



iPWS Cup 2024 Rule Book Ver 1.1

Ver 1.1 July 25, 2024

Ver 1.0 July 14, 2024

iPWS Cup 2024 Steering Committee



About iPWS Cup

- PWS (Privacy Workshop) : held annually in Japan since 2015.
 - <https://www.iwsec.org/pws/2024/>
- PWS Cup : A competition for anonymization and its attack techniques, held as one of the PWS events.
- iPWS Cup : International version of the PWS Cup, starting in 2023.

Competition Schedule
July 26 - Sep. 10, 2024

Presentation
Sep. 20, 2024

Kyoto International Conference Center, Kyoto, Japan



iPWSCUP2024
Data Anonymization Competition
in conjunction with IWSEC2024

Competition Schedule
July 26 - Sep. 10, 2024

Presentation
Sep. 20, 2024
Kyoto International Conference Center, Kyoto, Japan

Entry
July 12 - 24, 2024

NOTE : Registration for IWSEC2024 is also required (for a fee).

iPWS Cup 2023
Data Anonymization Competition
at IWSEC 2023 in Yokohama, Japan

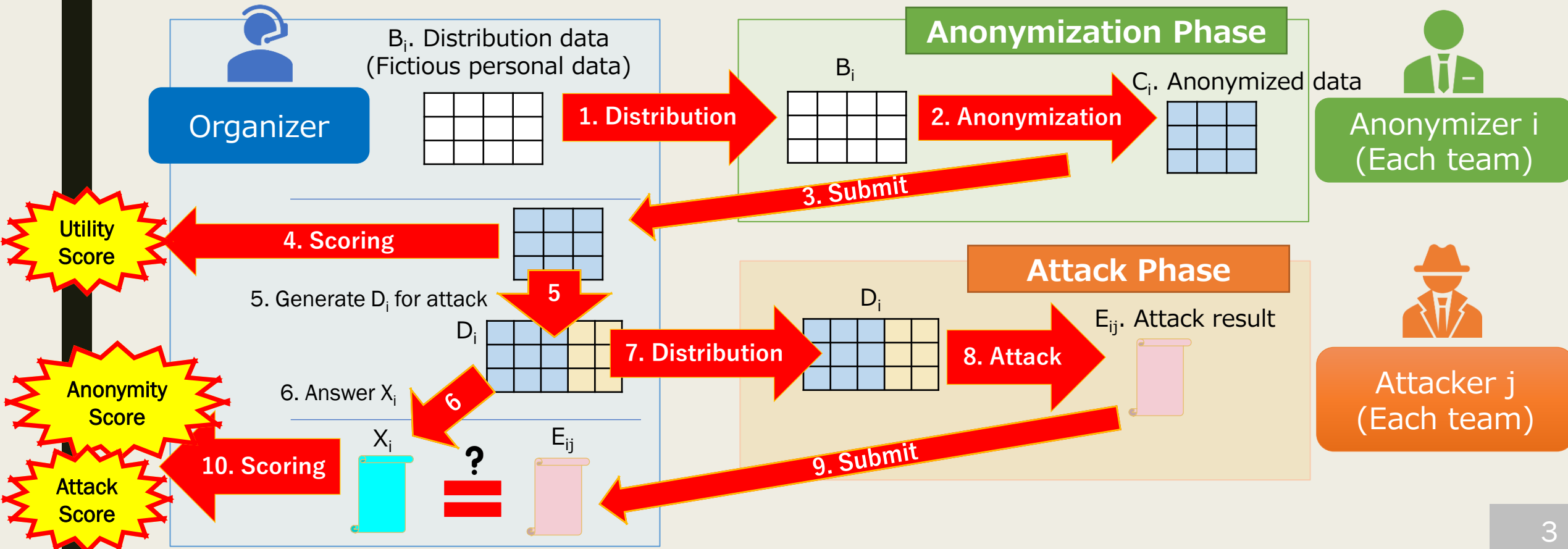
Schedule (tentative)
Registration : May 14 - June 7
Pre. Round : June 12 - July 7 @ Online
Final Round : July 17 - Aug. 21 @ Online
Presentation : Aug. 29 @ IWSEC 2023

