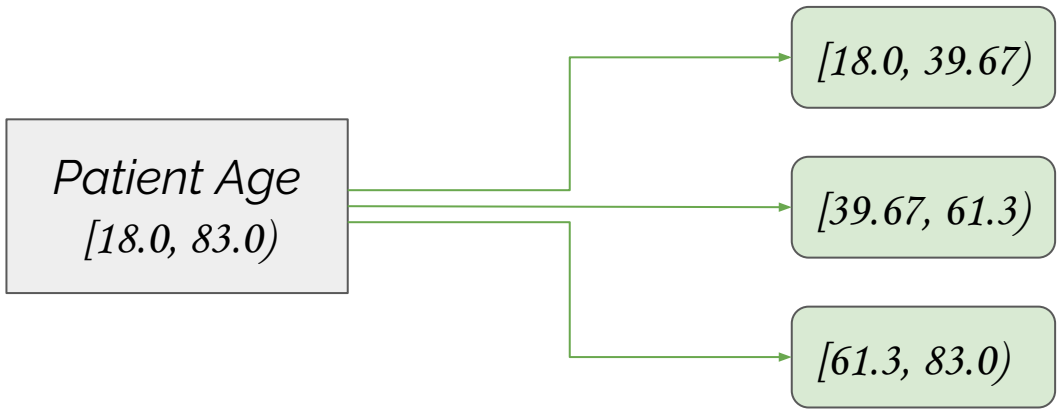


2) ML models/Decision Trees

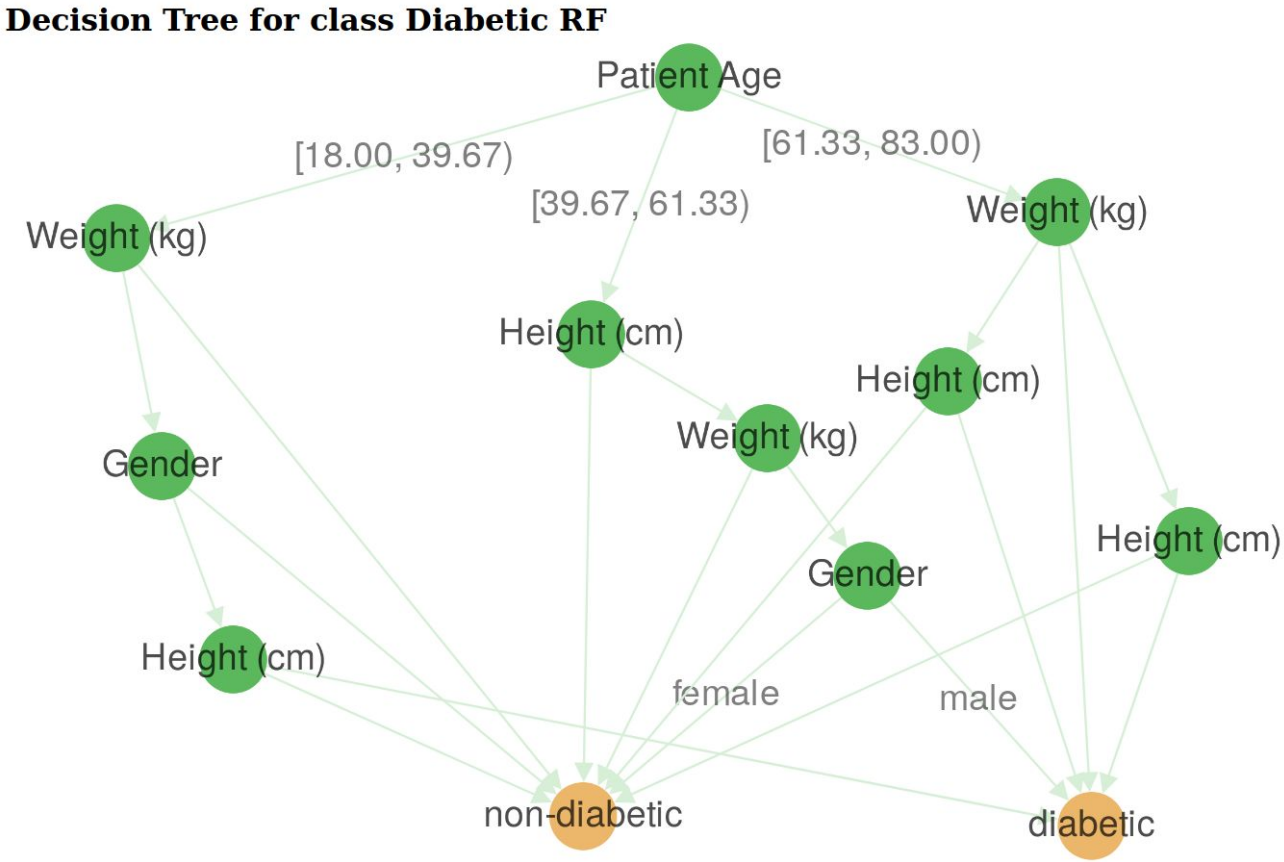


2) ML models/Decision Trees (ID3)

