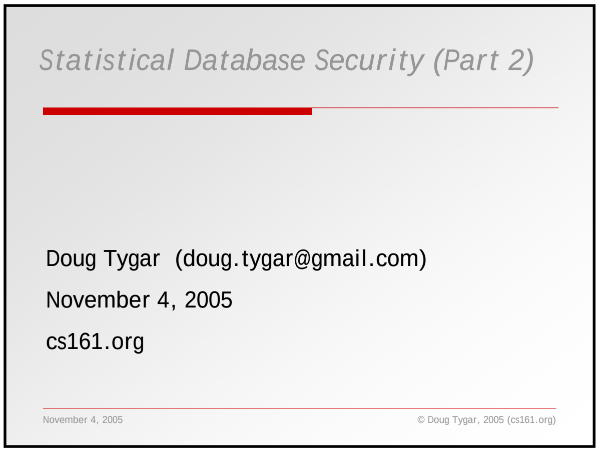
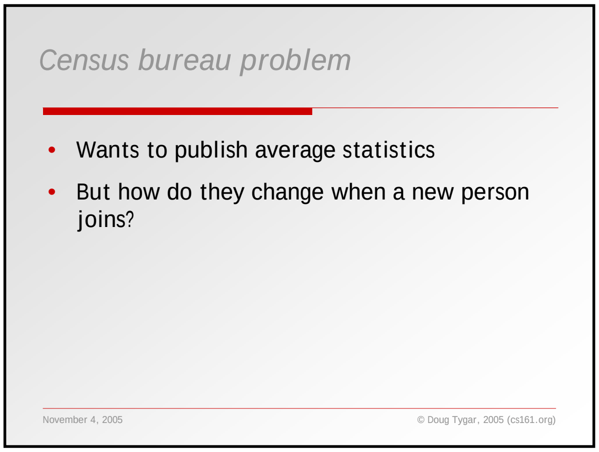


Statistical Database Security (Part 2)

Doug Tygar (doug.tygar@gmail.com)

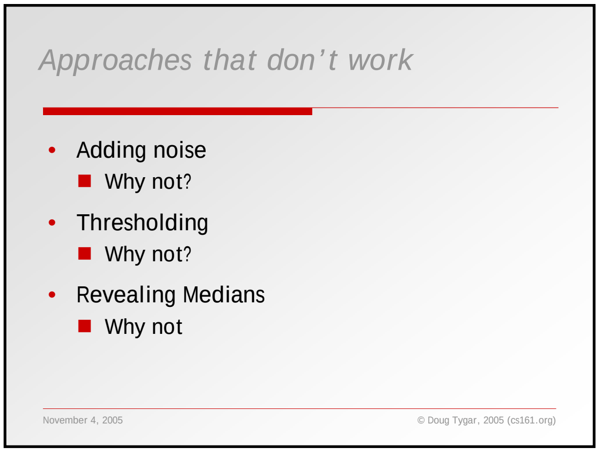
November 4, 2005

cs161.org



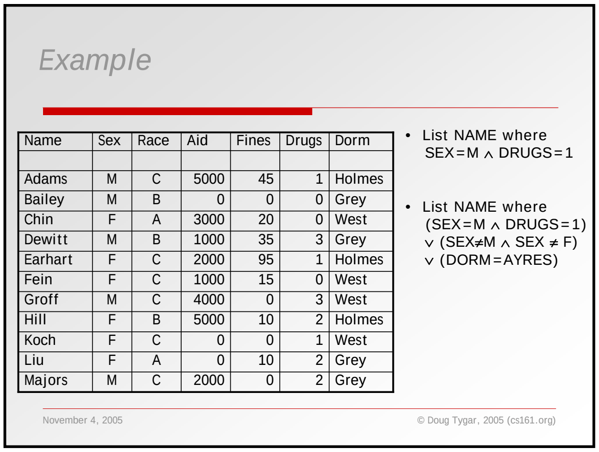
Census bureau problem

- Wants to publish average statistics
- But how do they change when a new person joins?



Approaches that don't work

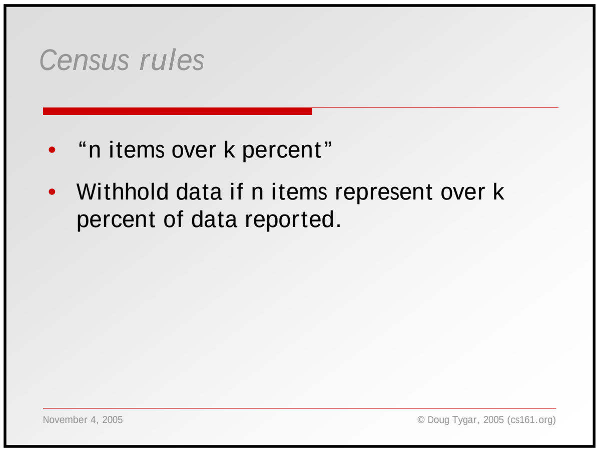
- Adding noise
  - Why not?
- Thresholding
  - Why not?
- Revealing Medians
  - Why not