iPWS Cup 2023

IWSEC 2023

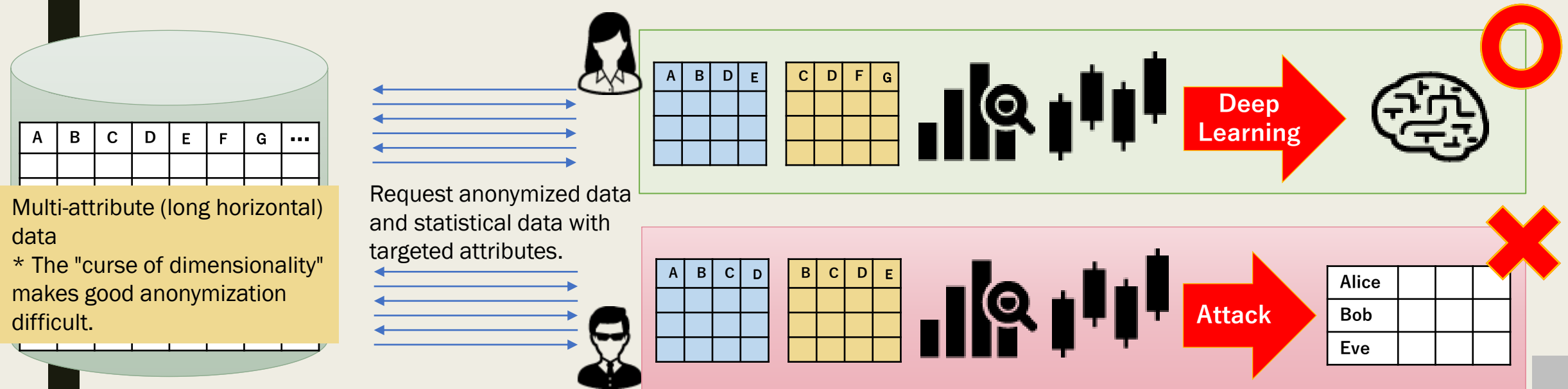
Basic Flow of the iPWS Cup

- All participating teams take part in both the anonymization and attack phases.
- Anonymization phase : Submit anonymized (fictitious) personal data provided by the organizer.
- Attack phase : Submit the attack result on the anonymized data of other teams.
- The organizer announces the results of each team's utility, anonymity and attack scores.



Theme of iPWS Cup 2024

- "Database (DB) reconstruction attacks" have become a serious problem in recent years.
- Using the movie review data "MovieLens" <https://grouplens.org/datasets/movielens/> to create highly useful anonymized data while preventing the DB reconstruction attacks and personal identification attacks.



Data to Use

- MovieLens 1M Dataset (Released 2/2003) <https://grouplens.org/datasets/movielens/>
 - movies.dat
 - Movie ID, Title, Genres
 - 3,952 films, 18 genres (Action, Adventure, Animation, ...) [multiple choice]
 - ratings.dat
 - User ID, Movie ID, Rating (1 – 5), Timestamp
 - 6,040 Users, 1,000,209 records
 - All users have rated at least 20 films (max. 2,314 films).
 - users.dat
 - User ID and Basic attributes (Gender (M/F), Age, Occupation, ZIP-code)
 - Age : 1(Under 18), 18(18-24), 25(25-34), 35(35-44), 45(45-49), 50(50-55), 56(56+)
 - Occupation : 0(other or not specified), 1(academic/educator), ..., 20(writer)
 - ZIP-code : 5 digits

Data to Use (Cont.)

- The MovieLens 1M Dataset is modified as follows.
 - Extract only films with "Fantasy" in genre.
 - 46 films, 4,850 views
 - Consists of the following 51 attributes
 - User ID (Plans to convert to fictitious names.)
 - Basic attributes : Gender (M/F), Age (7 types), Occupation (21 types), ZIP-code (Up to the 3rd digit)
 - Rating (1-5) for each of the 46 films *No viewing is considered to be 0.
 - Create a single csv file with one record per person (51 columns).
 - 1,920 records with a Rating of 5 or more films were extracted.
 - The above processed data is published as "Original data A".
 - Filename of A : "A.csv"
- Generate "distribution data B_i " for each team from the original data A by means of data synthesis.
 - The contents of other teams' distribution data should not be known.
 - Number of records increases from 1,920 to 10,000.
 - Filename of a sample B_i : "sampleBi.csv"
 - Note that B_i differs significantly from A.
 - For fairness, B_i allows each team to choose.

Genres	Films	Views
Action	402	6,012
Adventure	234	5,894
Animation	71	4,808
Children's	179	5,283
Comedy	938	6,031
Crime	171	5,662
Documentary	90	2,243
Drama	1,168	6,037
Fantasy	46	4,850
Film-Noir	33	4,150
Horror	251	5,300
Musical	84	4,754
Mystery	83	5,133
Romance	367	5,961
Sci-Fi	220	5,911
Thriller	375	5,989
War	139	5,769
Western	56	4,100

Image of A

Movie ID (46 film IDs)

