Fighting Coordinated Attackers with Cross-Organizational Information Sharing

Ethan Blanton

November 30, 2006

Mark Allman, Ethan Blanton, Vern Paxson, and Scott Shenker

The Problem

- “Botnets” represent a real problem on the Internet
- We’d rather find them *before* they DDoS, *etc*.
- Finding individual infected hosts is a better understood problem
  - Bro, Snort, . . .
- Correlating large numbers of infected hosts into a botnet requires a vantage on those hosts

The Problem

- “Botnets” represent a real problem on the Internet
- We’d rather find them *before* they DDoS, *etc*.
- Finding individual infected hosts is a better understood problem
  - Bro, Snort, . . .
- Correlating large numbers of infected hosts into a botnet requires a vantage on those hosts
- *What if we all work together to identify botnets?*

Fighting Crime

The Real World:

- Detectives
- Witnesses

Fighting Crime

The Real World:

- Detectives
- Witnesses

Our World:

- Detectives: Honeypots, IDSs, Firewalls
- Witnesses: Packet traces, Netflow records, Server logs

Fighting Crime (cont'd)

- Savvy monitors (detectives) are few, well-known, and trusted — *depth*
  - Say, installed at major ISPs or dense POPs
  - These guys are really smart about finding ne'er-do-wells
- Witnesses are many and untrusted — *breadth*
  - The flow records you're keeping anyway
  - These guys don't make value judgments

Fighting Crime (cont’d)

- Savvy monitors (detectives) are few, well-known, and trusted — *depth*
  - Say, installed at major ISPs or dense POPs
  - These guys are really smart about finding ne’er-do-wells
- Witnesses are many and untrusted — *breadth*
  - The flow records you’re keeping anyway
  - These guys don’t make value judgments
  - *“Confine yourself to the facts, please”*

Problems with Sharing

People are reluctant to share:

- Resource commitment
- Privacy concerns
  - User privacy
  - Operational privacy

Resource Commitment

We collect *lots of information* anyway

- For performance monitoring
- For provisioning
- For security audits
- For research

Resource Commitment

We collect *lots of information* anyway

- For performance monitoring
- For provisioning
- For security audits
- For research

Let's leverage that, not collect something new.

Privacy Concerns

- Detectives are trusted, but not too far
  - Witnesses can wait for a threshold of detectives to ask the same question before answering
  - Witnesses don't have to answer questions they don't like
- Loose Private Matching
  - Witnesses don't know what detectives are looking for
  - Detectives can't glean extra information
  - If both parties really are talking about the same thing, tightly scoped information is shared

Loose Private Matching

You can only answer my question if you know what I’m talking about

$$T_1 = \{d, t, p\}$$

Loose Private Matching

You can only answer my question if you know what I’m talking about

$$T_1 = \{d, t, p\}$$

- “Questions” are ambiguous

$$Q = H(T_1)$$

Loose Private Matching

You can only answer my question if you know what I'm talking about

$$T_1 = \{d, t, p\}$$

- "Questions" are ambiguous

$$Q = H(T_1)$$

- Answers are encrypted

$$A = E(T_1, d, ...)$$

Choosing H

- It should be hard to guess the inputs
- The probability of collision should be high

Choosing H

- It should be hard to guess the inputs
- The probability of collision should be high

Consider:

$$(d_0 \cdot d_1 \cdot d_2 \cdot d_3 \cdot t \cdot p_0 \cdot p_1)\ mod\, 256$$

Choosing H

- It should be hard to guess the inputs
- The probability of collision should be high

