SecretSVM: Privacy-Preserving Support Vector Machine with IoT

Yu-Chi Chen
Dept. CSE @ YZU

(joint work with Song-Yi Hsu and Xin Xie)

元智大學
Yuan Ze University

Outline

- Introduction
- Building Blocks
- Our System, SecretSVM
- Experiments and Conclusions

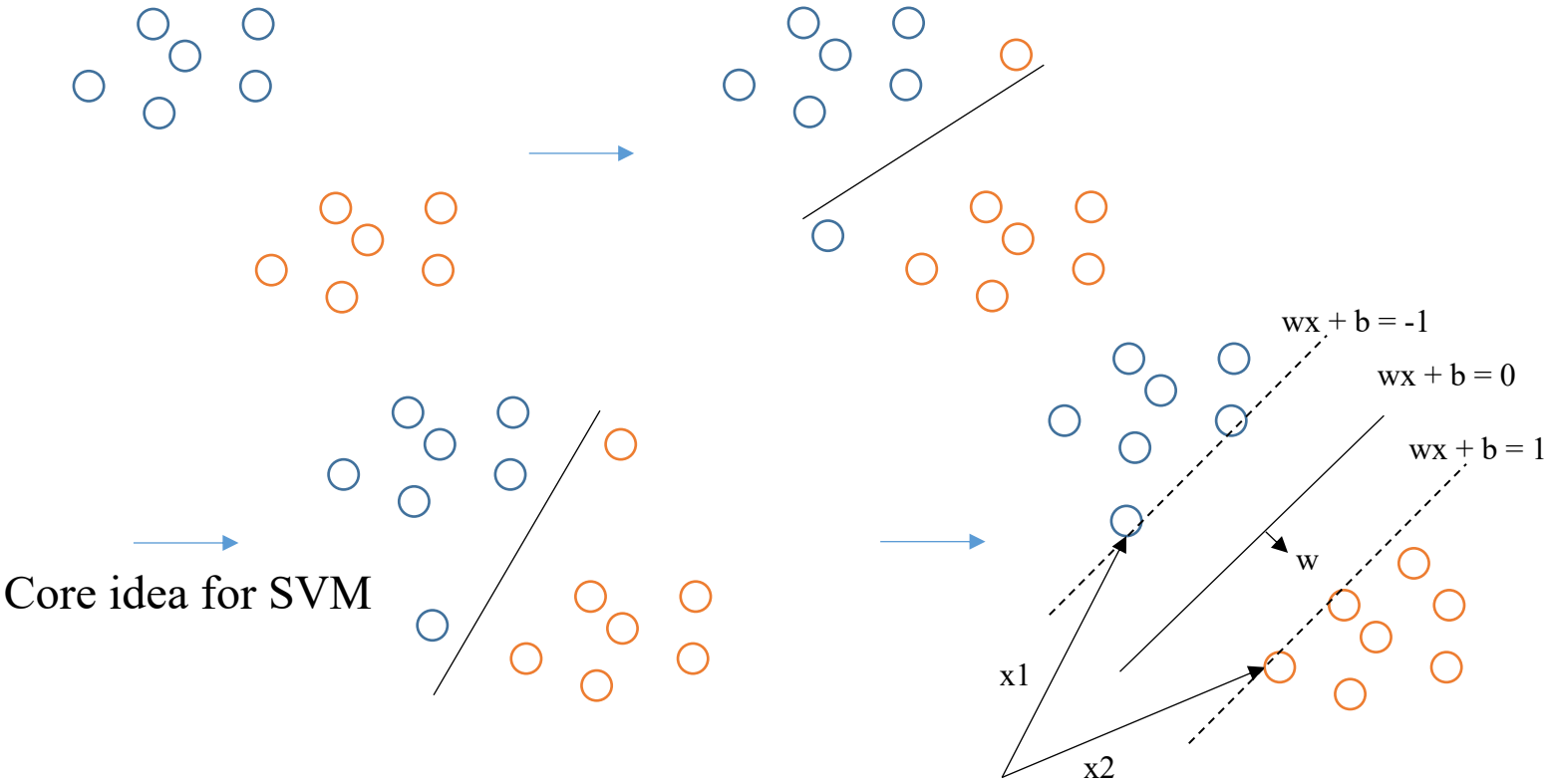
Warm-up

Some materials in this slide are credited by *Demon Slayer*

Nice Source

39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K
52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K
31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K
42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States	>50K
37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80	United-States	>50K
30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India	>50K
23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States	<=50K
32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-States	<=50K
40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?	>50K
34	Private	245487	7th-8th	4	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0	0	45	Mexico	<=50K
25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States	<=50K
32	Private	186824	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Male	0	0	40	United-States	<=50K
38	Private	28887	11th	7	Married-civ-spouse	Sales	Husband	White	Male	0	0	50	United-States	<=50K
43	Self-emp-not-inc	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States	>50K
40	Private	193524	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	60	United-States	>50K
54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0	20	United-States	<=50K
35	Federal-gov	76845	9th	5	Married-civ-spouse	Farming-fishing	Husband	Black	Male	0	0	40	United-States	<=50K
43	Private	117037	11th	7	Married-civ-spouse	Transport-moving	Husband	White	Male	0	2042	40	United-States	<=50K
59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0	40	United-States	<=50K

Support Vector Machine (SVM)



Core idea for SVM

The optimization problem of SVM as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \; s.t, y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, m.$$

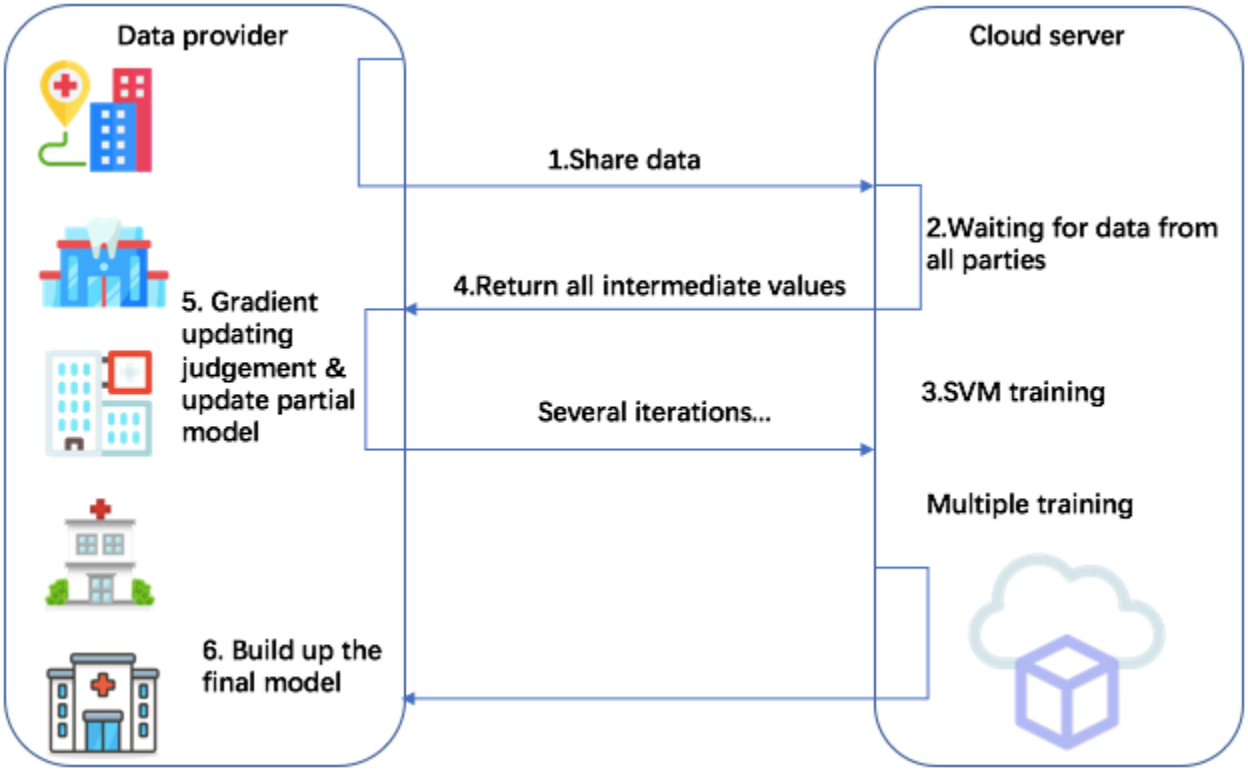
Classification is not always linear!