51 columns

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	User ID	Gender	Age	Occupation	ZIP-code	2	56	247	260	653	673	810	885	1009	1073	1097	1126	1525	1654	1702	1750	1881	1920	1967	20
2	0	M	1	0	296	0	3	4	5	1	5	4	5	3	2	5	0	5	4	2	5	4	4	2	
3	1	M	18	19	390	3	3	5	5	1	0	4	0	5	0	1	2	0	0	0	5	1	3	3	
4	2	M	25	15	27	3	0	4	5	0	0	1	2	1	3	2	4	1	0	1	1	0	2	4	
5	3	F	1	2	316	4	3	1	5	5	0	0	5	3	1	4	1	3	3	0	5	4	4	0	
6	4	M	45	9	725	4	1	0	1	5	4	0	0	1	4	1	1	0	0	4	2	2	2	0	
7	5	F	18	8	968	4	2	3	1	3	1	2	1	1	3	4	2	3	3	2	5	5	0	0	
8	6	M	50	9	45	2	4	4	3	2	0	2	0	3	5	0	2	2	5	0	1	5	1	5	
9	7	M	50	14	517	2	3	3	4	0	2	0	4	4	1	1	0	4	1	5	2	2	3	4	
10	8	M	18	14	525	2	1	5	3	4	1	0	4	2	1	4	2	4	0	2	1	5	0	4	
11	9	M	18	14	530	0	0	1	2	4	3	0	5	0	1	5	5	0	5	5	1	2	0	2	
12	10	M	25	10	695	2	1	4	2	0	3	4	3	0	2	0	0	5	1	0	3	3	4	1	
13	11	M	56	3	378	4	1	0	1	5	3	0	1	4	2	1	0	1	2	1	1	3	2	2	
14	12	M	18	12	20	1	0	5	2	3	0	2	3	1	5	0	1	4	2	5	5	0	4	2	
15	13	M	25	1	886	0	1	2	0	1	4	0	1	4	4	2	2	5	1	5	4	2	3	2	
16	14	F	18	14	829	1	1	4	5	3	0	4	4	5	5	2	0	2	3	5	1	0	4	2	
17	15	M	18	15	910	4	1	4	4	5	0	3	3	2	0	2	0	0	3	3	0	4	1	3	
18	16	M	50	1	596	2	2	5	2	0	0	4	4	1	5	4	5	0	5	0	2	0	0	5	
19	17	M	35	16	947	1	0	3	2	3	0	0	5	1	2	5	3	0	4	2	3	1	4	2	
20	18	F	25	10	888	4	1	0	5	0	1	1	2	4	1	3	1	1	3	5	4	0	3	1	
21	19	M	18	20	546	0	2	4	2	3	2	0	4	5	2	3	1	5	2	4	1	3	2	3	
22	20	F	1	10	715	5	1	1	3	5	1	0	3	3	1	0	3	2	5	3	2	0	1	4	
23	21	M	1	2	488	2	1	4	5	0	0	4	2	2	2	0	0	2	4	1	2	4	5	0	

1,920 rows (excluding header lines)

List of Films in the Genre "Fantasy"

ID	Title	Genres	Views
2	Jumanji (1995)	Adventure Children's Fantasy	701
56	Kids of the Round Table (1995)	Adventure Children's Fantasy	9
247	Heavenly Creatures (1994)	Drama Fantasy Romance Thriller	477
260	Star Wars, Episode IV - A New Hope (1977)	Action Adventure Fantasy Sci-Fi	2991
653	Dragonheart (1996)	Action Adventure Fantasy	612
673	Space Jam (1996)	Adventure Animation Children's Comedy Fantasy	563
810	Kazaam (1996)	Children's Comedy Fantasy	120
885	Bogus (1996)	Children's Drama Fantasy	43
1009	Escape to Witch Mountain (1975)	Adventure Children's Fantasy	291
1073	Willy Wonka and the Chocolate Factory (1971)	Adventure Children's Comedy Fantasy	1313
1097	E.T. the Extra-Terrestrial (1982)	Children's Drama Fantasy Sci-Fi	2269
1126	Drop Dead Fred (1991)	Comedy Fantasy	317
1525	Warriors of Virtue (1997)	Action Adventure Children's Fantasy	44
1654	FairyTale, A True Story (1997)	Children's Drama Fantasy	87
1702	Flubber (1997)	Children's Comedy Fantasy	302
1750	Star Kid (1997)	Adventure Children's Fantasy Sci-Fi	63
1881	Quest for Camelot (1998)	Adventure Animation Children's Fantasy	68
1920	Small Soldiers (1998)	Animation Children's Fantasy War	364
1967	Labyrinth (1986)	Adventure Children's Fantasy	554
2017	Babes in Toyland (1961)	Children's Fantasy Musical	162
2021	Dune (1984)	Fantasy Sci-Fi	789
2043	Darby O'Gill and the Little People (1959)	Adventure Children's Fantasy	158

ID	Title	Genres	Views
2086	One Magic Christmas (1985)	Drama Fantasy	29
2087	Peter Pan (1953)	Animation Children's Fantasy Musical	594
2093	Return to Oz (1985)	Adventure Children's Fantasy Sci-Fi	276
2100	Splash (1984)	Comedy Fantasy Romance	1163
2105	Tron (1982)	Action Adventure Fantasy Sci-Fi	970
2138	Watership Down (1978)	Animation Children's Drama Fantasy	305
2143	Legend (1985)	Adventure Fantasy Romance	355
2174	Beetlejuice (1988)	Comedy Fantasy	1495
2193	Willow (1988)	Action Adventure Fantasy	802
2253	Toys (1992)	Action Comedy Fantasy	440
2399	Santa Claus, The Movie (1985)	Adventure Children's Fantasy	223
2628	Star Wars, Episode I - The Phantom Menace (1999)	Action Adventure Fantasy Sci-Fi	2250
2797	Big (1988)	Comedy Fantasy	1491
2872	Excalibur (1981)	Action Drama Fantasy Romance	742
2968	Time Bandits (1981)	Adventure Fantasy Sci-Fi	1010
3393	Date with an Angel (1987)	Comedy Fantasy	51
3438	Teenage Mutant Ninja Turtles (1990)	Action Children's Fantasy	534
3439	Teenage Mutant Ninja Turtles II, The Secret of the Ooze (1991)	Action Children's Fantasy	251
3440	Teenage Mutant Ninja Turtles III (1993)	Action Children's Fantasy	188
3466	Heart and Souls (1993)	Comedy Fantasy	219
3479	Ladyhawke (1985)	Adventure Fantasy Romance	542
3489	Hook (1991)	Adventure Fantasy	722
3877	Supergirl (1984)	Action Adventure Fantasy	182
3889	Highlander, Endgame (2000)	Action Adventure Fantasy	135