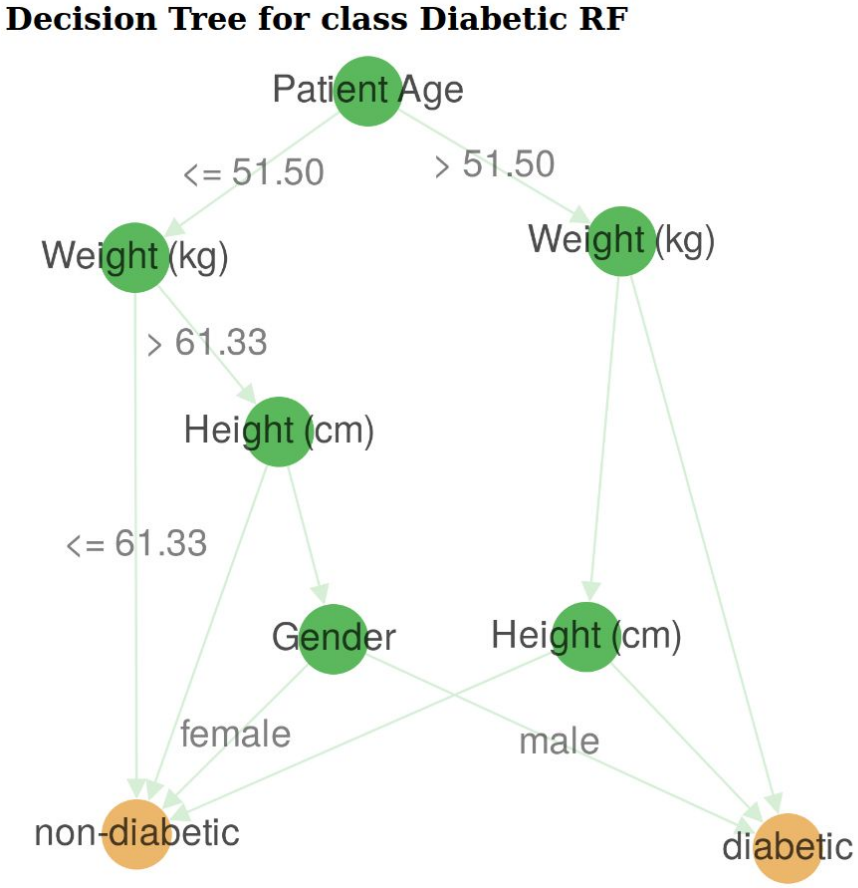
Ranges defined by the user (*e.g.* for $\beta = 3$)



2) ML models/Decision Trees (ID3)



2) ML models/Decision Trees (C4.5)



Datasets

Typical “raw” data

Format	Standard tabular
Data type	Continuous
Available dataset	Cardiovascular Imaging
Physical format	CSV

Typical “raw” data

Format	Standard tabular
Data type	Continuous
Available dataset	Cardiovascular Imaging
Physical format	CSV

Heart rate	Height (cm)	Weight (kg)	LVEDV (ml)	...
90	146.8412592	61.94370569	118.3623341	...
82	139.3567129	41.50638292	133.3962064	...
58	189.0026979	72.23268446	146.2021485	...
65	182.6839067	72.79223487	136.8615622	...
76	160.1900907	57.07295106	124.6417011	...
87	138.6691087	51.00651618	138.4764149	...

Semantically annotated data

Format	Set of terms
Data type	Categorical
Available dataset	MeSH (NLM / PubMed)
Physical format	JSON

Semantically annotated data

Format	Set of terms
Data type	Categorical
Available dataset	MeSH (NLM / PubMed)
Physical format	JSON

Anatomy [A] ⊕
Organisms [B] ⊕
Diseases [C] ⊖
- Bacterial Infections and Mycoses [C01] ⊕
- Virus Diseases [C02] ⊕
- Parasitic Diseases [C03] ⊕
- ...
- Digestive System Fistula [C06.267] ⊕
- Digestive System Neoplasms [C06.301] ⊕
- Gastrointestinal Diseases [C06.405] ⊕
- **Liver Diseases [C06.552]** ⊖
  - *alpha 1-Antitrypsin Deficiency [C06.552.074]*
  - Cholestasis, Intrahepatic [C06.552.150] ⊕
  - Chemical and Drug Induced Liver Injury [C06.552.195] ⊕
  - **Fatty Liver [C06.552.241]** ⊖
    - *Fatty Liver, Alcoholic [C06.552.241.390]*

Chemicals and Drugs [D] ⊕
Analytical, Diagnostic and Therapeutic Techniques, and Equipment [E] ⊕
...

Dataset type	Typical “raw” data	Semantically annotated
Format	Standard tabular	Set of terms
Data type	Continuous	Categorical
Available dataset	Cardiovascular Imaging	MeSH (NLM / PubMed)
Physical format	CSV	JSON

Cardiovascular Imaging

Heart rate	Height (cm)	Weight (kg)	LVEDV (ml)	LVESV (ml)	LVSV (ml)	...
90	146.8412592	61.94370569	118.3623341	37.69783342	80.66450068	...
82	139.3567129	41.50638292	133.3962064	41.75157147	91.64463497	...
58	189.0026979	72.23268446	146.2021485	46.46389429	99.73825425	...
65	182.6839067	72.79223487	136.8615622	38.83590358	98.02565863	...
76	160.1900907	57.07295106	124.6417011	36.94373843	87.69796263	...
87	138.6691087	51.00651618	138.4764149	37.89984229	100.5765726	...
105	172.9552102	83.90900534	118.4186404	41.07606853	77.34257187	...
84	174.2890575	68.04371317	160.7649009	52.59209694	108.172804	...
87	151.5365984	55.7666763	104.7862129	35.30753682	69.47867607	...
89	173.063218	92.23323664	139.9417832	39.21955081	100.7222324	...