Core Condition in SVM

The optimization problem of SVM as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \; s.t. \; y_i(w^T x_i + b) \geq 1, \; i = 1, 2, \ldots, m.$$

Move to Cloud Computing Era



Recap of Issues from Literature!

Achieving
data privacy

Encryption

~~Homomorphic encryption~~

Secret sharing

Our proposal:

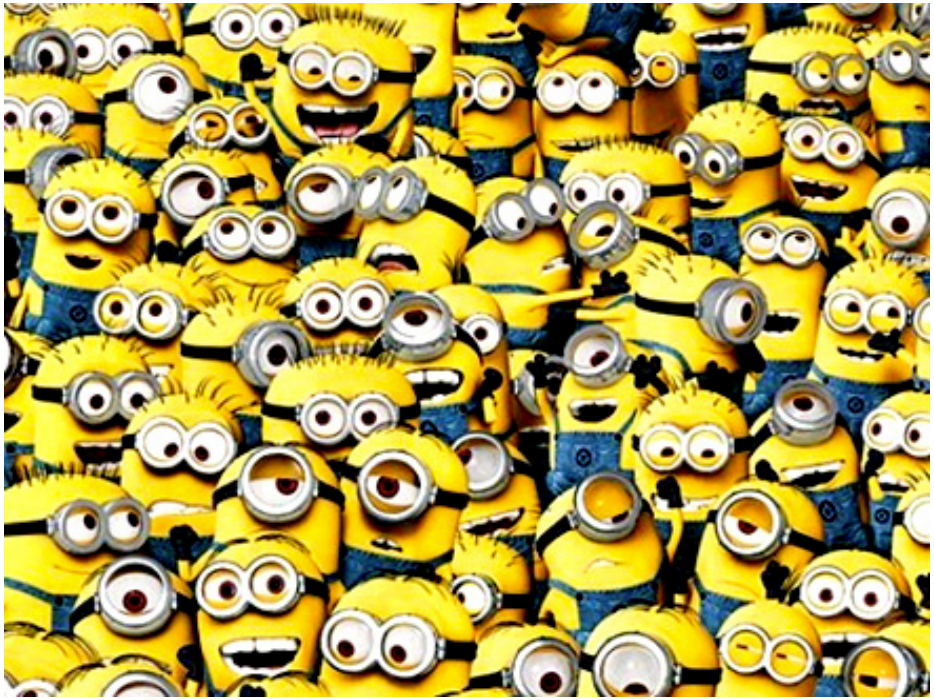
Preventing Collusion
(Cloud and Adversary)

~~Two Server~~

Many nodes jointly obtains
the model

in our IoT scenario

The truth is ...



Big Problem in the Proposal

Achieving
data privacy

→

Encryption

~~Homomorphic encryption~~

Secret sharing

Our proposal:

Preventing Collusion
(Cloud and Adversary)

→

~~Two Server~~

Many nodes jointly obtains
the model

Very Lucky!

It’s not the first time to emulate a server among distributed nodes

Proof-of-Work (in Bitcoin)