How to Create and Distribute B_i (Synthetic Data)

- 100 synthetic data (data ID 00-99) are created from the original data A using PrivBayes.
 - PrivBayes <https://github.com/DataResponsibly/DataSynthesizer/blob/master/DataSynthesizer/lib/PrivBayes.py>
- Disclose the hash value (SHA256) of each synthetic data with the data ID.
- Each team chooses three data IDs of their choice.
 - In the case of duplicates, the order of submission will be the earliest first and the remaining teams will change to the free data ID immediately afterwards (99 followed by 00).
- If three data IDs are determined for all teams without duplication, the organizer distributes the three synthetic data corresponding to the data IDs to each team as distribution data 1, 2 and 3.
- Each team checks the integrity of the distribution data 1, 2, 3 from the hash values if necessary.
- Each team freely selects one data from the distributed data 1, 2 and 3 that is easy to anonymize, designates it as B_i and submits the data ID of B_i together with the anonymized data in the anonymization phase (The data ID will be disclosed after the anonymization phase has started.).

[Anonymization Phase] Processing of Anonymizer i

1. Extract subset data $B_i^{(0)}-B_i^{(9)}$ of the following 10 patterns from the distribution data B_i .
 - Basic Attributes (BAs) : Gender, Age, Occupation, ZIP-code
 - $B_i^{(0)}-B_i^{(9)}$ will be distributed as csv files by the organizer.
 - 1. BAs and "Action" (260, 653, 1525, 2105, 2193, 2253, 2628, 2872, 3438, 3439, 3440, 3877, 3889)
 - 2. BAs and "Adventure" (2, 56, 260, 653, 673, 1009, 1073, 1525, 1750, 1881, 1967, 2043, 2093, 2105, 2143, 2193, 2399, 2628, 2968, 3479, 3489, 3877, 3889)
 - 3. BAs and "Animation" (673, 1881, 1920, 2087, 2138)
 - 4. BAs and "Children's" (2, 56, 673, 810, 885, 1009, 1073, 1097, 1525, 1654, 1702, 1750, 1881, 1920, 1967, 2017, 2043, 2087, 2093, 2138, 2399, 3438, 3439, 3440)
 - 5. BAs and "Comedy" (673, 810, 1073, 1126, 1702, 2100, 2174, 2253, 2797, 3393, 3466)
 - 6. BAs and "Drama" (247, 885, 1097, 1654, 2086, 2138, 2872)
 - 7. BAs and "Romance" (247, 2100, 2143, 2872, 3479)
 - 8. BAs and "Sci-Fi" (260, 1097, 1750, 2021, 2093, 2105, 2628, 2968)
 - 9. BAs and "Musical" and "Thriller" and "War" (247, 1920, 2017, 2087)
 - 10. BAs and View Top 10 films (260, 1097, 2628, 2174, 2797, 1073, 2100, 2968, 2105, 2193)
2. For each of the above $B_i^{(0)}-B_i^{(9)}$, the processing is freely carried out and submitted as anonymized data $C_i^{(0)}-C_i^{(9)}$.
 - Each value may be freely changed within the value range, but not to values outside the value range (e.g. changing age to 10-year increments).

[Anonymization Phase] Utility and Sample Anonymity

- Anonymizer i creates anonymized data $C_i^{(0)}-C_i^{(9)}$ for $B_i^{(0)}-B_i^{(9)}$ such that the following "utility" and "sample anonymity" scores are as high as possible.
 - As the anonymity score cannot be calculated until the end of the attack phase, the sample anonymity will be introduced instead.
- **Utility Score** : $(1-MAE_{\text{worst}}) \times 100$ (0 to 100 points)
 - MAE_{worst} is the worst value of MAE (Mean Absolute Error) values of all cross-tabulations of BAs x Rating and Rating x Rating obtained from $(B_i^{(j)}, C_i^{(j)})$ for $j=0,1,\dots,9$.
- **Sample Anonymity Score** : $(1-\text{SampAtk}_{\text{succ}}) \times 100$ (0 to 100 points)
 - $\text{SampAtk}_{\text{succ}}$ is the attack success rate by the sample attack code described on page 18.
 - Sample anonymity scores will be disclosed during the anonymization phase.


Crosstabulation table for Gender & Film 260 in $B_i^{(0)}$, where $a_{*,*}$ is the frequency.

	0	1	2	3	4	5
F	$a_{F,0}$	$a_{F,1}$	$a_{F,2}$	$a_{F,3}$	$a_{F,4}$	$a_{F,5}$
M	$a_{M,0}$	$a_{M,1}$	$a_{M,2}$	$a_{M,3}$	$a_{M,4}$	$a_{M,5}$

Crosstabulation table for Gender & Film 260 in $C_i^{(0)}$, where $b_{*,*}$ is the frequency.

