Trustworthy ML Project - Practical Methods of Inventory Anonymization and Privacy

Charles Cook

Rensselaer Polytechnic Institute

April 6th, 2022

Motivation: A Trader's Life

- Virtual Marketplaces are burgeoning environments of commerce
  - ex. Steam (TF2/CSGO Item Trading)
- Curators of rare/valuable items in these marketplaces are often bothered and accosted regularly by Scammers
  - Scammers seek to con Curators out of their items, or hijack their accounts totally
  - Aggravated if the Curator's inventory is publicly visible
    - ex. Only options are Steam are Public, Friends Only, or Private
- Curators who engage in trading often need to publicly display their Inventory, but this can bring negative attention
- **How can Curators present their Inventory in such a way to minimize Scammer attention but maintain status with trading Appraisers?**

Problem Statement: Games Dataset

- The combined techniques of k-Anonymization and β-Sampling *(Li et. al., 2012)* can achieve Differential Privacy *(Dwork, 2006)*
  - Recap: D.P. **protects** individual **identities** in a dataset while **maintaining** some level of **utility**
- Experimental Dataset: NES & SNES video games
  - Sensitive Attributes: **Title**, **Price**
  - Quasi-Identifies: **Genre**, **Developer(s)**, **Publisher(s)**, **Release Year**, **Platform**
    - *(had to manually add in (estimated) Prices and Genres)*
  - Superset of 223 total titles *(Items of interest)*
    - *(total of over 1,000 on Wikipedia's lists)
  - Collection set of 68 titles in Curator's *Inventory*
  - Collection/Interest Ratio of ~30.5%