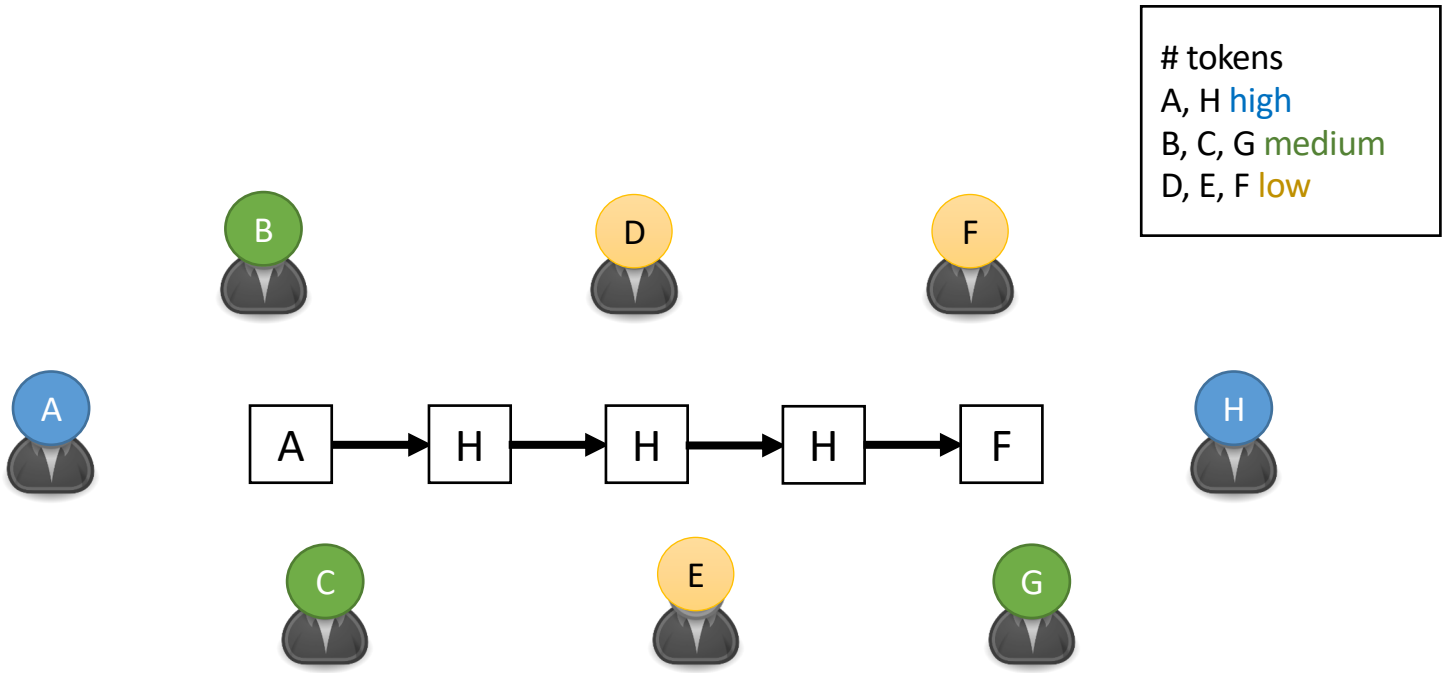
Competition

One winner → others waste energy

Competition and PoW

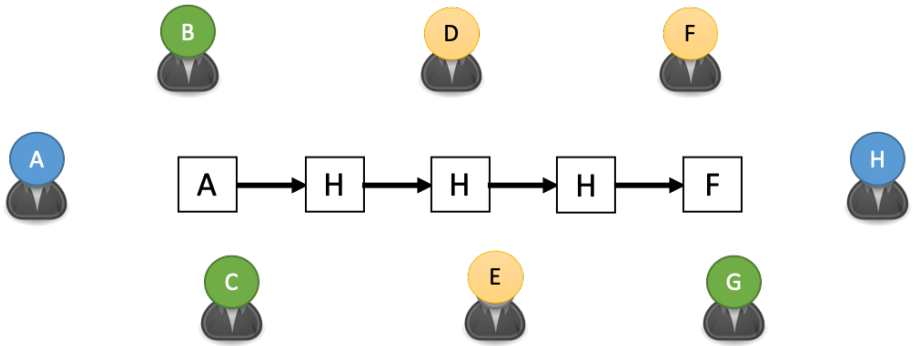
Alternative: Proof-of-Stake

Blocks are “mined” according to the amount of “tokens” he/she holds



Quick Question?

Which one is better in IoT?



Outline

- Introduction
- Building Blocks
- Our System, SecretSVM
- Experiments and Conclusions

Secret Sharing

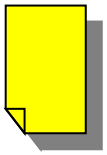
Dealer

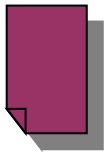
Takes a secret as input, and generates n shares for n parties, such that any k shares (k <n) can recover the original secret.

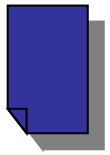
Share1

Share2

Share3





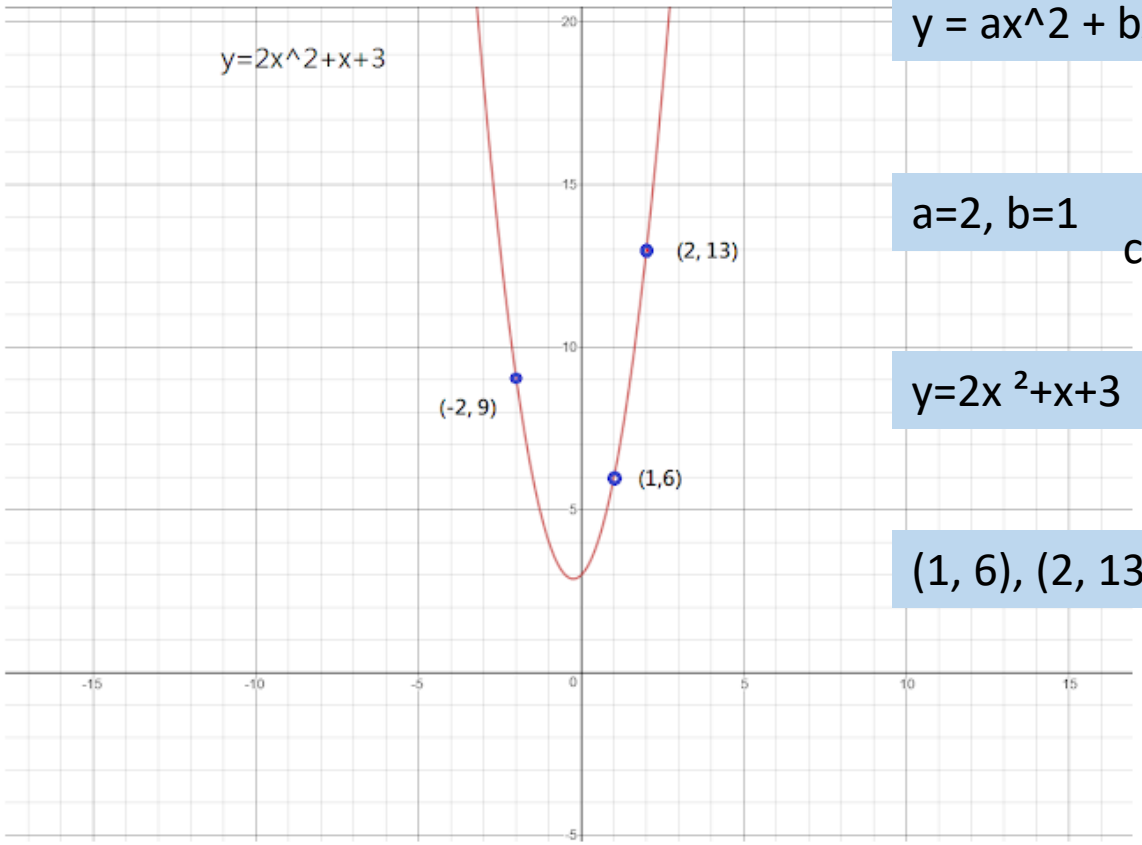


Secret Sharing with k = 3 & n = 3

(1, 6)

(2, 13)

(-2, 9)



y = ax^2 + bx + c

a=2, b=1

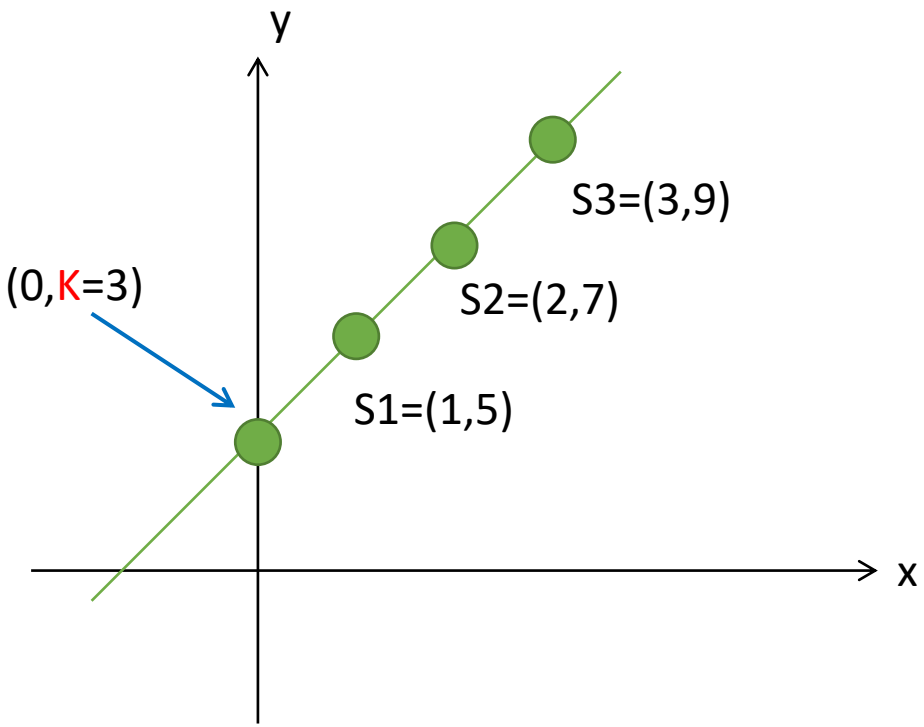
c = 3 is our secret!