	0	1	2	3	4	5
F	$b_{F,0}$	$b_{F,1}$	$b_{F,2}$	$b_{F,3}$	$b_{F,4}$	$b_{F,5}$
M	$b_{M,0}$	$b_{M,1}$	$b_{M,2}$	$b_{M,3}$	$b_{M,4}$	$b_{M,5}$

These symbols represent the MAE of cross-tabulation of Gender x Movie 260 obtained from $B_i^{(0)}$ and $C_i^{(0)}$.


$$MAE_{(0),\text{Gender},260} = \frac{1}{20000} \sum_{i=0}^5 (|a_{F,i} - b_{F,i}| + |a_{M,i} - b_{M,i}|)$$

[Attack Phase] Processing of the Organizer

- Randomly select 50 records from each team's distribution data B_i .
- Separate User ID & BAs (Gender, Age, Occupation, ZIP-code) and the ratings for 46 films and randomly shuffle all records in the ratings data.
- For each record of the ratings data, one place is selected at random and painted black.
- A set of anonymized data $C_i^{(0)}-C_i^{(9)}$ for each team and the User ID & BAs and processed (random shuffling + black filled) ratings data are used as the attack data D_i and sent to attacker j .

Randomly selected 50 records from B_i .

User ID & BAs					Ratings data								
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	User ID	Gender	Age	Occupatic	ZIP-code	2	56	247	260	653	673	810	885
2	0	M	1	0	296	0	3	4	5	1	5	4	5
3	1	M	18	19	390	3	3	5	5	1	0	4	0
4	2	M	25	15	27	3	0	4	5	0	0	1	2
5	3	F	1	2	316	4	3	1	5	5	0	0	5
6	4	M	45	9	725	4	1	0	1	5	4	0	0
7	5	F	18	8	968	4	2	3	1	3	1	2	1
8	6	M	50	9	45	2	4	4	3	2	0	2	0
9	7	M	50	14	517	2	3	3	4	0	2	0	4
10	8	M	18	14	525	2	1	5	3	4	1	0	4
11	9	M	18	14	530	0	0	1	2	4	3	0	5
12	10	M	25	10	695	2	1	4	2	0	3	4	3



Rating data processed by random shuffling and black paining.

User ID & BAs					Processed ratings data								
	A	B	C	D	E		F	G	H	I	J	K	L
1	User ID	Gender	Age	Occupatic	ZIP-code	2	56	247	260	653	673	810	885
2	0	M	1	0	296	2	4	■	3	2	0	2	0
3	1	M	18	19	390	4	3	1	5	5	■	0	5
4	2	M	25	15	27	3	3	5	■	1	0	4	0
5	3	F	1	2	316	2	1	5	3	4	1	0	■
6	4	M	45	9	725	0	3	4	5	■	5	4	5
7	5	F	18	8	968	0	0	■	2	4	3	0	5
8	6	M	50	9	45	4	■	3	1	3	1	2	1
9	7	M	50	14	517	■	3	3	4	0	2	0	4
10	8	M	18	14	525	3	0	4	5	0	0	■	2
11	9	M	18	14	530	4	1	0	1	■	4	0	0
12	10	M	25	10	695								

[Attack Phase] Processing of Attacker j

- Identification Attack : Guess the number of randomly shuffled records (0-50 points).
- DB Reconstruction Attack : Restore the values of 50 blacked-out areas (0-50 points).
- As attack result data E_{ij} for Anonymizer i, generate the data with 50 rows and 2 columns.
 - (Identification attack result) In row k of column 1, estimate and enter the row number of User ID & BAs corresponding to the k-th row of the processed ratings data in D_i .
 - (DB reconstruction attack result) In row k of column 2, estimate and enter the value of the blacked-out part of the k-th row of the ratings data in D_i .
- Submit a single csv file consisting of attack result data $E_{1j}, E_{2j}, E_{3j}, \dots, E_{Nj}$ where N is the number of teams.
 - (2i-1) and 2i-th columns represent an identification attack result and DB reconstruction attack result for the i-th team.
 - The columns for your team's attack result should be blank.
- Attack result data can be submitted up to three times. *Finally, any of the three attack results can be selected for final submission.

Randomly selected 50 records from B_i .

	User ID & BAs					Ratings data							
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	User ID	Gender	Age	Occupatic	ZIP-code	2	56	247	260	653	673	810	885
2	0	M	1	0	296	0	3	4	5	1	5	4	5
3	1	M	18	19	390	3	3	5	5	1	0	4	0
4	2	M	25	15	27	3	0	4	5	0	0	1	2
5	3	F	1	2	316	4	3	1	5	5	0	0	5
6	4	M	45	9	725	4	1	0	1	5	4	0	0
7	5	F	18	8	968	4	2	3	1	3	1	2	1
8	6	M	50	9	45	2	4	4	3	2	0	2	0
9	7	M	50	14	517	2	3	3	4	0	2	0	4
10	8	M	18	14	525	2	1	5	3	4	1	0	4
11	9	M	18	14	530	0	0	1	2	4	3	0	5
12	10	M	25	10	695	2	1	4	2	0	3	4	3



Rating data processed by random shuffling and black paining.