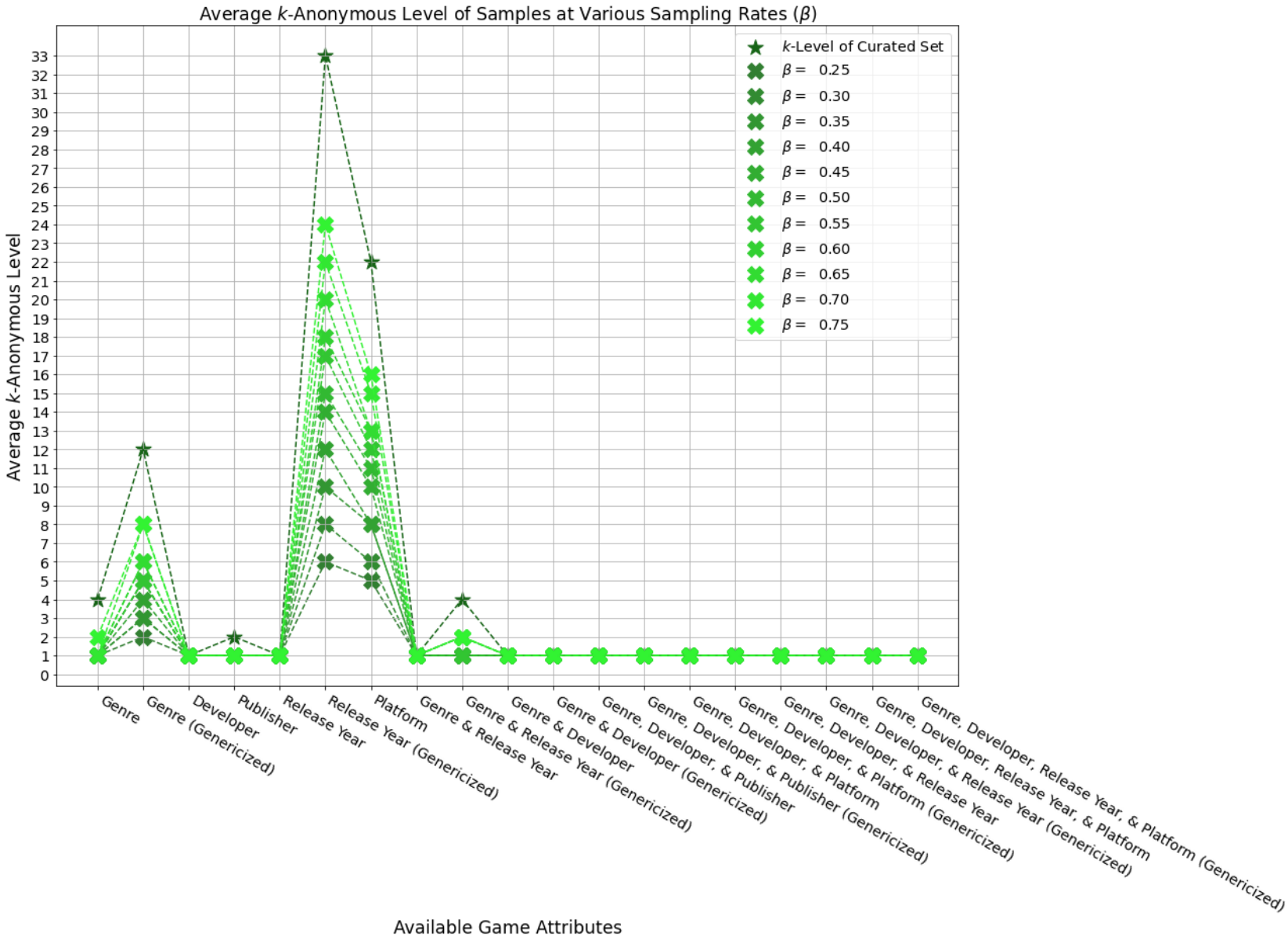
Game Dataset Pretty-Printed

['86, NES]	[Platformer]	(15)	Balloon Fight	(Nintendo Research & Development 1)
['86, NES]	[Platformer]	(15)	Donkey Kong	(Nintendo Research & Development 1)
['86, NES]	[Platformer]	(10)	Mario Bros.	(Nintendo Research & Development 1)
['86, NES]	[Fighting]	(8)	Urban Champion	(Nintendo Research & Development 1)
['87, NES]	[Platventure]	(25)	Mega Man	(Capcom)
['87, NES]	[Platventure]	(35)	Castlevania	(Konami)
['87, NES]	[Platventure]	(25)	Kid Icarus	(Nintendo Research & Development 1)
['87, NES]	[Platventure]	(30)	Metroid	(Nintendo Research & Development 1)
['87, NES]	[Fighting]	(20)	Mike Tyson's Punch-Out!!	(Nintendo Research & Development 3)
['87, NES]	[Adventure]	(35)	The Legend of Zelda	(Nintendo Research & Development 4)
['87, NES]	[Racing]	(10)	Rad Racer	(Square)
['88, NES]	[Shooter]	(5)	Bionic Commando	(Capcom)
['88, NES]	[Shooter]	(25)	Contra	(Konami)
['88, NES]	[Adventure]	(15)	Metal Gear	(Konami)
['88, NES]	[Sports]	(15)	Skate or Die!	(Konami)
['88, NES]	[Shooter]	(10)	Galaga: Demons of Death	(Namco)
['88, NES]	[Shooter]	(8)	Xevious	(Namco)
['88, NES]	[Platformer]	(2)	Super Mario Bros./Duck Hunt	(Nintendo)
['88, NES]	[Platformer]	(15)	Super Mario Bros. 2	(Nintendo Research & Development 4)
['88, NES]	[Adventure]	(20)	Zelda II: The Adventure of Link	(Nintendo Research & Development 4)