y=2x ²+x+3

(1, 6), (2, 13), (-2, 9)

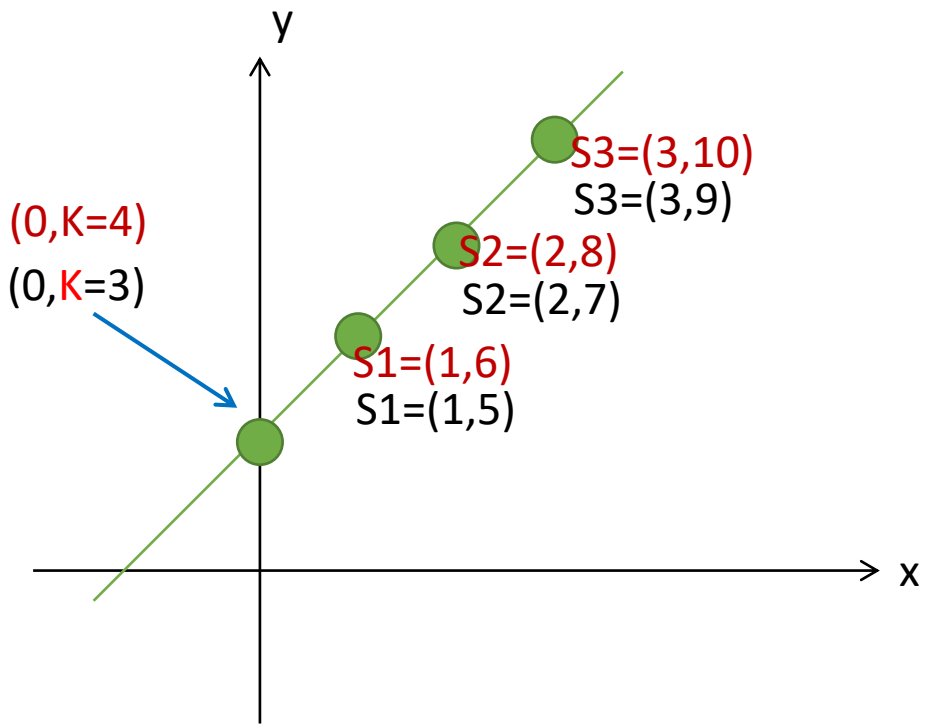
How about k=2 & n = ...

Secret Sharing with k = 2 & n = …



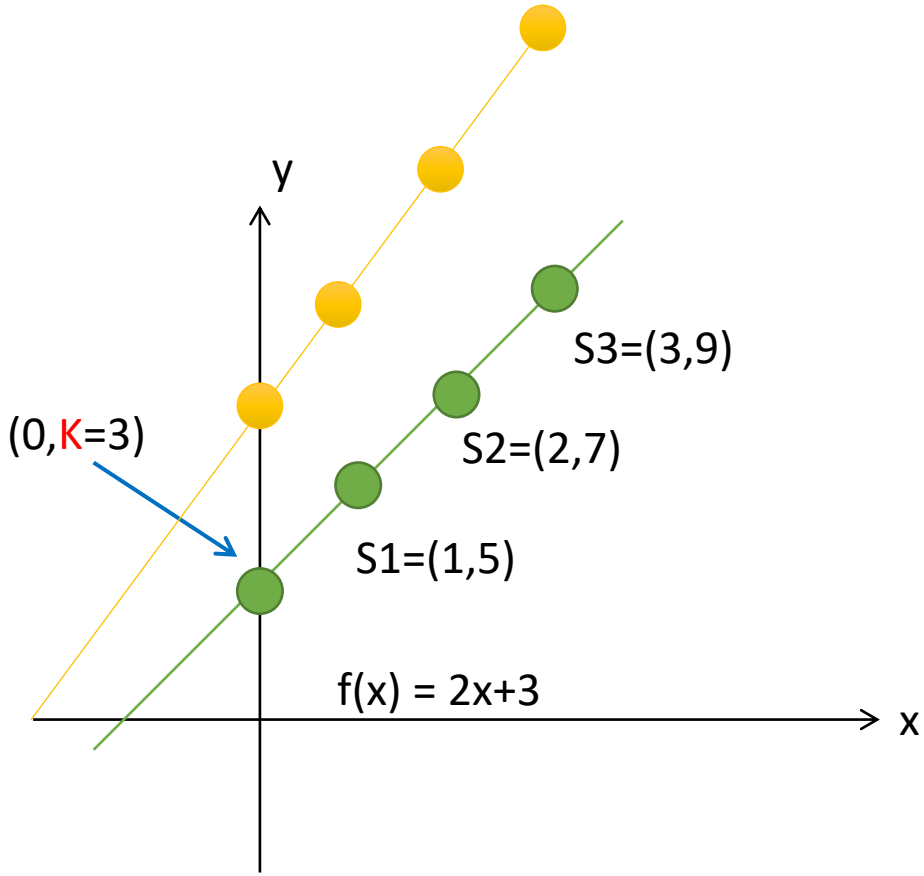
For more information, please contact Prof. Tso

If Operating over Shares?



Exercise 1: Add 1 in all shares

If Operating over Shares?



S3=(3,18)

S2=(2,14)

S1=(1,10)

(0,K=?)