	User ID & BAs					Processed ratings data							
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	User ID	Gender	Age	Occupatic	ZIP-code	2	56	247	260	653	673	810	885
2	0	M	1	0	296	2	4	■	3	2	0	2	0
3	1	M	18	19	390	4	3	1	5	5	■	0	5
4	2	M	25	15	27	3	3	5	■	1	0	4	0
5	3	F	1	2	316	2	1	5	3	4	1	0	■
6	4	M	45	9	725	0	3	4	5	■	5	4	5
7	5	F	18	8	968	0	0	■	2	4	3	0	5
8	6	M	50	9	45	4	■	3	1	3	1	2	1
9	7	M	50	14	517	■	3	3	4	0	2	0	4
10	8	M	18	14	525	3	0	4	5	0	0	■	2
11	9	M	18	14	530	4	1	0	1	■	4	0	0
12	10	M	25	10	695								

[Attack Phase] Score

- **Anonymity Score** : $100 - \{\text{Highest score of another team's attack (identification + DB reconstruction)}\}$ (0-100 points)
- **Overall Score** : Anonymity score + Utility score (0-200 points)
- **Attack Score** : Value of added points for attacks against the top five teams in the overall score (0-500 points)
 - Your team's attack score is the highest score by another team attacking your team.

(The following are reminders.)

- Identification Attack : Guess the number of randomly shuffled records (0-50 points).
- DB Reconstruction Attack : Restore the values of 50 blacked-out areas (0-50 points).
- The scores determined in the anonymization phase :
 - **Utility Score** : $(1 - \text{MAE}_{\text{worst}}) \times 100$ (0 to 100 points)
 - $\text{MAE}_{\text{worst}}$ is the worst value of MAE (Mean Absolute Error) values of all cross-tabulations of BAs x Rating and Rating x Rating obtained from $(B_i^{(j)}, C_i^{(j)})$ for $j=0,1,\dots,9$.
 - **Sample Anonymity Score** : $(1 - \text{SampAtk}_{\text{succ}}) \times 100$ (0 to 100 points)
 - $\text{SampAtk}_{\text{succ}}$ is the attack success rate by the sample attack code described on page 18.
 - Sample anonymity scores will be disclosed during the anonymization phase.

Awards

- 1st to 3rd place awards
 - Anonymity score + Utility score
 - The number of places awarded may vary slightly depending on the number of participating teams.
- Best attack award
 - Awarded to the team with the highest attack score.
- Best presentation award
 - Awarded to the team with the best presentation at the iPWS Cup 2024 session on 20th September.
- Best data scientist award
 - Awarded to the team that actually proposes the most useful analysis method using your anonymized data.
 - Overall assessment of the originality and practicality of the analysis method and the utility of the analysis using anonymized data.
 - It is OK if you prepare before the presentation on the day of the event; it is optional whether you present or not.
- Gifts
 - Certificate : all the above-mentioned award-winning teams.
 - Secondary prize (something related to Kyoto) : 1st to 3rd place, best attack, best presentation, and best scientist teams.
 - Prize money (Visa Gift Card totaling 100,000 yen) *1st to 3rd place and best attack teams.