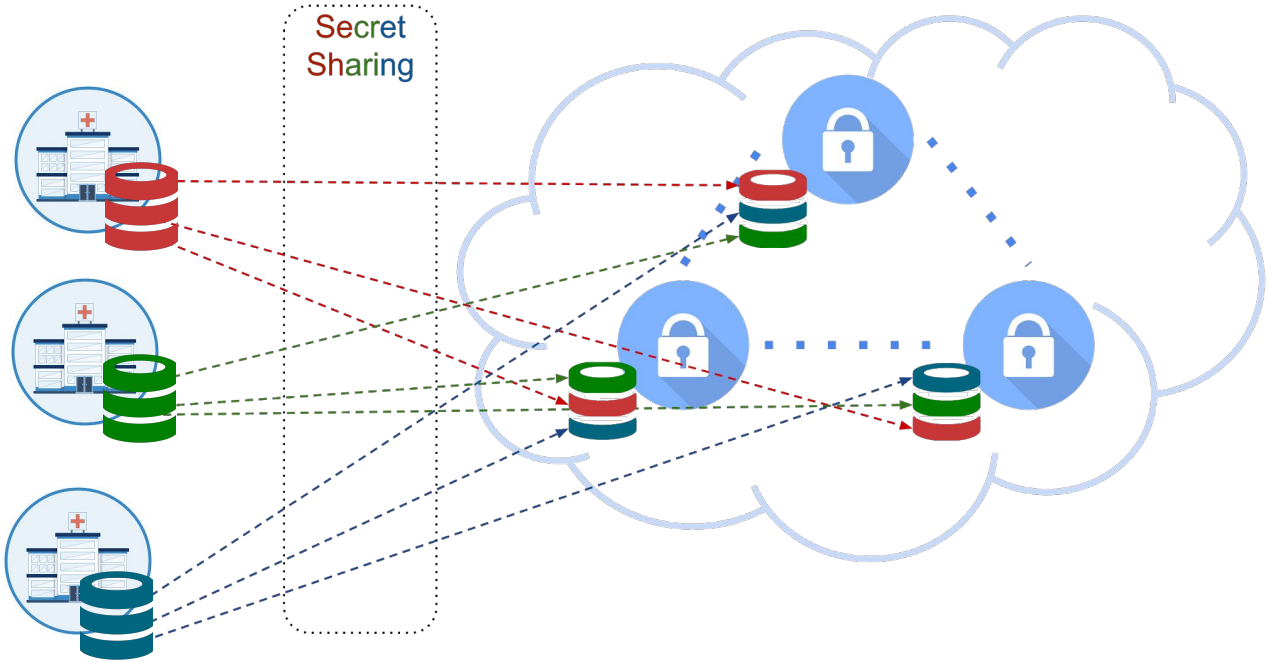
MeSH Tree

Anatomy [A] ⊕
Organisms [B] ⊕
Diseases [C] ⊖
- Bacterial Infections and Mycoses [C01] ⊕
- Virus Diseases [C02] ⊕
- Parasitic Diseases [C03] ⊕
- ...
- Digestive System Fistula [C06.267] ⊕
- Digestive System Neoplasms [C06.301] ⊕
- Gastrointestinal Diseases [C06.405] ⊕
- **Liver Diseases [C06.552]** ⊖
  - *alpha 1-Antitrypsin Deficiency [C06.552.074]*
  - Cholestasis, Intrahepatic [C06.552.150] ⊕
  - Chemical and Drug Induced Liver Injury [C06.552.195] ⊕
  - **Fatty Liver [C06.552.241]** ⊖
    - *Fatty Liver, Alcoholic [C06.552.241.390]*

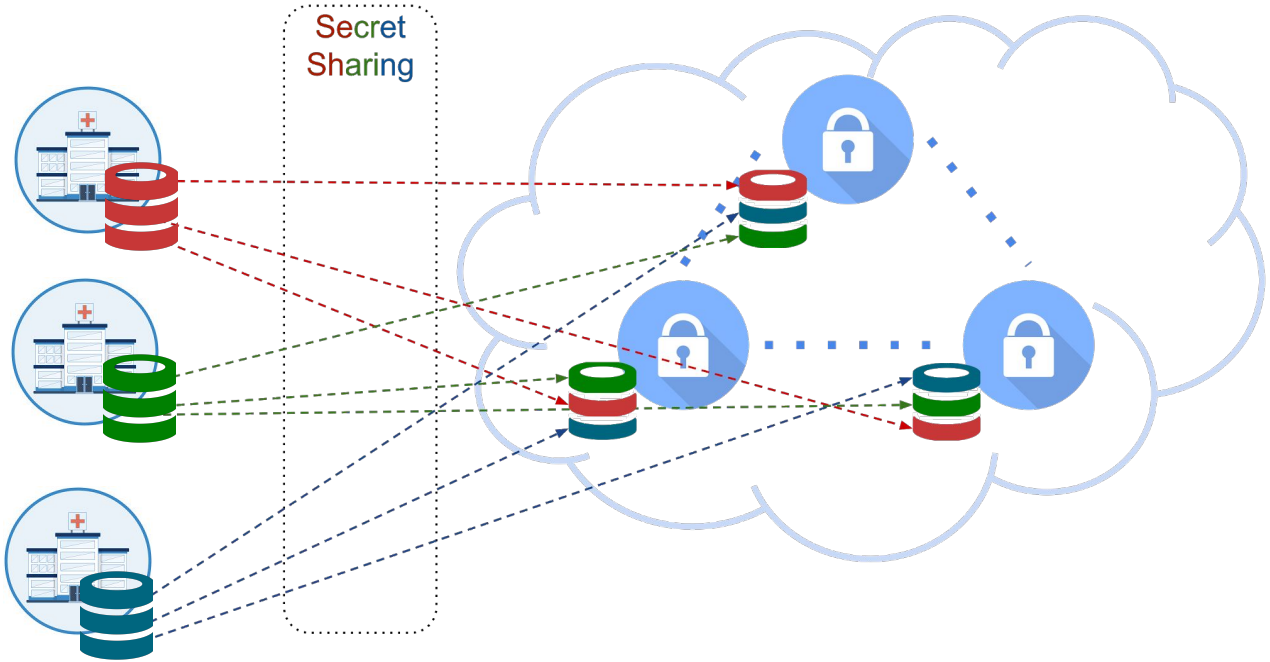
Chemicals and Drugs [D] ⊕
Analytical, Diagnostic and Therapeutic Techniques, and Equipment [E] ⊕
...