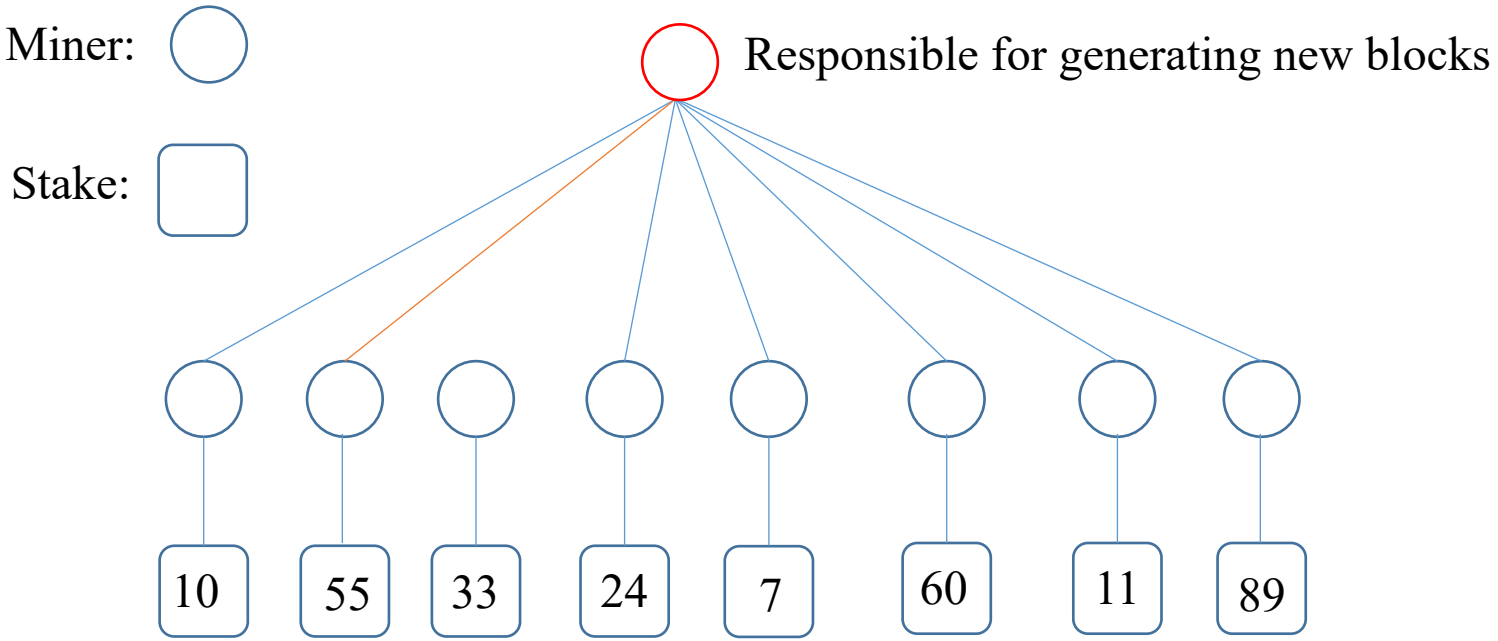
f'(x) = 4x+6

Exercise 2: Multiply 2 in all shares

PoS Consensus

Advantage:

1. Decentralization
2. Save energy



Outline

- Introduction
- Building Blocks
- Our System, SecretSVM
- Experiments and Conclusions

Our System

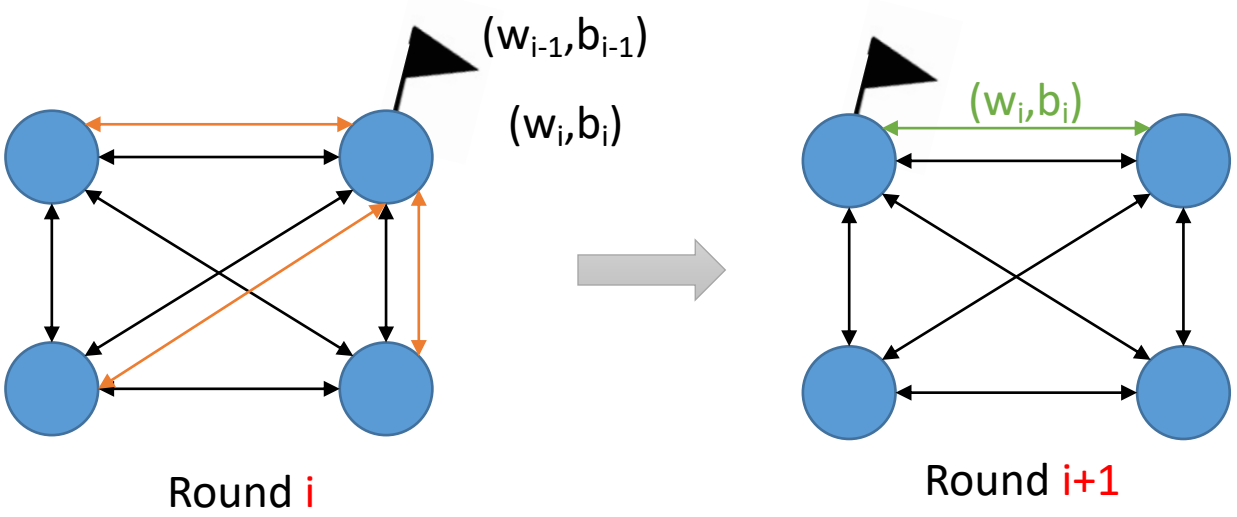
IoT data provider

Leader

SVM training protocol

PoS protocol (finding a leader at round i)

Message delivery protocol (between leaders at rounds i and i+1)



Collusion versus Decentralization

High Level Picture of SecretSVM

(x, y)

Node (user)

SVM training protocol

(w, b)

Leader

小弟們跟著我走！
豬突猛進

The optimization problem of SVM as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \; s.t, y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, m.$$

Addition over (2,2)-Secret Sharing

其實就是兩條多項式相加而已！ 但是真正在做的時候 多項式都會mod一個質數

a = 2

secret: a

b = 1

secret: b

$[b]_1=(1,4)$

$[b]_1$

$[a]_2$

$[a]_2=(2,4)$

Generate $[a]_1$, $[a]_2$ $[a]_1=(1,3)$, $[a]_2=(2,4)$

Obtain $[a+b]_1$

$[a+b]_1=(1,3+4) = (1,7)$

Generate $[b]_1$, $[b]_2$

Obtain $[a+b]_2$

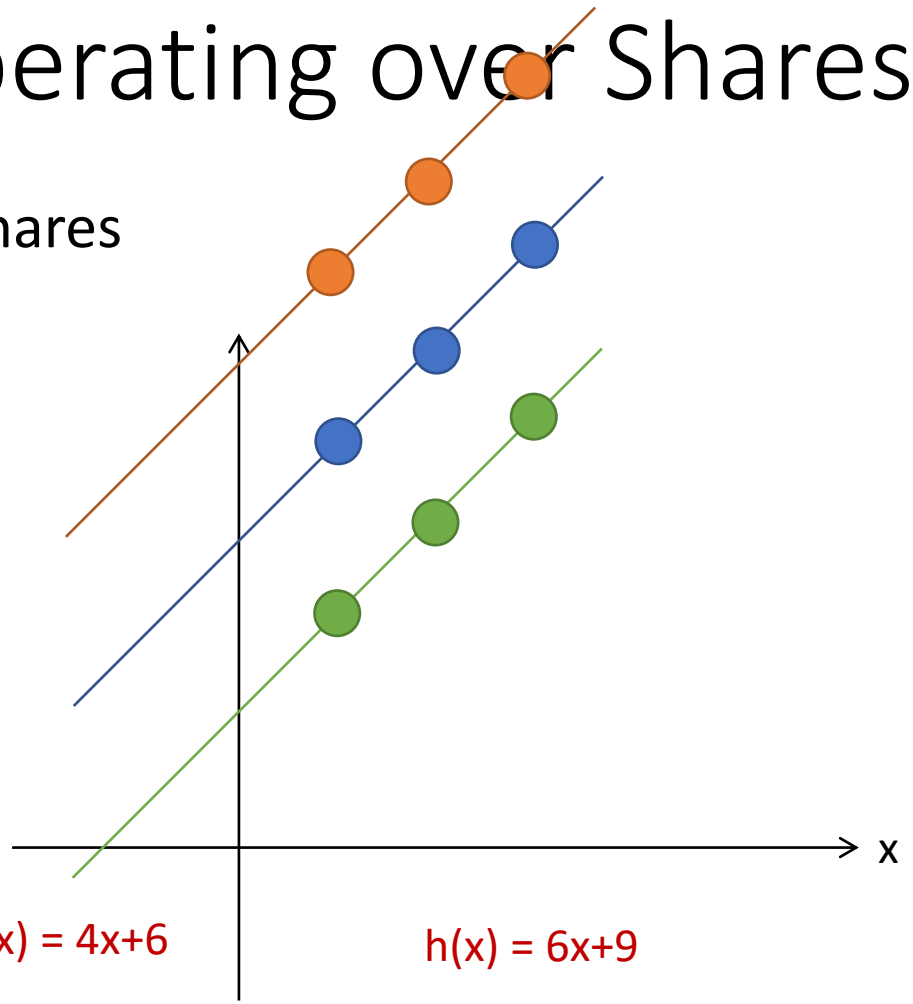
$[b]_1=(1,4)$, $[b]_2=(2,7)$

f'(x) = 4x+3

$[a+b]_2=(2,4+7) = (2,11)$

If Operating over Shares?