Example

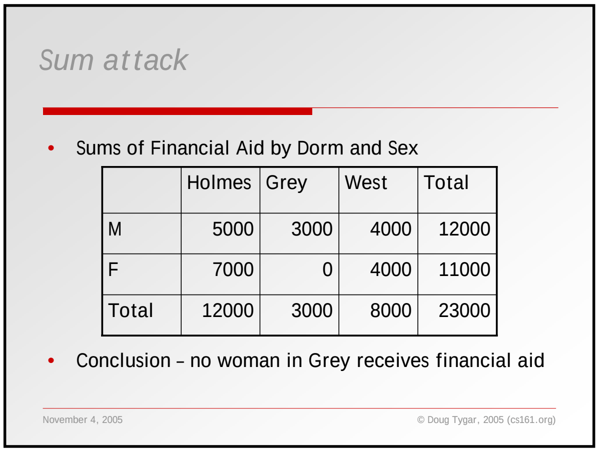
Name	Sex	Race	Aid	Fines	Drugs	Dorm
Adams	M	C	5000	45	1	Holmes
Bailey	M	B	0	0	0	Grey
Chin	F	A	3000	20	0	West
Dewitt	M	B	1000	35	3	Grey
Earhart	F	C	2000	95	1	Holmes
Fein	F	C	1000	15	0	West
Groff	M	C	4000	0	3	West
Hill	F	B	5000	10	2	Holmes
Koch	F	C	0	0	1	West
Liu	F	A	0	10	2	Grey
Majors	M	C	2000	0	2	Grey

- List NAME where SEX=M ∧ DRUGS=1
- List NAME where (SEX=M ∧ DRUGS=1) ∨ (SEX≠M ∧ SEX ≠ F) ∨ (DORM=AYRES)



Census rules

- "n items over k percent"
- Withhold data if n items represent over k percent of data reported.

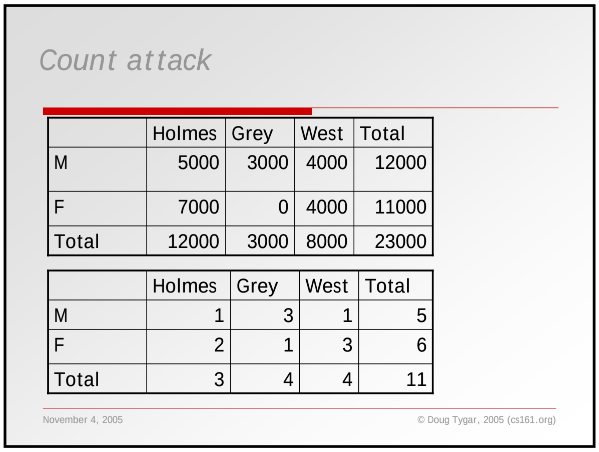


Sum attack

- Sums of Financial Aid by Dorm and Sex

	Holmes	Grey	West	Total
M	5000	3000	4000	12000
F	7000	0	4000	11000
Total	12000	3000	8000	23000

- Conclusion – no woman in Grey receives financial aid

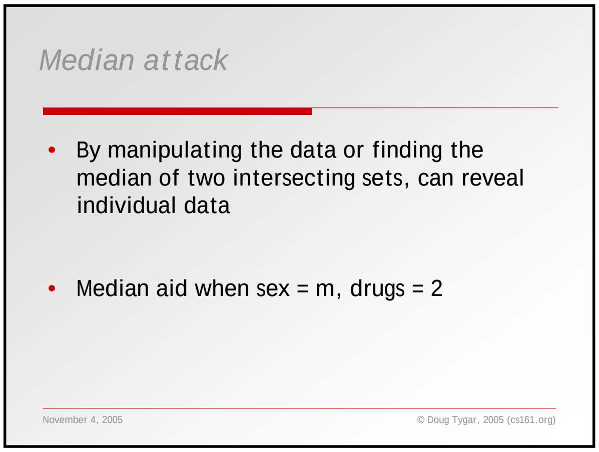


Count attack

	Holmes	Grey	West	Total
M	5000	3000	4000	12000
F	7000	0	4000	11000
Total	12000	3000	8000	23000