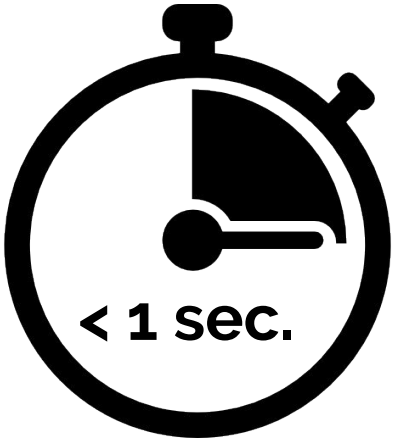
Evaluations

0) Importing to MPC-Cloud – Secret-sharing

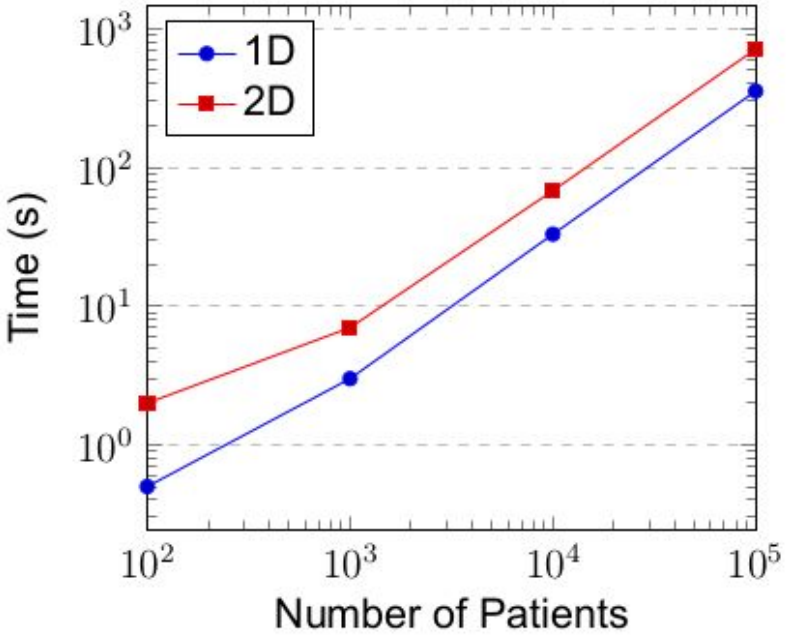


0) Importing to MPC-Cloud – Secret-sharing

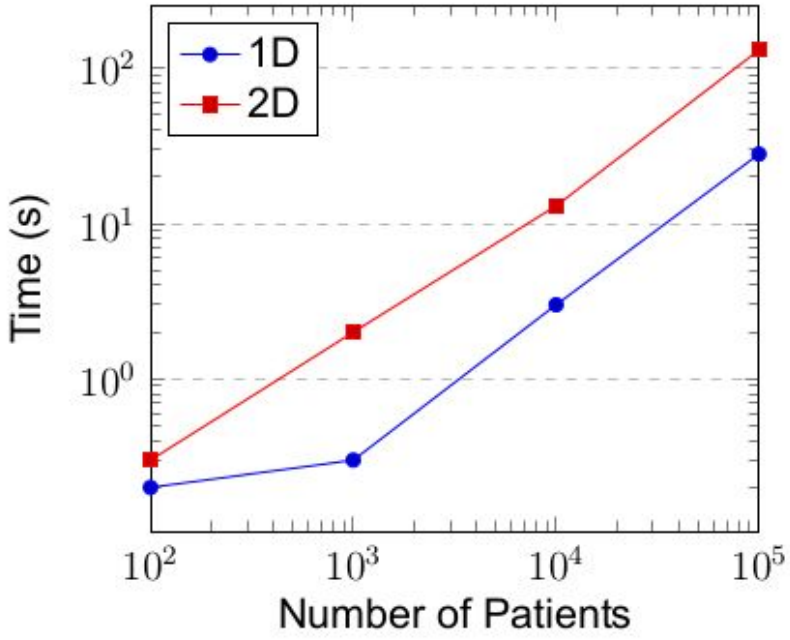




1) Aggregators (1D & 2D)

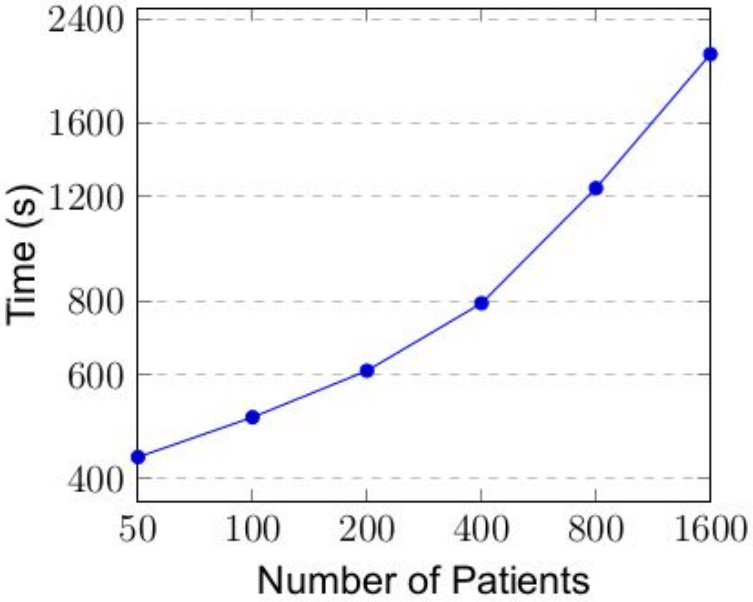


Continuous Histogram Timings

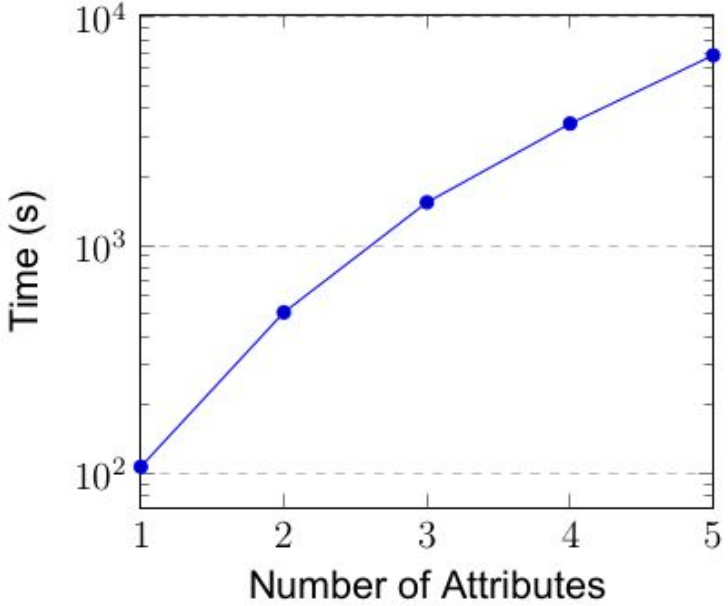


Categorical Histogram Timings

2) ML models/Decision Trees (ID3)