Exercise 3: Add two shares



f(x) = 2x+3	g(x) = 4x+6	h(x) = 6x+9
S3=(3,9)	S3=(3,18)	S3=(3,27)
S2=(2,7)	S2=(2,14)	S2=(2,21)
S1=(1,5)	S1=(1,10)	S1=(1,15)

Multiplication?

secret: a

$[a]_1, [b]_1,$
$[r]_1, [r']_1, [rr']_1$

Compute $[a-r]_1$ & $[b-r']_1$

Obtain $e = a - r$ & $p = b - r'$

$[a-r]_1$
$[b-r']_1$

$[a-r]_2$
$[b-r']_2$

secret: b

$[a]_2, [b]_2,$
$[r]_2, [r']_2, [rr']_2$

Compute $[a-r]_2$ & $[b-r']_2$

Obtain $e = a - r$ & $p = b - r'$

$$[ab]_1 = [rr']_1 + e[r']_1 + p[r]_1 + ep$$

$$= [rr']_1 + [(a-r)r']_1 + [(b-r')r]_1 + [(b-r')(a-r)]_1$$
$$= [rr' + (a-r)r' + (b-r')r + (b-r')(a-r)]_1$$

$$[ab]_2 = [rr']_2 + e[r']_2 + p[r]_2 + ep$$

Very easy to verify this by Exercises 1, 2, and 3

$y(wx + b) \geq 1$

Comparison

input: $[a]_1, [b]_1$

Goal: $a > b$ or $\leq$
where $a, b < \ell$

$[c]_2, [d]_2$

input: $[a]_2, [b]_2$

Pick random c, d $(2\ell c + d < q)$

Generate $[c]_1$, $[c]_2$ & $[d]_1$, $[d]_2$

Key point!

$[s]_2 = [a - b + \ell]_2 * [c]_2 + [d]_2$
$[h]_2 = [c]_2 + [d]_2$

$[s]_1 = [a - b + \ell]_1 * [c]_1 + [d]_1$
$[h]_1 = [c]_1 + [d]_1$

$[s]_2, [h]_2$

Reconstruct s, h
If $s > h$, then $a > b$

$>$ or $\leq$

Randomness from?