['89, NES]	[Platventure]	(40)	Mega Man 2	(Capcom)
['89, NES]	[RPG]	(2)	Dragon Warrior	(Chunsoft)
['89, NES]	[Shooter]	(8)	Gyruss	(Konami)
['89, NES]	[Puzzle]	(15)	Tetris	(Nintendo Research & Development 1)
['90, NES]	[Platventure]	(45)	Mega Man 3	(Capcom)
['90, NES]	[Platventure]	(50)	Castlevania III: Dracula's Curse	(Konami)
['90, NES]	[Shooter]	(25)	Super C	(Konami)
['90, NES]	[Puzzle]	(15)	Dr. Mario	(Nintendo Research & Development 1)
['90, NES]	[Adventure]	(25)	StarTropics	(Nintendo Research & Development 3)
['90, NES]	[Platformer]	(20)	Super Mario Bros. 3	(Nintendo Research & Development 4)
['90, NES]	[RPG]	(35)	Final Fantasy	(Square)
['91, SNES]	[Fighting]	(35)	Final Fight	(Capcom)
['91, SNES]	[Platventure]	(45)	Super Castlevania IV	(Konami)
['91, SNES]	[Racing]	(25)	F-Zero	(Nintendo)
['91, SNES]	[Sports]	(20)	Pilotwings	(Nintendo)
['91, SNES]	[Platformer]	(20)	Super Mario World	(Nintendo)
['91, SNES]	[RPG]	(30)	Final Fantasy II	(Square)
['92, NES]	[Platventure]	(45)	Mega Man 4	(Capcom)
['92, NES]	[Platventure]	(50)	Mega Man 5	(Capcom)
['92, SNES]	[Fighting]	(30)	Street Fighter II: The World Warrior	(Capcom)
['92, SNES]	[Adventure]	(35)	The Legend of Zelda: A Link to the Past	(Nintendo)
['92, SNES]	[Racing]	(80)	Super Mario Kart	(Nintendo)
['92, SNES]	[RPG]	(40)	Final Fantasy Mystic Quest	(Square)