ID3 Timings with variable patients for 3 attributes



ID3 Timings with variable number of attributes

Demo

https://mhmd.madgik.di.uoa.gr/

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Summary

- **End-to-end** infrastructure: Provide meaningful result to doctors, researchers and patients
- User & Programming **interface**
- **Dynamic deployment** of hospitals
- Mathematically **guaranteed privacy** against semi-honest adversaries
- Development of **essential privacy-preserving algorithms** in the MPC setting
- Dataset **flexibility** (categorical & continuous)

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Future Work

- Optimize decision-tree classifiers by computing **sufficient statistics**
- Develop more **sophisticated algorithms** and ML models
  e.g. Stochastic Gradient Descent or Deep Neural Networks
- Replace Sharemind **MPC engine** with SPDZ, FRESCO, etc
- Add **Differential Privacy** to ensure output privacy of algorithms
- Incorporate a **Blockchain** infrastructure for transparency and auditing of requested computations.
- Add **Zero-Knowledge Proofs** to ensure correct computation.

Thank you!!

https://github.com/Athena-MHMD/smpc-analytics

ATHENA

Questions?

Thank you!!

https://github.com/Athena-MHMD/smpc-analytics

Semantic annotation to Tabular

{ Temporomandibular Joint Dysfunction Syndrome,

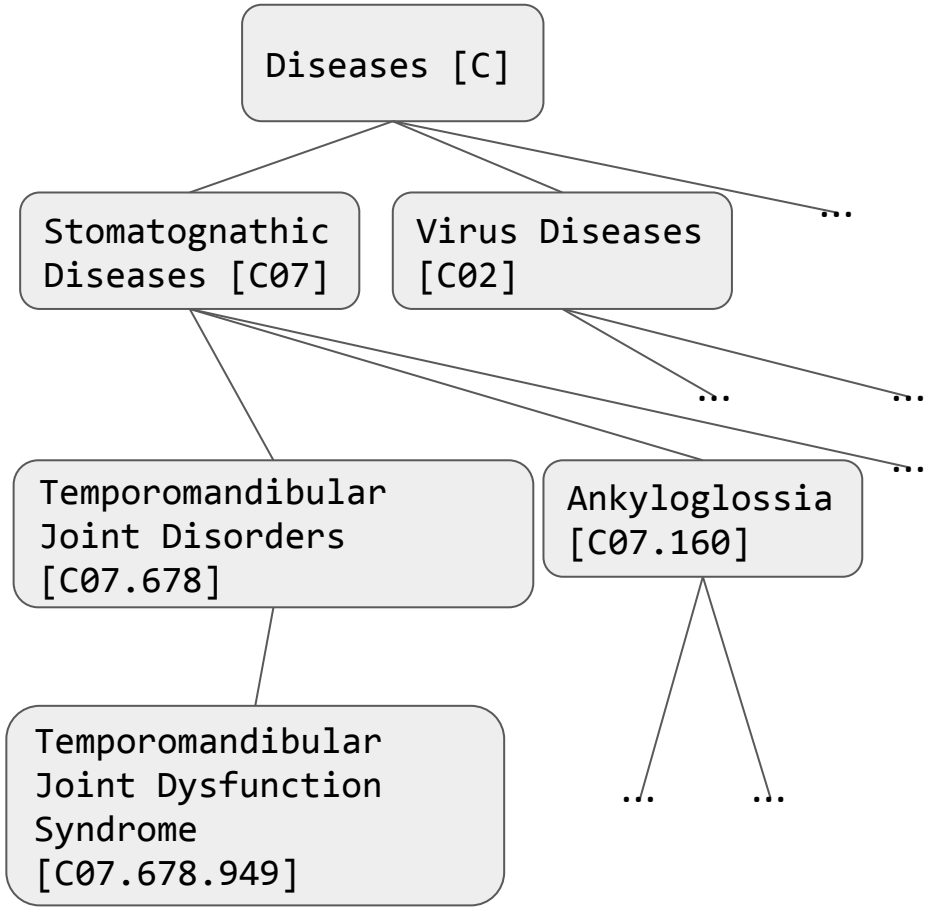
Liver Neoplasms, Adult, Alcoholics, Asian Americans }