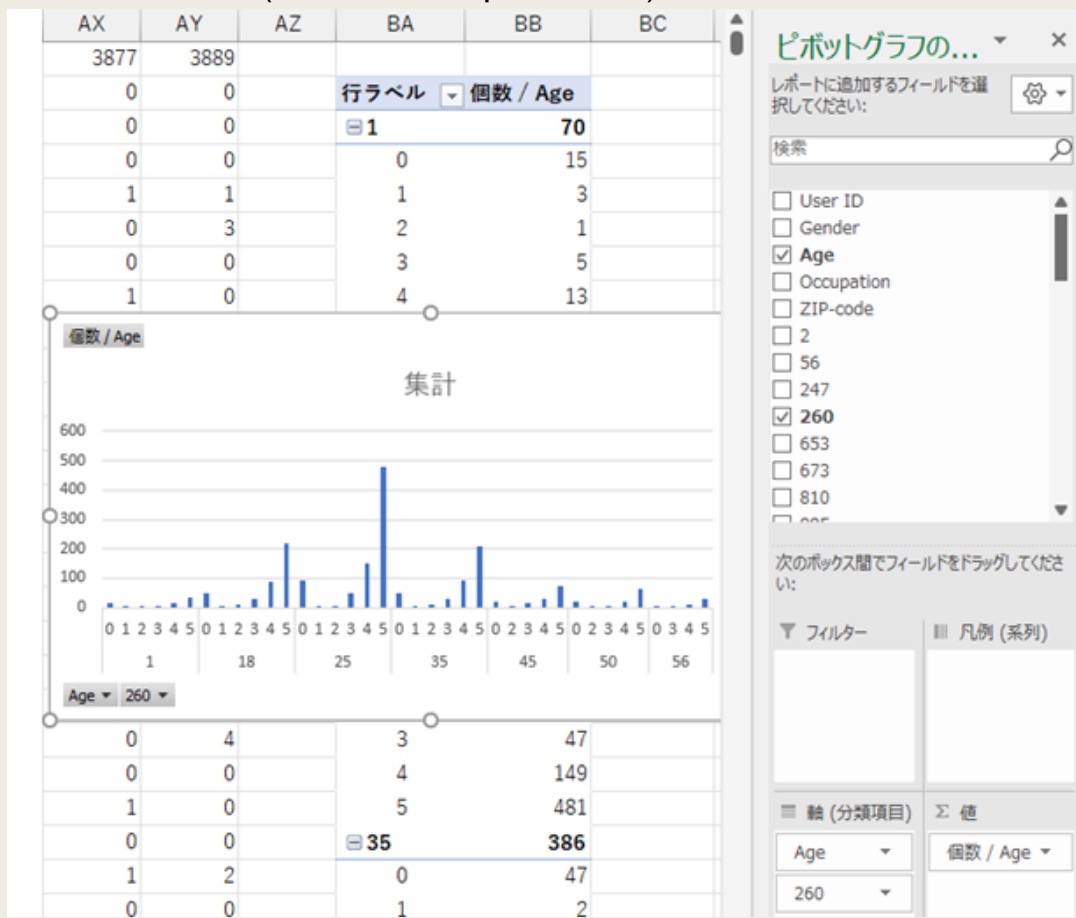
Sample Data Image

	A	B	C	D	E	F	G	H	I	J
1	Name	Gender	Age	Occupation	ZIP-code	2	56	247	260	653
2	Waylan Kirton	M	25	8	794	0	3	4	3	5
3	Jamal Seamon	M	50	8	322	0	4	0	3	2
4	Garvy Abyss	M	35	8	975	3	1	3	2	3
5	Arline Morales	F	1	9	140	4	0	5	4	5
6	Babbie Lorroway	F	1	3	942	1	2	1	2	0
7	Corissa Parham	F	56	11	180	0	3	1	0	4

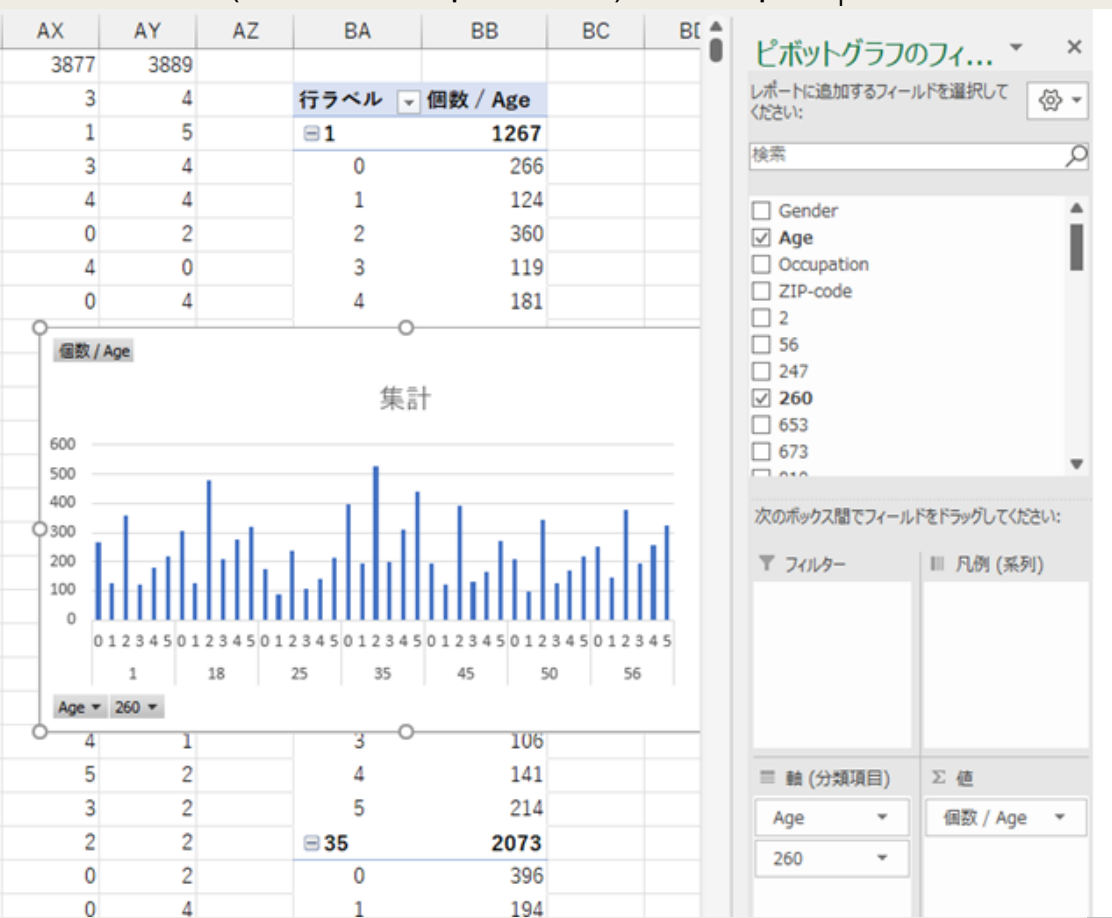
■ sampleBi.csv

- Synthetic data with 10,001 rows (including the header row) and 51 columns generated from A.

Histogram of the crosstabulation table of "Age" and "Film 260" (Star Wars Episode IV) in A.



Histogram of the crosstabulation table of "Age" and "Film 260" (Star Wars Episode IV) in sampleB.