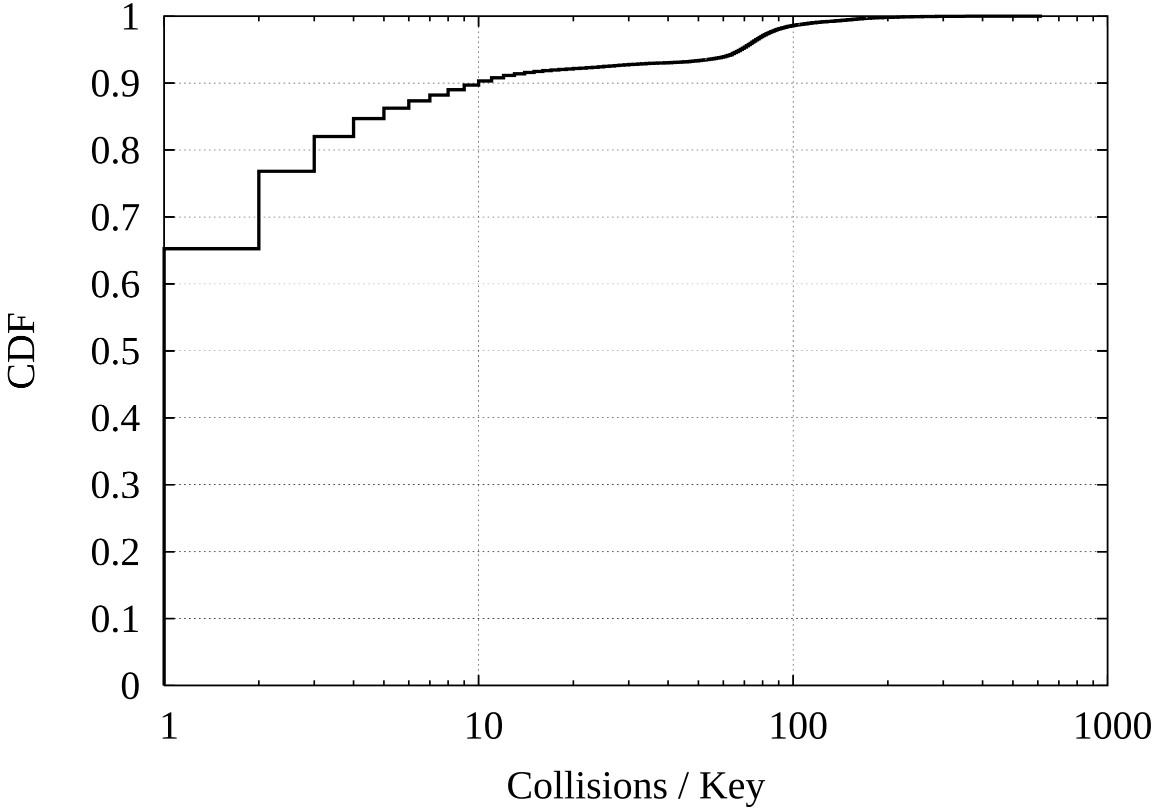
Consider:

$$(d_0 \cdot d_1 \cdot d_2 \cdot d_3 \cdot t \cdot p_0 \cdot p_1) \; mod \, 256$$

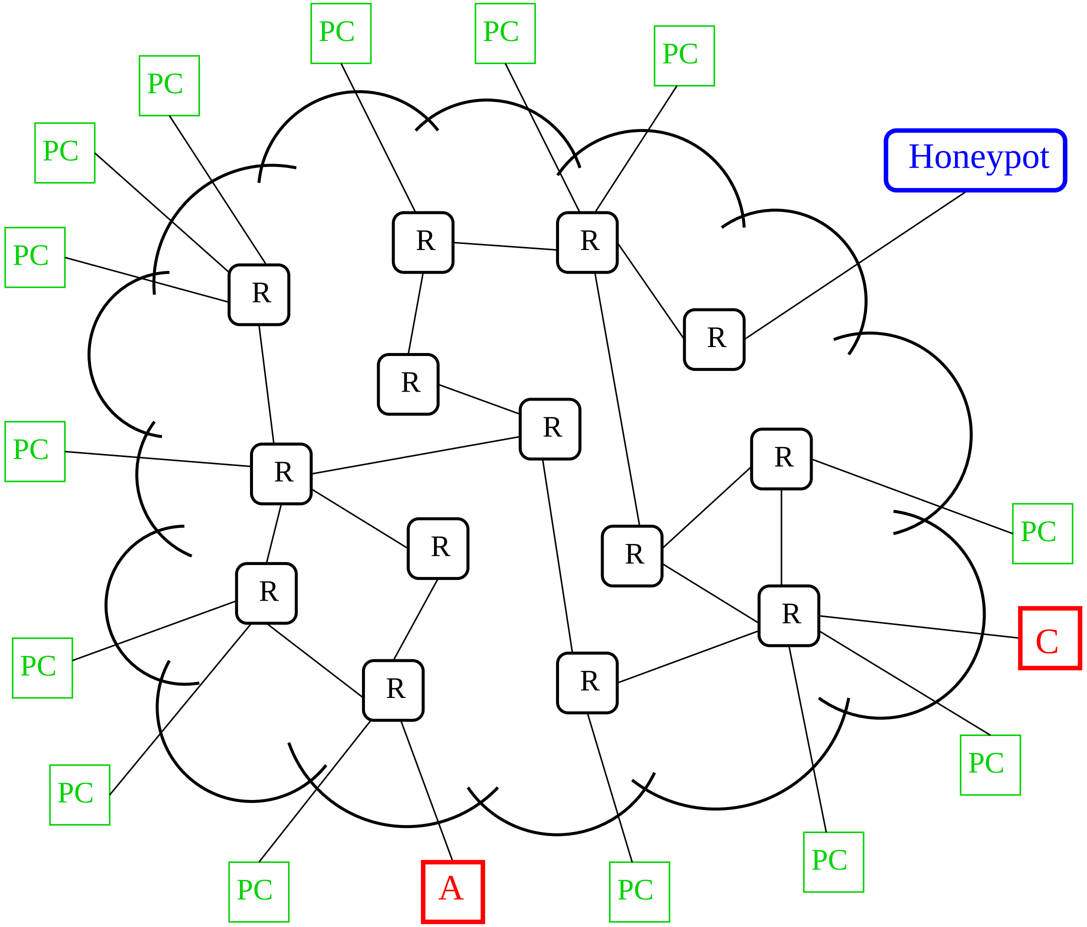
$$a.b.c.d = a.c.b.d = a.2 \cdot b.c/2.d$$