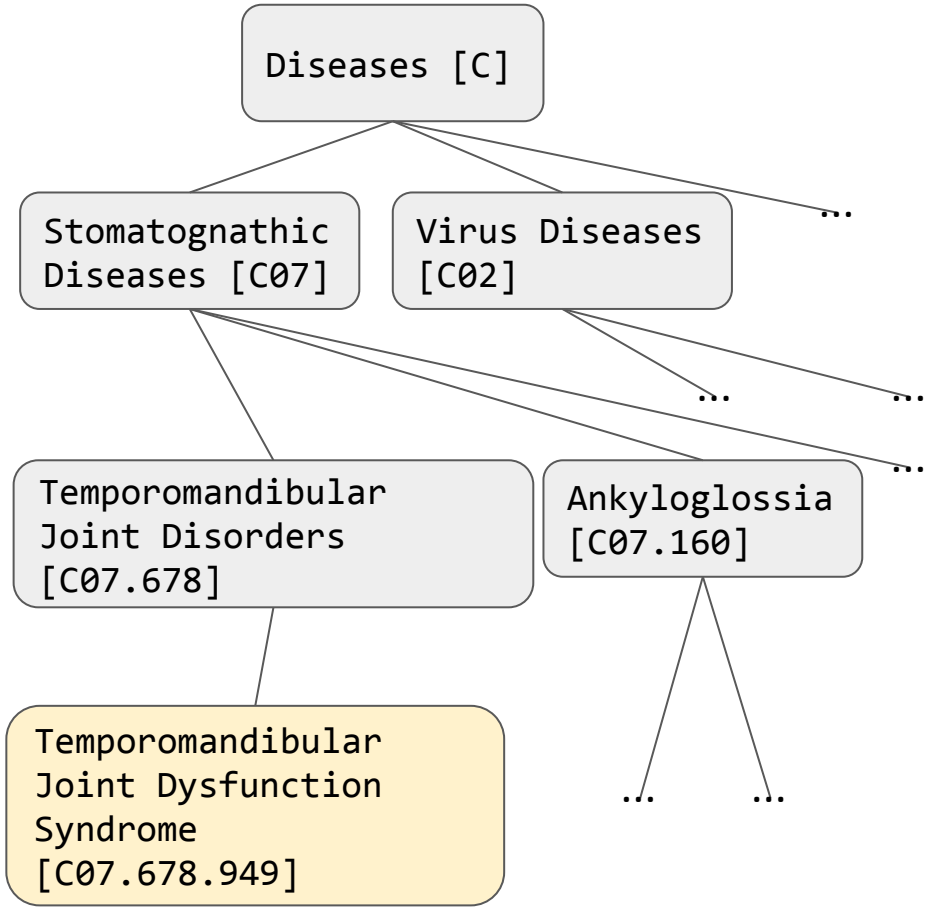
MeSH terms

{ Temporomandibular Joint Dysfunction Syndrome,

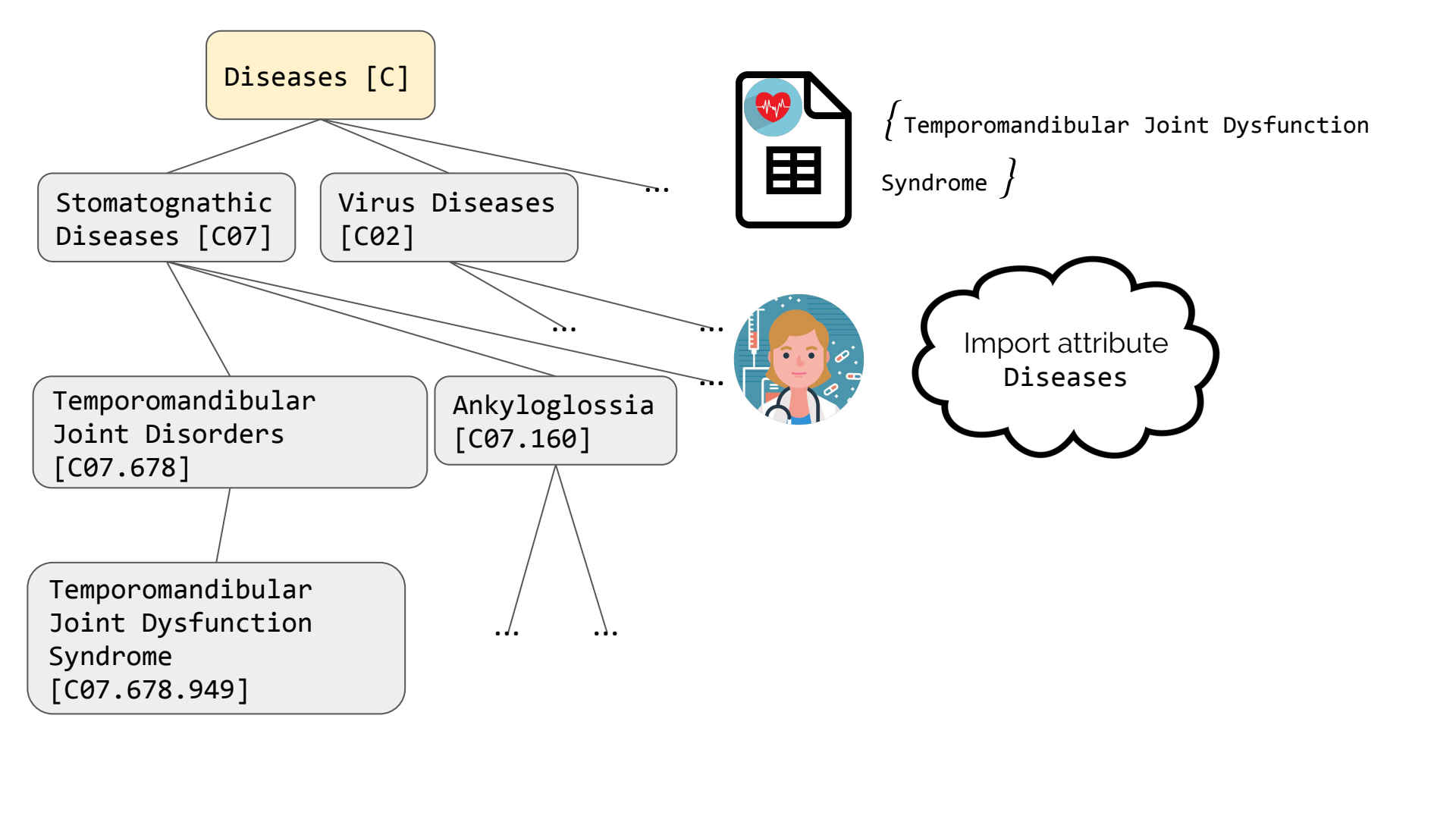
Liver Neoplasms, Adult, Alcoholics, Asian Americans }