	Holmes	Grey	West	Total
M	1	3	1	5
F	2	1	3	6
Total	3	4	4	11

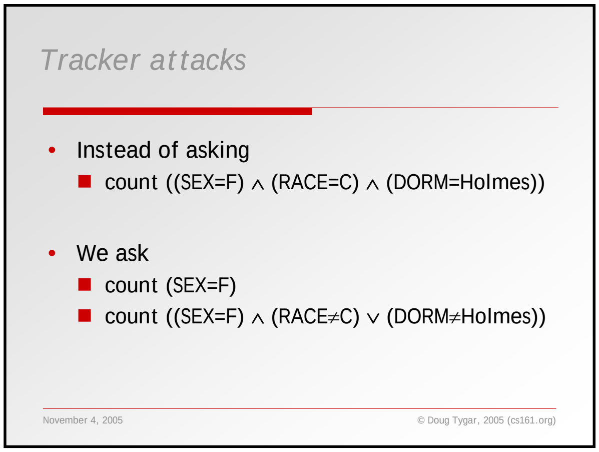
November 4, 2005 © Doug Tygar, 2005 (cs161.org)



Median attack

- By manipulating the data or finding the median of two intersecting sets, can reveal individual data
- Median aid when sex = m, drugs = 2

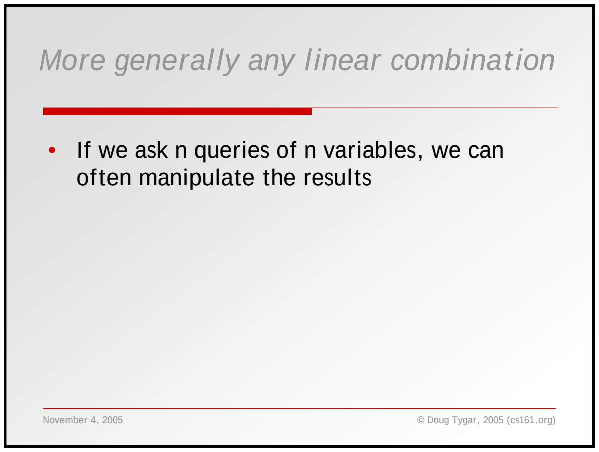
November 4, 2005 © Doug Tygar, 2005 (cs161.org)



Tracker attacks

- Instead of asking
  - count ((SEX=F) ∧ (RACE=C) ∧ (DORM=Holmes))
- We ask
  - count (SEX=F)
  - count ((SEX=F) ∧ (RACE≠C) ∨ (DORM≠Holmes))

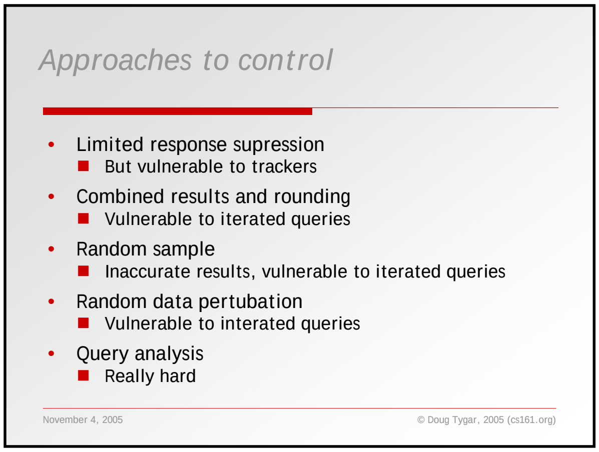
November 4, 2005 © Doug Tygar, 2005 (cs161.org)



More generally any linear combination

- If we ask n queries of n variables, we can often manipulate the results

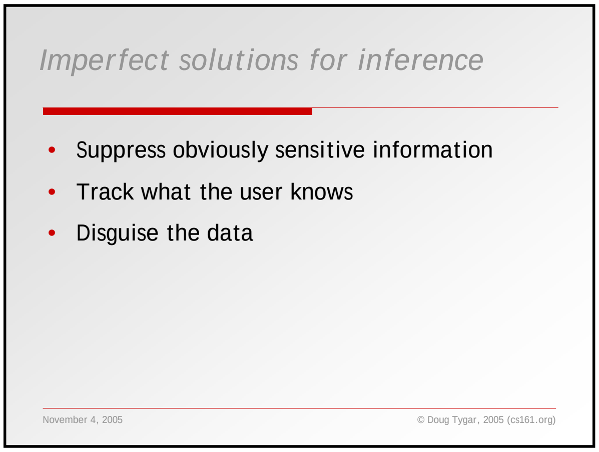
November 4, 2005 © Doug Tygar, 2005 (cs161.org)



Approaches to control

- Limited response supression
  - But vulnerable to trackers
- Combined results and rounding
  - Vulnerable to iterated queries
- Random sample
  - Inaccurate results, vulnerable to iterated queries
- Random data pertubation
  - Vulnerable to interated queries
- Query analysis
  - Really hard

November 4, 2005 © Doug Tygar, 2005 (cs161.org)



Imperfect solutions for inference

- Suppress obviously sensitive information
- Track what the user knows
- Disguise the data

November 4, 2005 © Doug Tygar, 2005 (cs161.org)