Choosing H (cont'd)

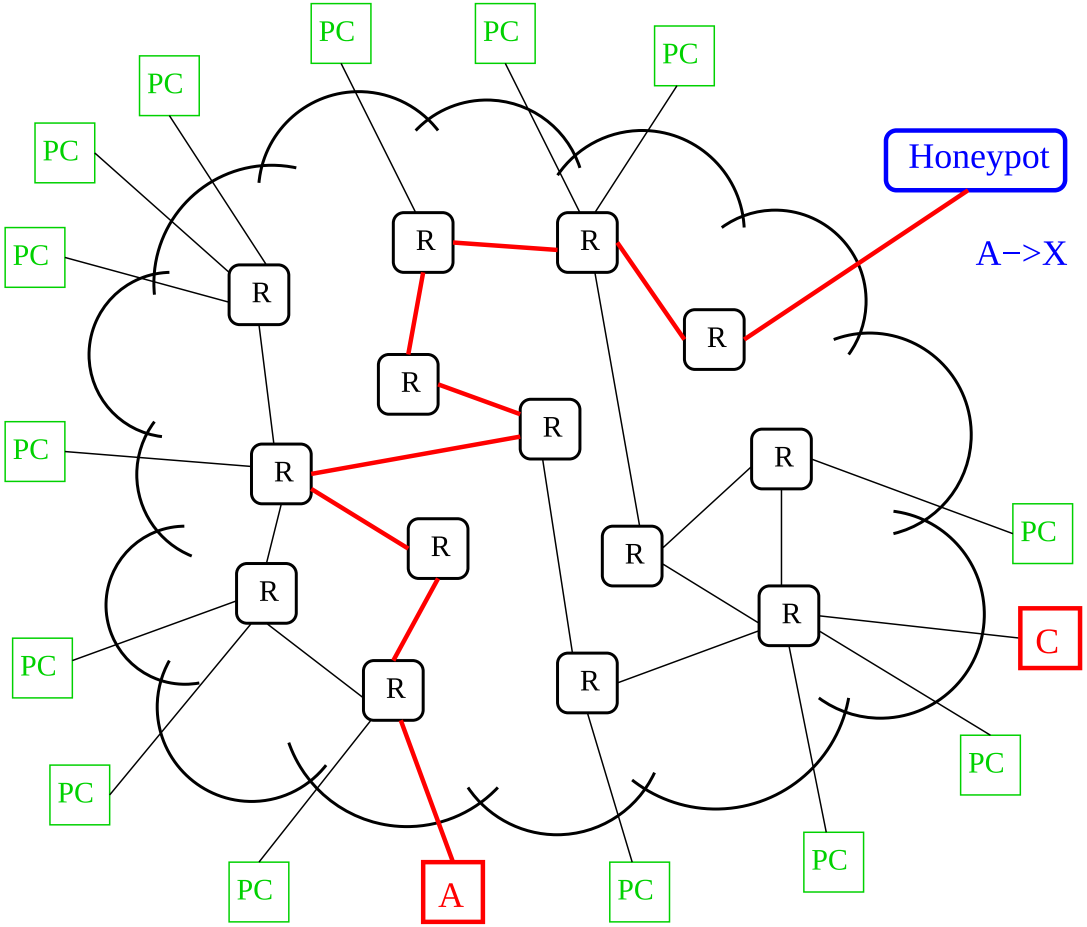
Only 11% of connections hash to a unique key