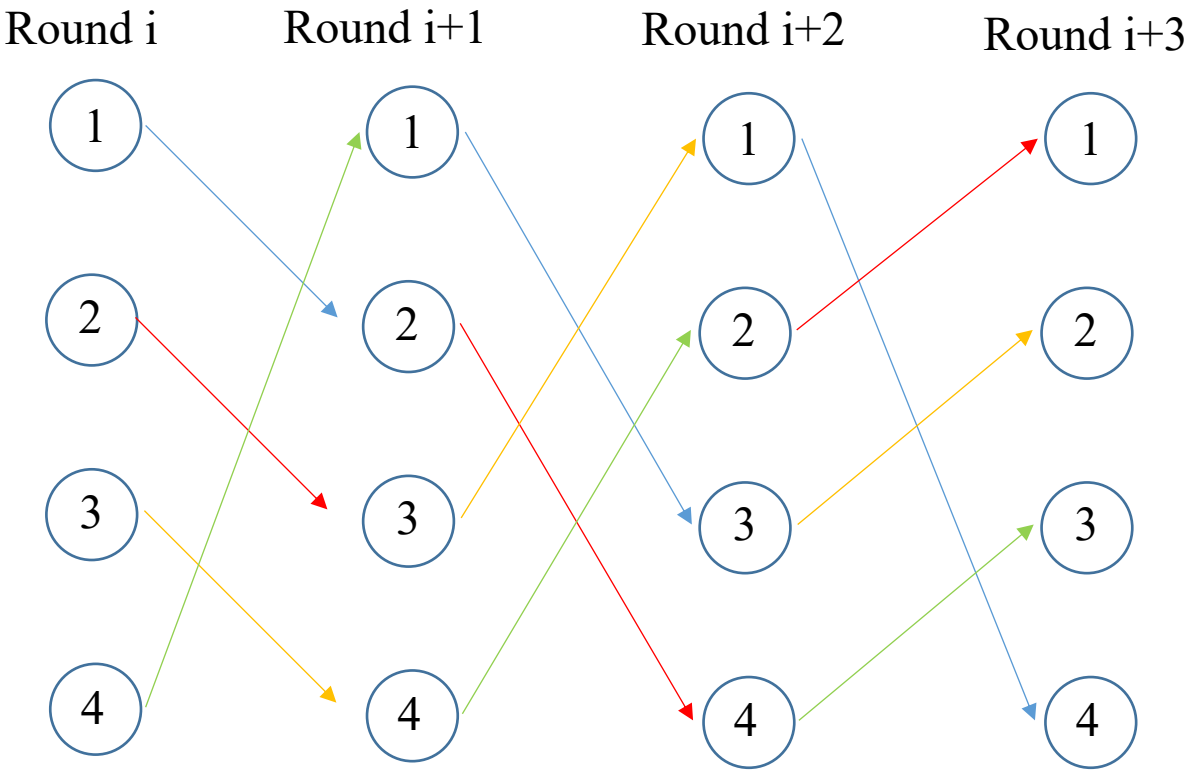
secret: a

$[a]_1, [b]_1,$
$[r]_1, [r']_1, [rr']_1$

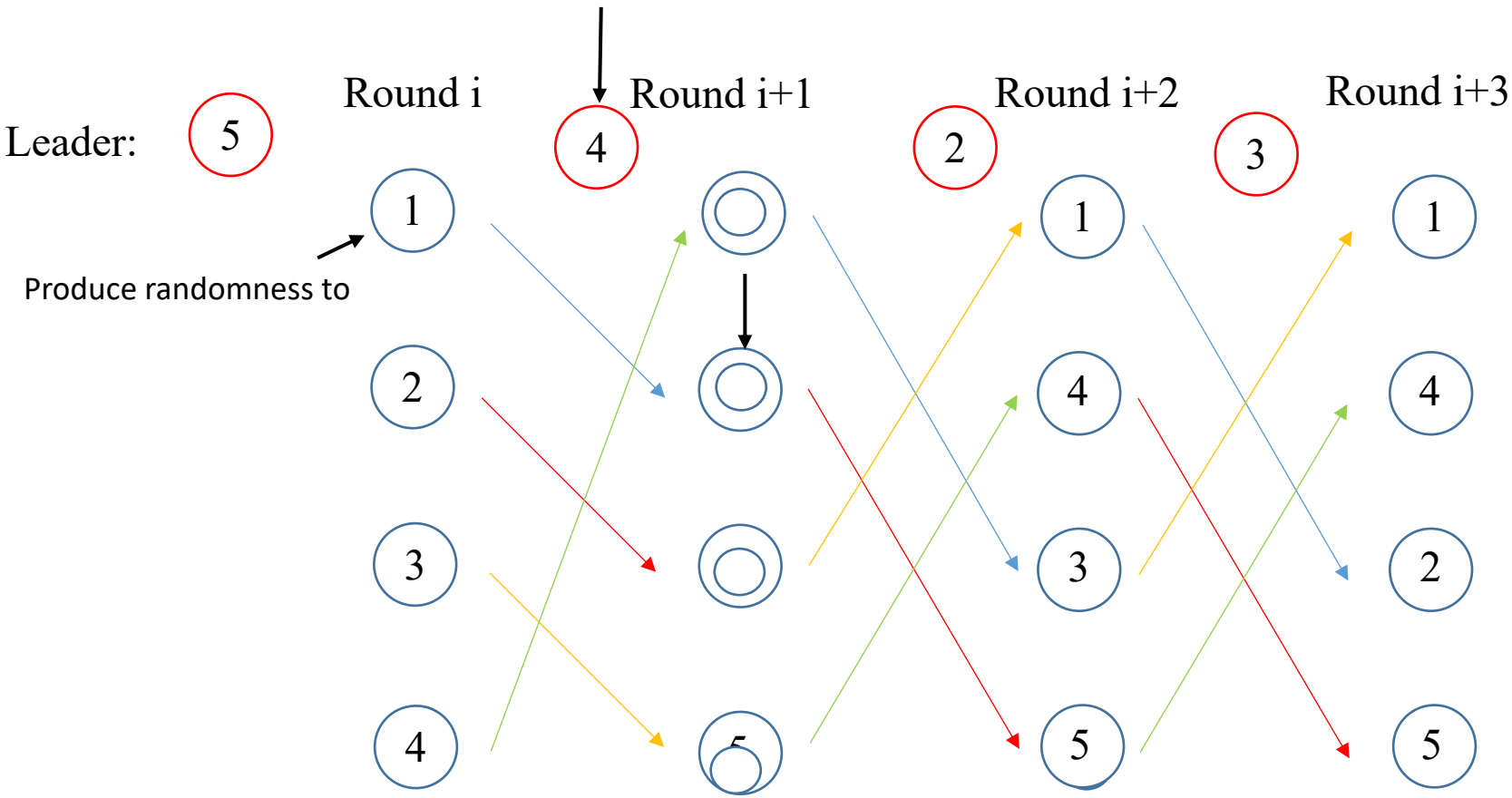
secret: b

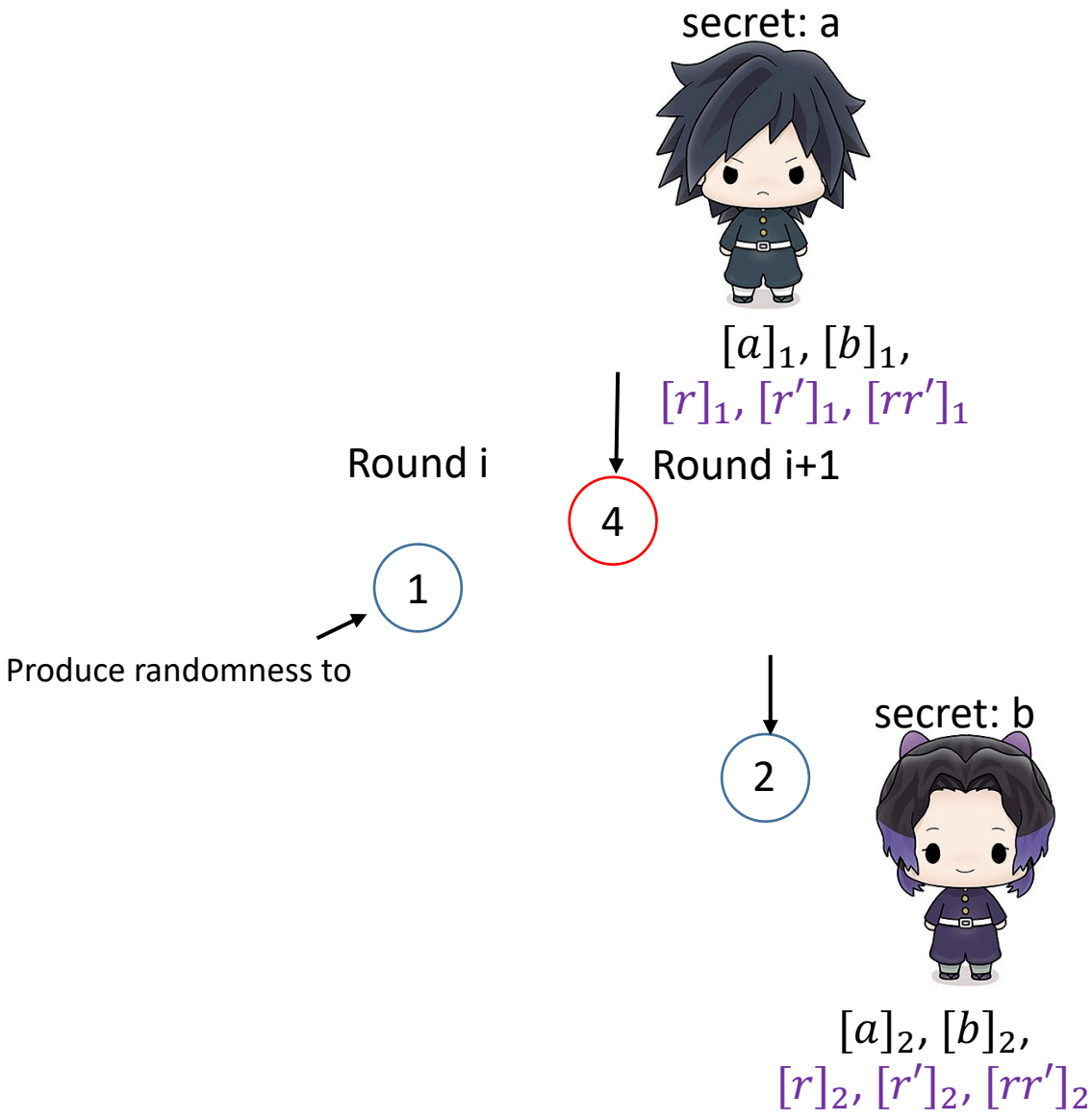
$[a]_2, [b]_2,$
$[r]_2, [r']_2, [rr']_2$