Problem Statement: Inventory Anonymization

- Anonymization Through **Suppression** & **Generalization**
  - Title & Price always withheld, optionally one or more QIs as well
  - Also optionally, Genre despecified & Release Year turned into Release Decade
    - 10 specific Genres to 3 Super-Genres
  - k-Anonymous level is computed after Suppression/Generalization, not set as a target
  - **Hypothesis/Heuristic**: Higher k-Levels mean both Scammer & Appraiser do worse
    - We want to balance this such that Scammer does bad but Appraiser does good
- Sampling of Anonymized Inventory
  - 25% to 75%, 5% increments
- Publish Sampled-Anonymous Inventory, Size (Cardinality) of Inventory, and full Interest Superset
  - Possibly leave inference of Superset up for Scammer/Appraiser to do

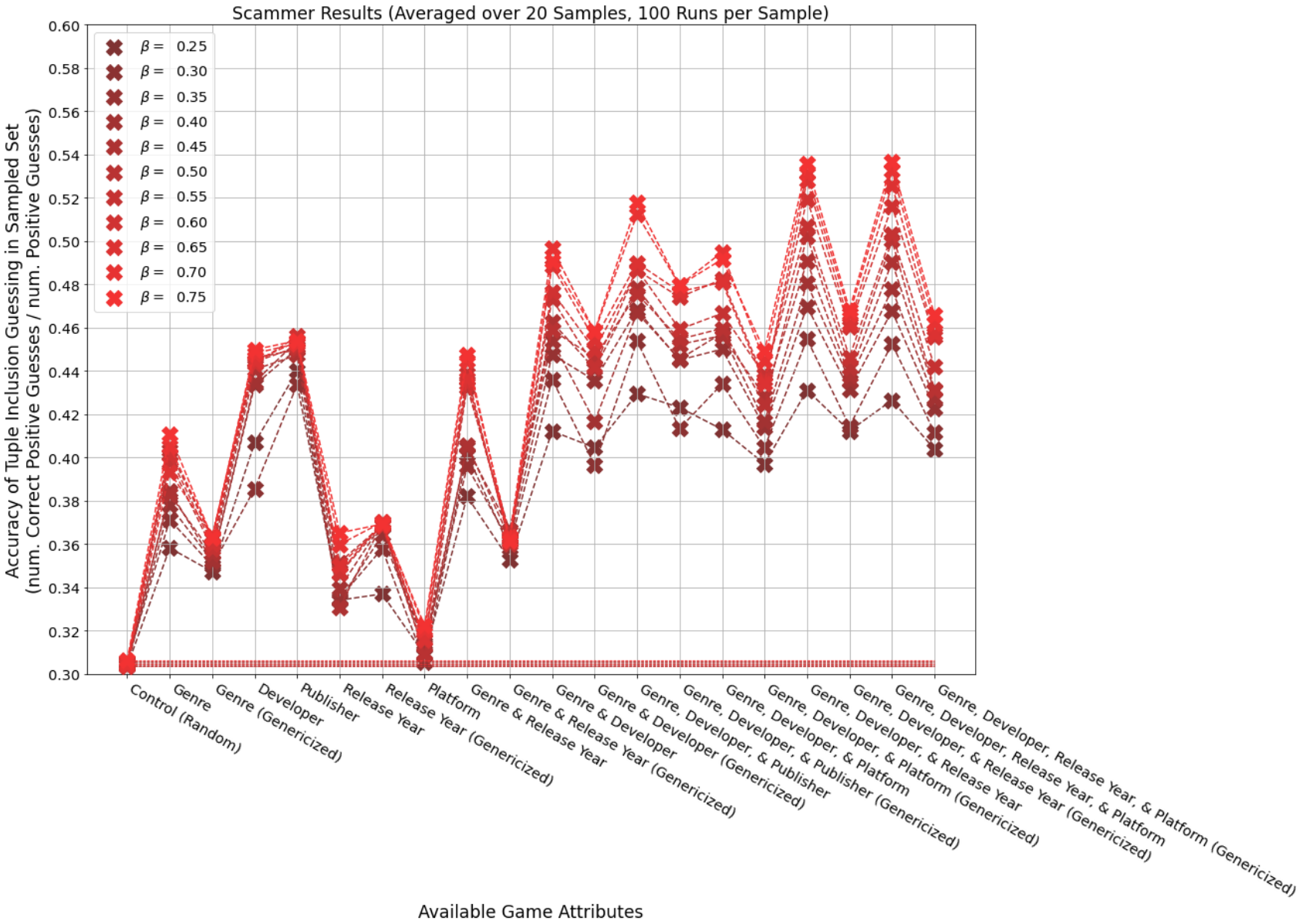
Anonymization Statistics: k-Levels



Metric 1: Scammer Deanonymization

- Scammer tries to identify if games are in the Inventory based on Samp.-Anon. set
- **Scammer Algorithm**
  - Anonymize the Interest Superset to match public attributes in Samp.-Anon. set
  - On all games in anonymous Interest Superset:
    - Flip a coin
    - Guess (proportionally) randomly on heads
    - On tails, compare proportions on the game between superset and sampled set
    - Guess **yes** if game occurs as much or more in sampled set against superset
  - Present guesses when done to an oracle, learn ratio of correct **yes** guesses to all **yes** guesses
    - *(Proportionally random: number of yes guesses equal to size of Inventory (see Ratio two slides back))