An Overview

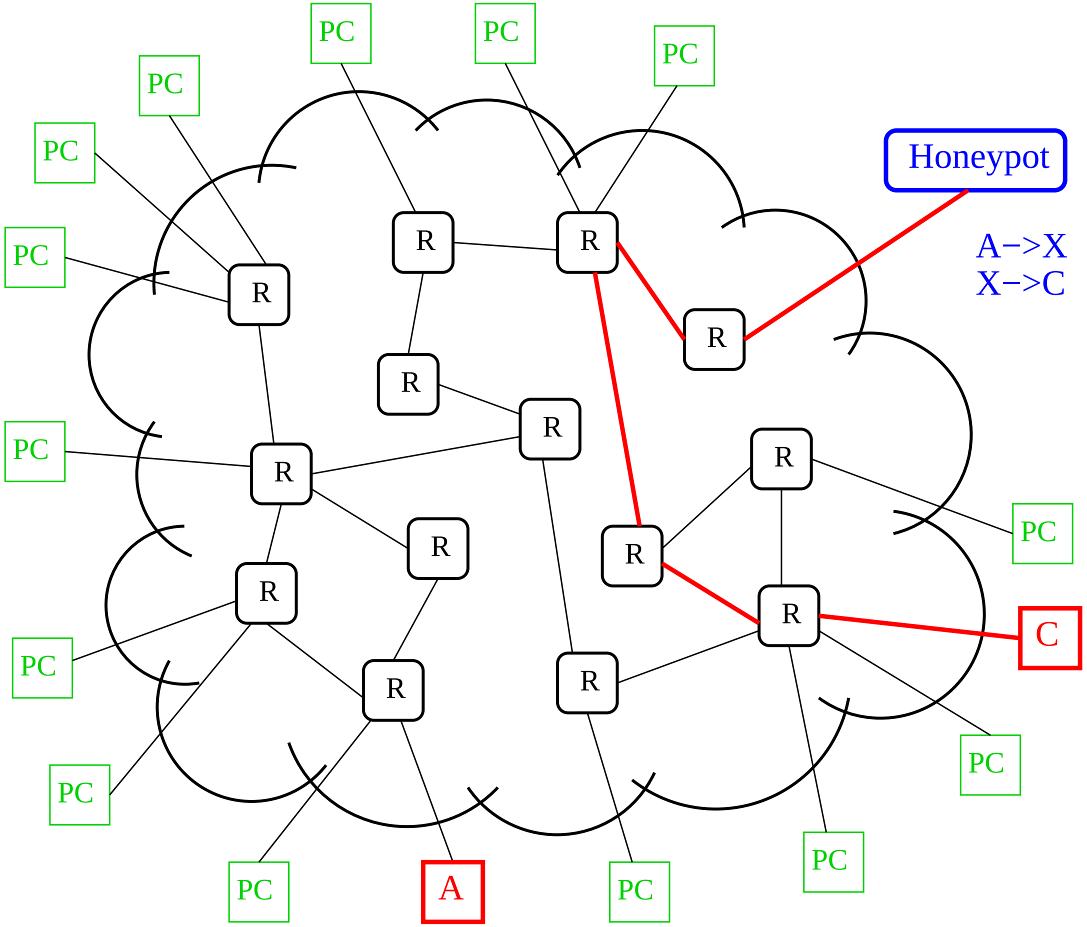

PC
PC
PC
PC
PC
PC
Honeypot
R
R
R
R
R
R
R
PC
R
R
R
R
PC
PC
R
R
R
R
C
PC
PC
PC
PC
A
PC
PC

An Overview (cont'd)