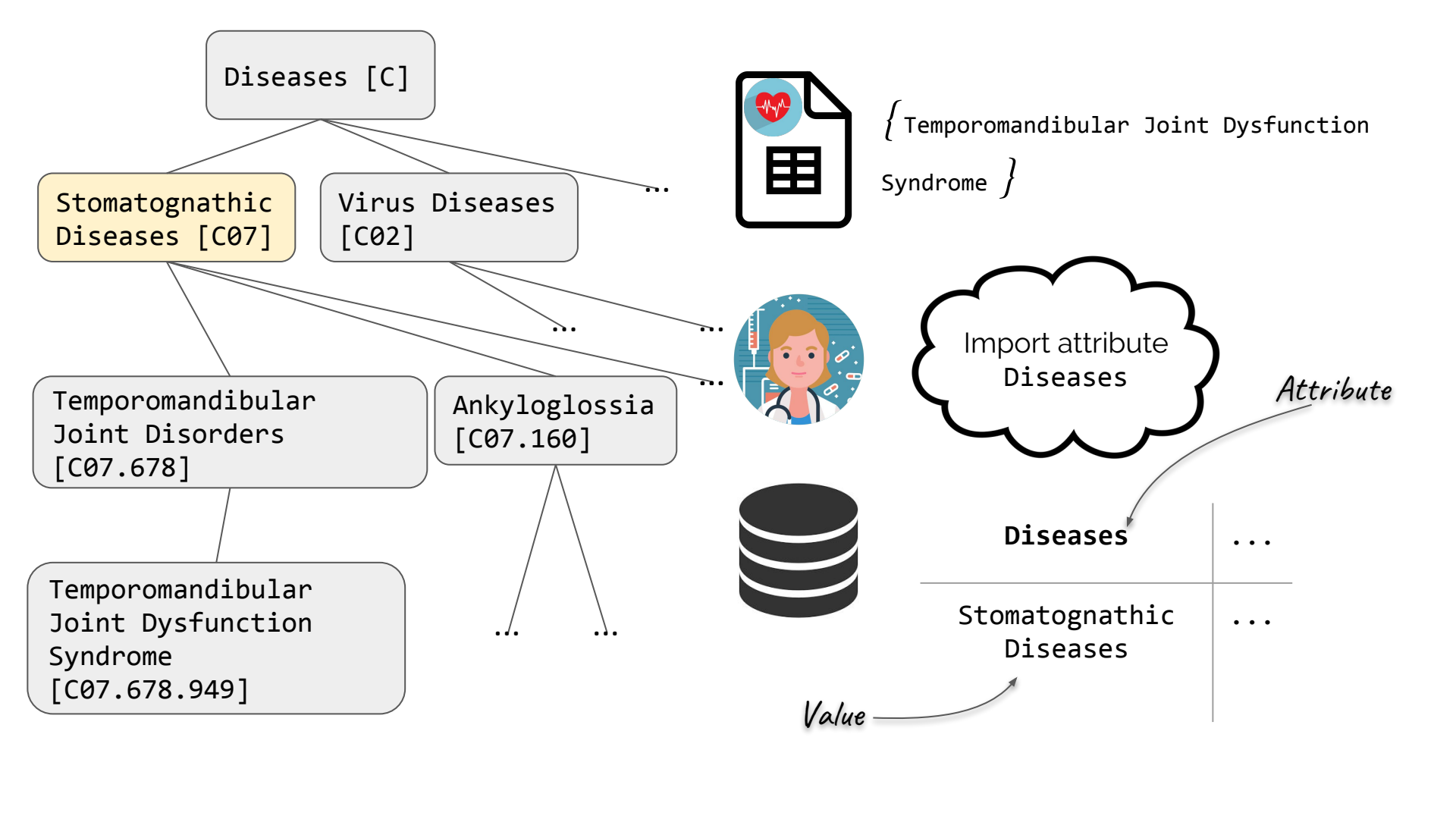
MeSH terms

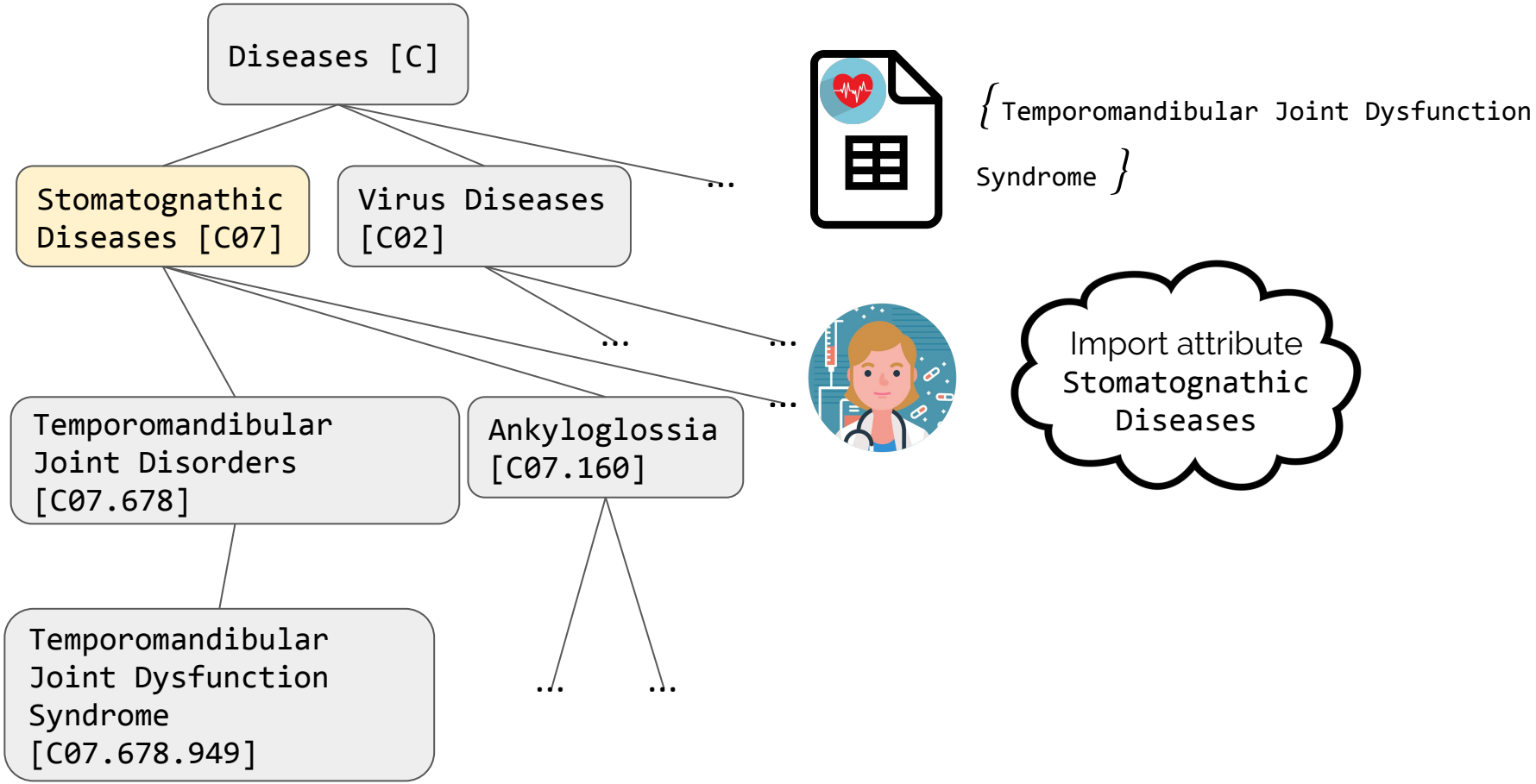

Diseases [C]
Stomatognathic Diseases [C07]
Virus Diseases [C02]
...
...
...
...
Temporomandibular Joint Disorders [C07.678]
Ankyloglossia [C07.160]
Temporomandibular Joint Dysfunction Syndrome [C07.678.949]
...
...