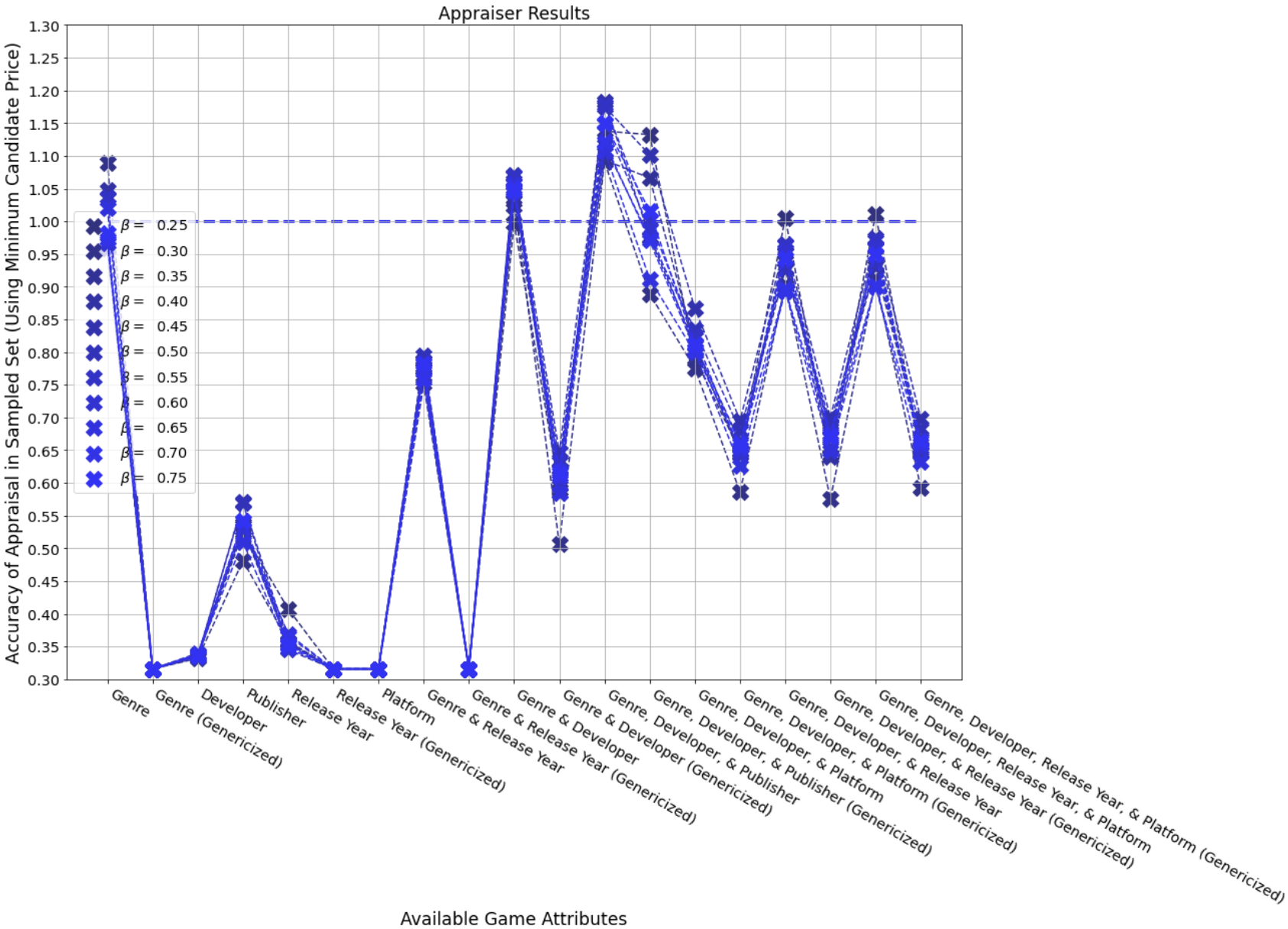
Scammer Results



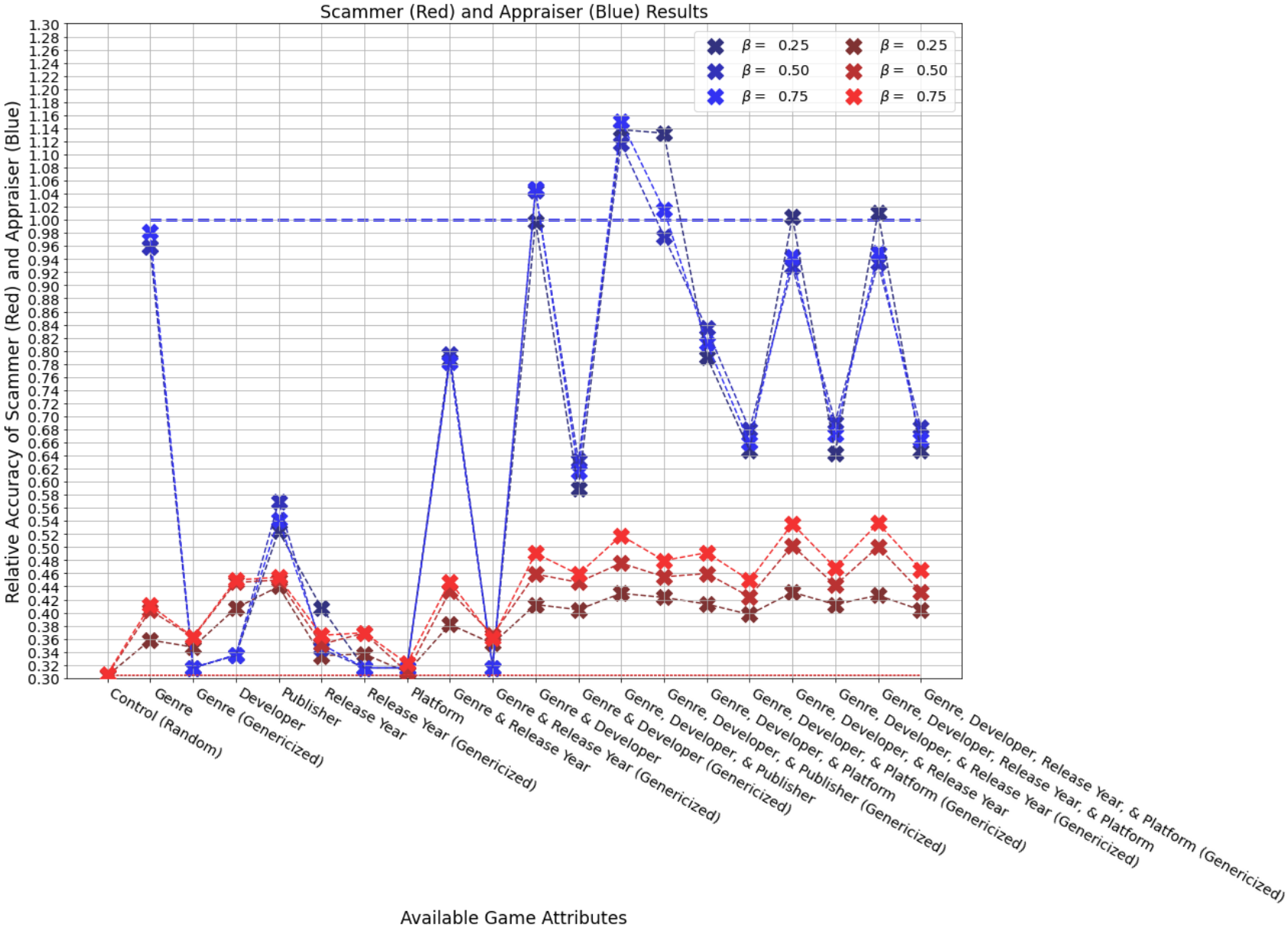
Metric 2: Appraiser Valuation

- Appraiser tries to compute the Full Value of the Inventory based on Samp.-Anon. set
- **Appraiser Algorithm**
  - Anonymize the Interest Superset to match public attributes in Samp.-Anon. set
  - On all games in Samp.-Anon. set:
    - Find all matching games in Interest Superset
    - Record the **minimum price** of matching games as Candidate Price
  - Compute the sum of all Candidate Prices
  - Multiply Candidate Price Sum by ratio of size of Inventory to Samp.-Anon. set
  - Present Scaled Candidate Price Sum to an oracle, learn ratio of prior to Actual Price Sum

Appraiser Results



Results Side-By-Side



Conclusions & What Is To Be Done

- With a maximum of +15 Scammer Accuracy
  - (**Genre**) & (**Genre**, **Release Year**) maximize Appraiser Accuracy irrespective of β
  - **Genre** alone is the best
    - *(both without Generalization, only Suppression)*
- (**Genre**, **Developer(s)**) is also desireable if β is kept well below 0.5
- k-Level clearly related to utility of Scammer/Appraiser, but not alone decisive
- As mentioned in Motivation, using Steam items (TF2 Unusual Hats, CSGO Knives, etc) would yield larger inventories
  - Results could be less affected by small k-Levels
  - What would be analogous attributes to Genre & Developer?
- Github link:
  https://github.com/cSquaerd/inventoryAnonProject.git

Bibliography

- **[1]** Ninghui Li, Wahbeh Qardaji, and Dong Su. 2012. On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security (ASIACCS '12)*. Association for Computing Machinery, New York, NY, USA, 32–33. DOI:https://doi.org/10.1145/2414456.2414474
- **[2]** Latanya Sweeney. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 10, 5 (October 2002), 571–588. DOI:https://doi.org/10.1142/S021848850200165X
- **[3]** Cynthia Dwork. 2006. Differential privacy. In *Proceedings of the 33rd international conference on Automata, Languages and Programming - Volume Part II (ICALP'06)*. Springer-Verlag, Berlin, Heidelberg, 1–12. DOI:https://doi.org/10.1007/11787006_1

EOF