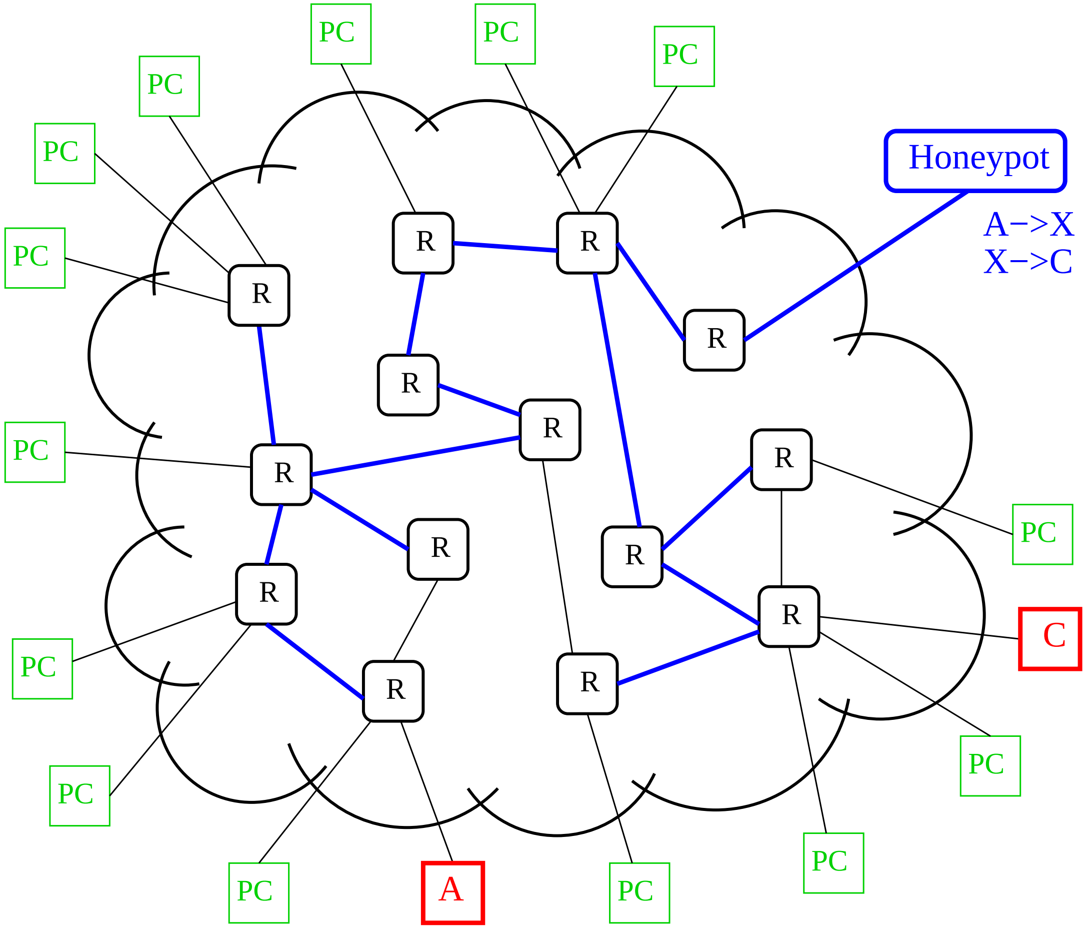
$$T_1 = \{s_A, t_1, p_1\}$$

An Overview (cont'd)