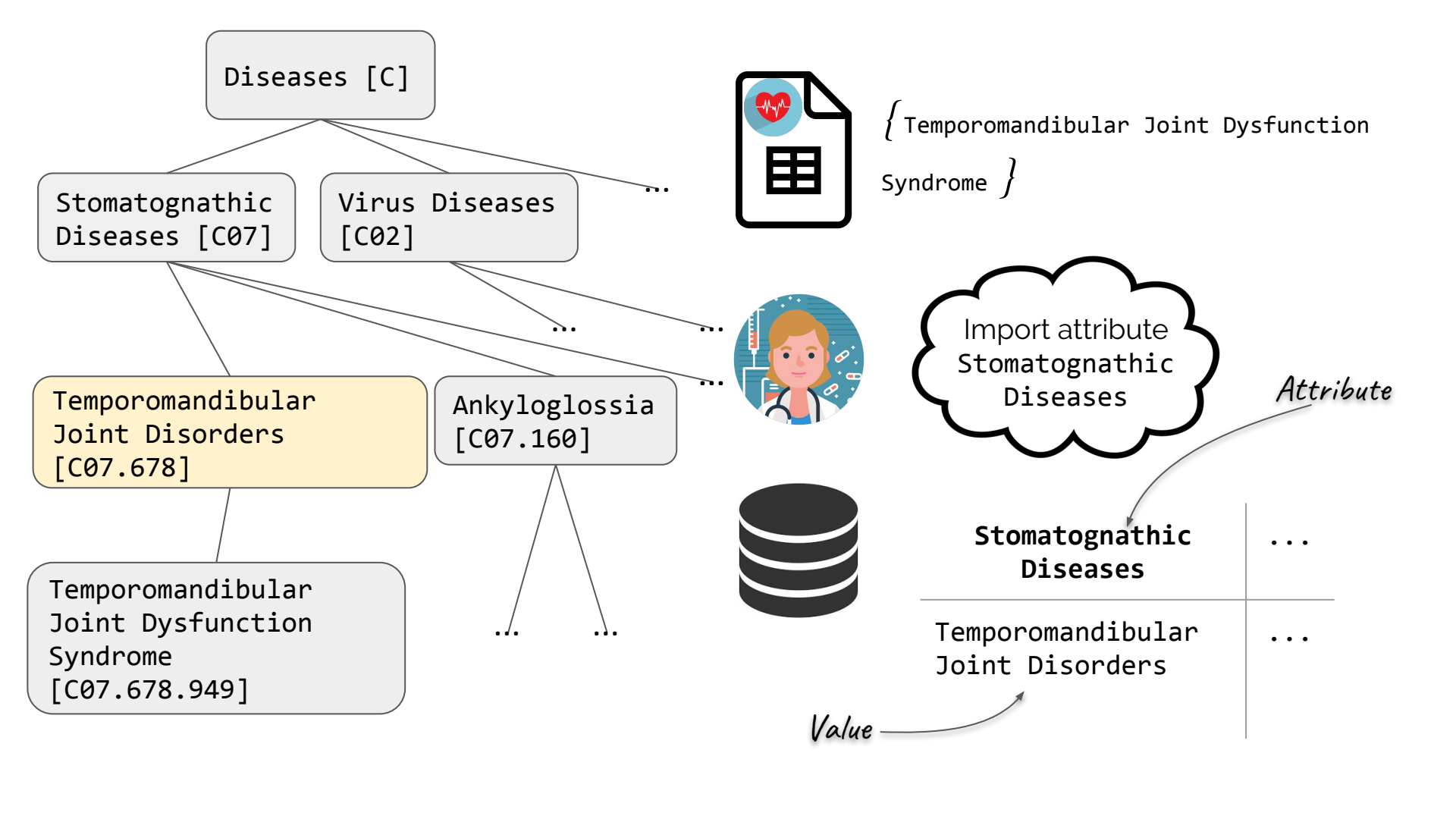


{ Temporomandibular Joint Dysfunction Syndrome }


Diseases [C]
Stomatognathic Diseases [C07]
Virus Diseases [C02]
...
Temporomandibular Joint Disorders [C07.678]
Ankyloglossia [C07.160]
Temporomandibular Joint Dysfunction Syndrome [C07.678.949]
{ Temporomandibular Joint Dysfunction Syndrome }
Import attribute Diseases


Diseases [C]
Stomatognathic Diseases [C07]
Virus Diseases [C02]
...
Temporomandibular Joint Disorders [C07.678]
Ankyloglossia [C07.160]
Temporomandibular Joint Dysfunction Syndrome [C07.678.949]
{ Temporomandibular Joint Dysfunction Syndrome }
Import attribute Diseases
Attribute
Diseases
...
Stomatognathic Diseases
Value


Diseases [C]
Stomatognathic Diseases [C07]
Virus Diseases [C02]
...
Temporomandibular Joint Disorders [C07.678]
Ankyloglossia [C07.160]
Temporomandibular Joint Dysfunction Syndrome [C07.678.949]
{ Temporomandibular Joint Dysfunction Syndrome }
Import attribute Stomatognathic Diseases


Diseases [C]
Stomatognathic Diseases [C07]
Virus Diseases [C02]
...
Temporomandibular Joint Disorders [C07.678]
Ankyloglossia [C07.160]
Temporomandibular Joint Dysfunction Syndrome [C07.678.949]
{ Temporomandibular Joint Dysfunction Syndrome }
Import attribute Stomatognathic Diseases
Attribute
Stomatognathic Diseases
...
Temporomandibular Joint Disorders
...
Value