Sample Codes (Python)

■ Anonymization

- Random shuffling (sampleRandomShuffle.py)
 - Read a csv file and randomly replace the order of records (rows).
 - However, the first line (header line) is not replaced.
- Swapping (swapping.py)
 - Read a csv file, specify the attribute name(s) and the number of times n , select two random pieces of data for the specified attribute and swap the values (repeat this n times).
- Recoding (recoding.py)
 - Read a csv file, specify the attribute name (only one), the value before the change and the value after the change, and batch convert the value before the change to the value after the change.
 - Note that this can be seen as a similar process to generalization, as the values before the change are merged with the values after the change. However, the MAE values in the cross-tabulation table become larger.
- Randomization (samleRandomization.py)
 - Read a csv file and randomize the data for a given attribute.
 - For example, when the data for attribute A is selected, one element from the domain of attribute A is randomly selected and replaced.

Sample Codes (Python)

■ Attack (sampleAttack.py)

- Perform identification and DB reconstruction attacks
- Identification attack
 - Read 50 records of User ID & BAs, 50 ratings records and $C_i^{(0)}-C_i^{(9)}$, connect 50 ratings records one by one for each record of User ID & BAs, calculate the Hamming distance to each record of $C_i^{(0)}-C_i^{(9)}$. The ratings record number with the smallest sum of the minimum Hamming distances of $C_i^{(0)}-C_i^{(9)}$ is estimated as the correct answer and output.
 - Note that each row of 50 records has only one column blacked out (marked with an asterisk '*').
- DB reconstruction attack
 - Output the data of the $C_i^{(0)}-C_i^{(9)}$ records for which the sum of the minimum Hamming distances for $C_i^{(0)}-C_i^{(9)}$ is the smallest, as calculated in the identification attack.
 - Note that each row of 50 records has only one column blacked out (marked with an asterisk '*').
- Output a csv file with the results of the identification attack marked in column 1 and the results of the DB reconstruction attack marked in column 2.

Administrative codes (also available to participating teams)

■ Score calculation

- Utility score (utilityScore.py)
 - Input B_i and $C_i^{(0)}-C_i^{(9)}$, calculate MAE of all cross-tabulations of BAs x Rating and Rating x Rating and use the worst value MAE_{worst} (0-1) to calculate and output utility score $S_{util,i} = (1 - MAE_{worst}) \times 100$.
- Attack score (attackScore.py)
 - Input attack result data E_{ij} and correct answer data X_i and output the number of matches.
 - E_{ij} and X_i have 50 rows and 2 columns.
- Sample anonymity score (sampleAnonymityScore.py)
 - Input attack data D_i and correct data X_i , run sampleAttack.py to obtain the attack result data E_{ij} , and use the output $SampAtk_{succ}$ (0-100) of attackScore.py to calculate the anonymity score $S_{anon,i} = 100 - SampAtk_{succ}$.

■ Checker

- C_i (checkCi.py)
 - Check for correct format of anonymized data C_i .
 - Checkers for distribution data B_i , attack data D_i , attack result data E_{ij} and correct answer data X_i will be created.
- Hash value generator (genHash.py)

■ Table operations

- Create $B_i^{(0)}, \dots, B_i^{(9)}$ from B_i . (split.py)

How to Participate

- See "How to participate in the iPWS Cup 2024" on the iPWS Cup 2024 HP.
 - <https://www.iwsec.org/pws/2024/cup24.html>
- Use the competition platform "CodaBench".



[Benchmarks](#) [Resources](#) [Queue Management](#)



IPWS CUP 2024

ORGANIZED BY: Kchida
CURRENT PHASE ENDS: Never
CURRENT SERVER TIME: 2024年7月14日 10:44 JST
Docker image: [codalab/codalab-legacy:py39](#)
Competition Report: <https://www.iwsec.org/pws/ipws2024/>

2 PARTICIPANTS

1 SUBMISSIONS

Aug 2024 Sep 2024

Get Started Phases

About iPWS Cup 2024

Attack Phase Result															
#	User	Entries	Date of Last Entry	Team Name	Attack Score	Privacy Scores after the attack									
					Attack Score ▲	Team 1 ▲	Team 2 ▲	Team 3 ▲	Team 4 ▲	Team 5 ▲	Team 6 ▲	Team 7 ▲	Team 8 ▲	Team 9 ▲	Team 10 ▲
1	Hikaru	13	08/21/23	08: THREE	0.3044 (1)	0.9800 (1)	0.0000 (1)	0.4600 (1)	0.9500 (1)	0.9500 (1)	0.8667 (2)	0.8833 (4)	- (9)	0.4100 (1)	0.7600 (2)
2	kchida	11	08/17/23	04: Gunmataro116luxury	0.2911 (2)	0.9800 (1)	0.0267 (3)	0.4667 (2)	- (9)	0.9533 (2)	0.8433 (1)	0.8833 (4)	0.8900 (1)	0.5533 (6)	0.7833 (4)

A screen shot of iPWS Cup 2023 (Attack phase result)

See you in Kyoto, Japan!

Organized by  



IWSEC Top

IWSEC 2024 Home

Call For Papers

Call For Posters

Important Dates

Submission

Keynote

Program

Proceedings

Venue

Excursion

Guidelines

Registration

iPWS Cup 2024

Committees

Contact Us

Sponsors

IWSEC 2024

The 19th International Workshop on Security
September 17 (Tue) – September 19 (Thu), 2024
Kyoto International Conference Center, Kyoto, Japan

Co-located with iPWS Cup 2024



<https://www.iwsec.org/2024/>