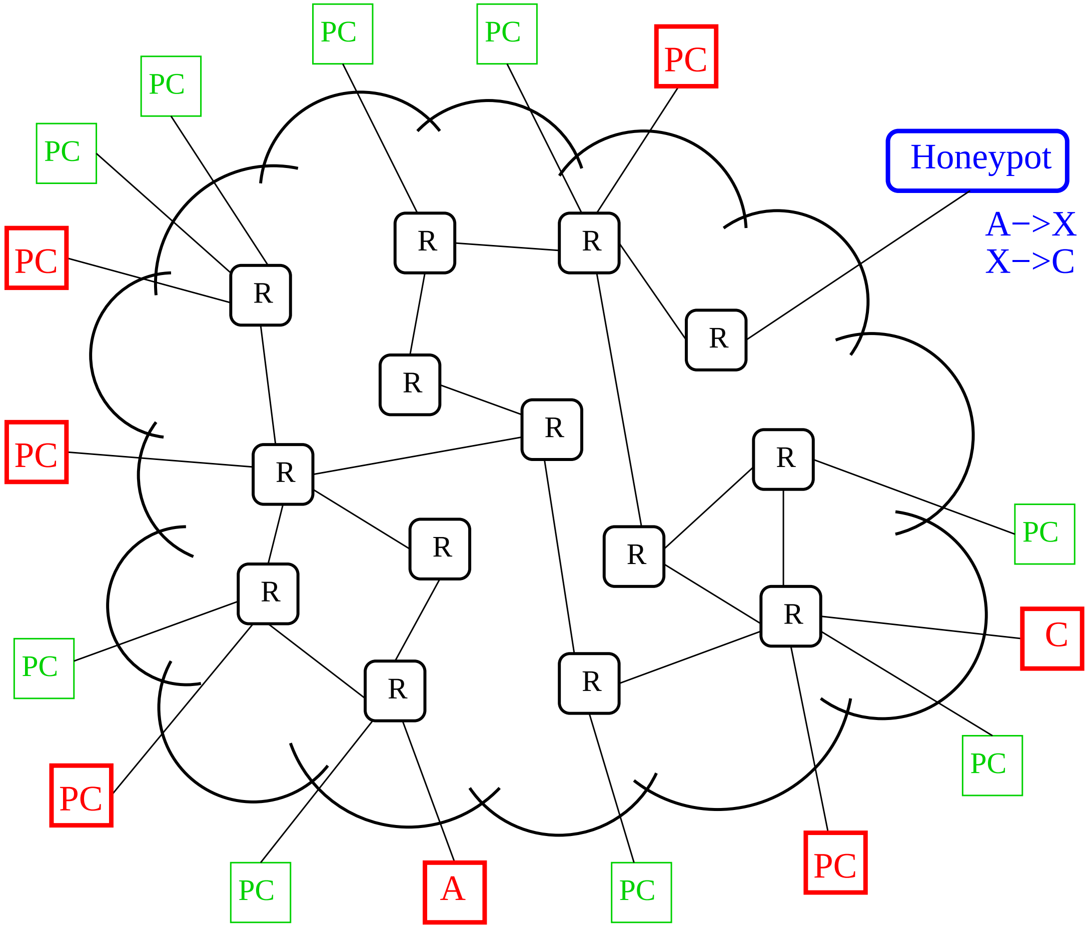
$$T_1 = \{s_A, t_1, p_1\}\,, \quad T_2 = \{d_C, t_2, p_2\}$$

An Overview (cont'd)



"$H(T_1) \cup H(T_2)$?"

An Overview (cont'd)



$$\text{“}E\,(T_1||T_2, \{PC_1, ..., PC_n\}).\text{”}$$

Cheating the System

- Detectives can ask all the wrong questions

“Have you seen any politicians frequenting that unsavory gentlemen’s club lately?”

Cheating the System

- Detectives can ask all the wrong questions

  “Have you seen any politicians frequenting that unsavory gentlemen’s club lately?”

- Witnesses can make up answers

  “I totally saw my competitor down at the docks last night”

Cheating the System

- Detectives can ask all the wrong questions

  "Have you seen any politicians frequenting that unsavory gentlemen's club lately?"

- Witnesses can make up answers

  "I totally saw my competitor down at the docks last night"

- Witnesses can withold answers

  "I've never seen the bathroom of that crackhouse, officer. ***Wait****, I mean,* ***what crackhouse****?"*

Detectives Cheating

There’s not a lot we can do about this

- Witnesses can threshold questions, and keep track of detectives that ask a lot of “funny” questions
- We have to assume that most detectives are trustworthy

Witnesses Cheating

This we can do more about

- Made-up answers won't decrypt properly, because the witness can't figure out T_i
- Bogus additions to legitimate responses (where the witness *does* know T_i) can be mitigated again with thresholds
- Withheld answers only hurt if no one else saw it

Sharing the Data

- Detectives could report to a central authority
  - “Network CDC”
- ISPs and enterprises could provide “Detective services” to their own networks
- Detectives could populate a global bad behavior database used for all sorts of bad activities

Conclusions and Future Work

Conclusions:

- We can use existing infrastructure to do collaboration
- The "barrier to entry" for this collaboration can be lowered with some tricks to improve privacy and reduce information leakage

Future Work:

- *"So full of holes it will probably generate plenty of discussion"*
- $H()$ is an open question — we threw a straw man out there
- Getting the queries right will require some more information
- Is it safe enough that people will *use* it?

Questions

Thanks for Listening

Any Questions?

http://www.cs.purdue.edu/homes/eblanton/slides/hotnets06.pdf