(Somewhat) Sorting Network

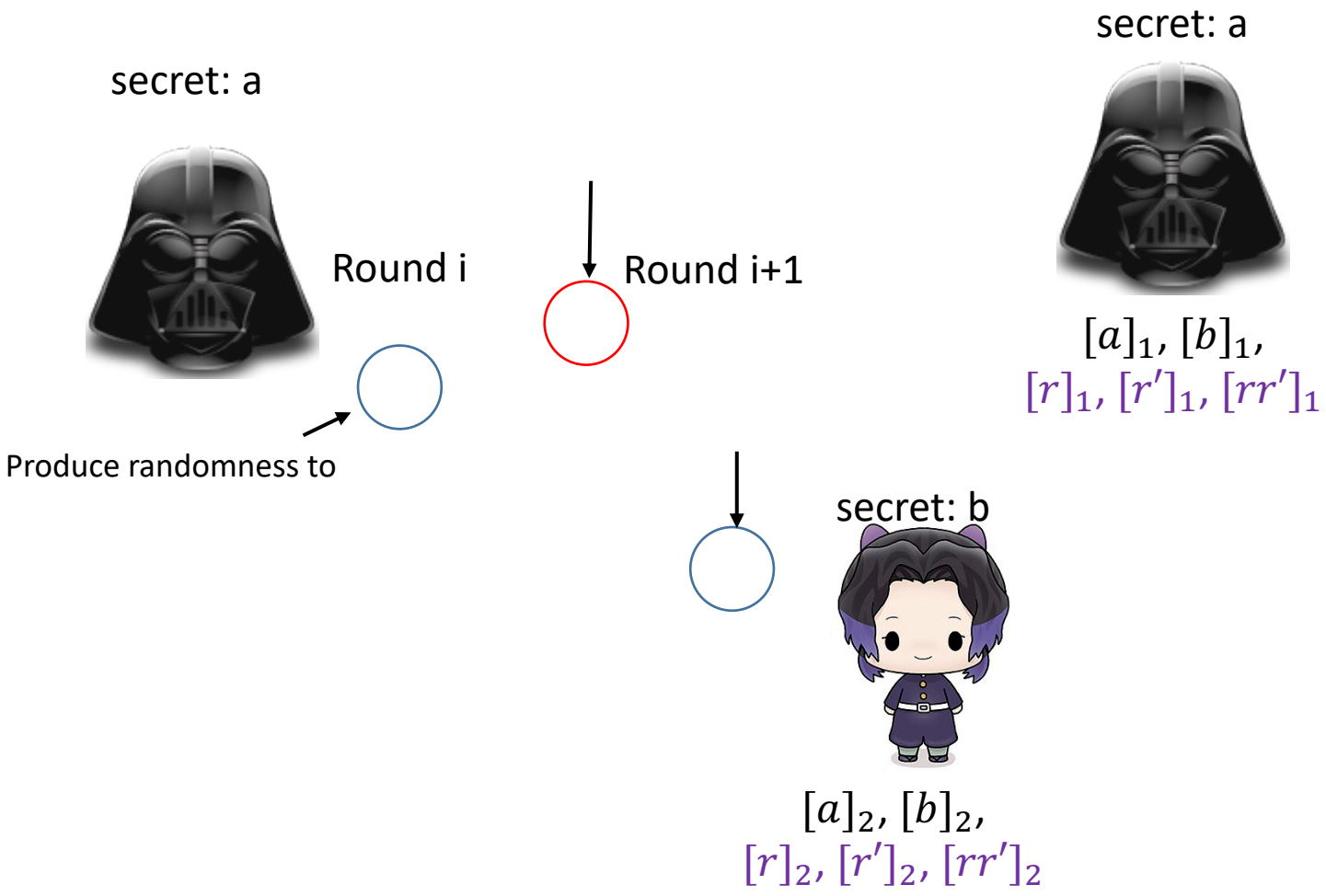


Producing Randomness w/ SSN




secret: a
$[a]_1, [b]_1,$
$[r]_1, [r']_1, [rr']_1$
Round i
Round i+1
4
1
Produce randomness to
2
secret: b
$[a]_2, [b]_2,$
$[r]_2, [r']_2, [rr']_2$

Security of our PR protocol?



Disadvantage of our PR

secret: a

$[a]_1, [b]_1,$
$[r]_1, [r']_1, [rr']_1$

secret: b

$[a]_2, [b]_2,$
$[r]_2, [r']_2, [rr']_2$

need a trust third party in (2,2)-secret sharing

PR Techniques w/o a Third Party

- (2,2)-secret sharing must involve the third party.

- How about (3,3)-secret sharing?
  - Good news! (3,3) can offer PR w/o any third party
  - Three participants can self-organize PR

User

secret: a

$[a]_1, [b]_1,$
$[r]_1, [r']_1, [rr']_1$

Modify our SecretSVM

Leader

secret: b

$[a]_2, [b]_2,$
$[r]_2, [r']_2, [rr']_2$

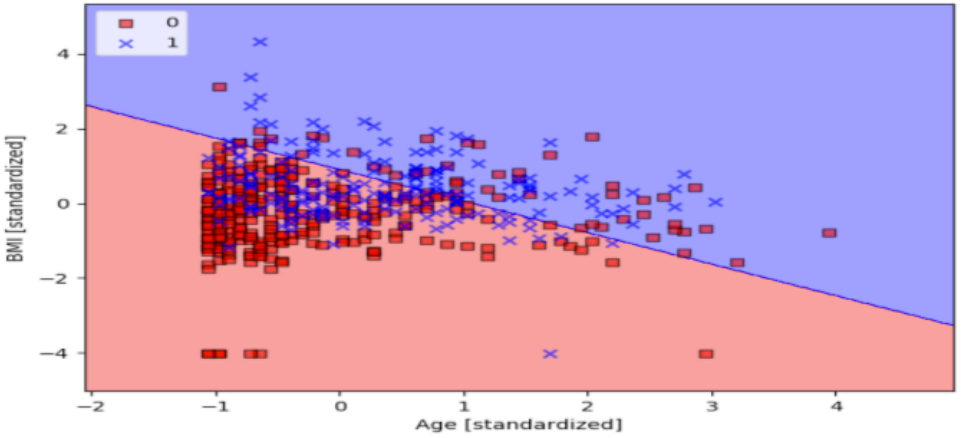
Assistant

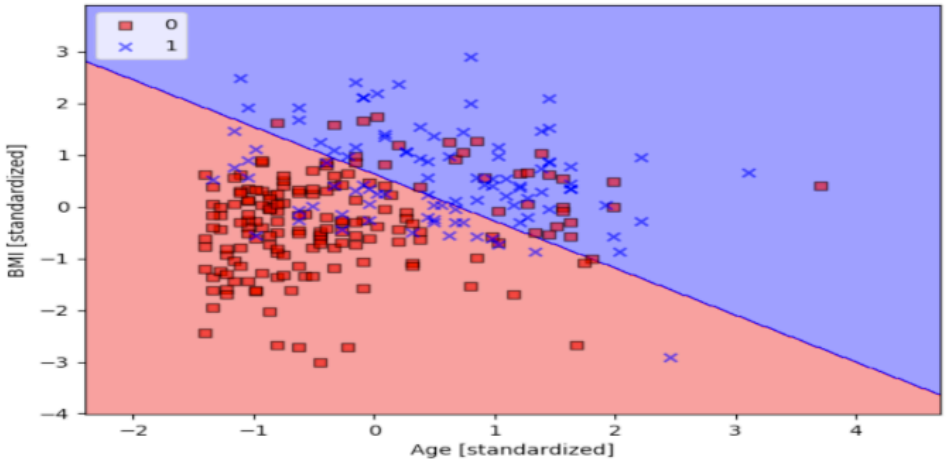
$[a]_3, [b]_3,$
$[r]_3, [r']_3, [rr']_3$

Outline

- Introduction
- Building Blocks
- Our System, SecretSVM
- Experiments and Conclusions

Experiments





Take-home Messages

- SecretSVM
  - Privacy ← secret sharing
  - Collusion prevention ← distributed consensus
  - Condition check in SVM ← secret sharing protocol
  - Reducing computation cost
    - more communication round
    - weak device with nice connection
    - IoT

Enjoy SecretSVM

Credited by Mïïn敏