REDJACK

How (Not) to Apply Differential Privacy to Anonymity Networks

Scott Coull
scott.coull@redjack.com

Differential Privacy

REDJACK

The gold standard in privacy definitions

- **Some popular interpretations:**
  - **No assumptions on the data**
  - **Protection even when attacker knows all but one of the rows in the database**
  - **Robust to arbitrary background knowledge**

Differential Privacy

REDJACK

The gold standard in privacy definitions

misinterpretations

- **Some popular ~~interpretations~~:**
  - **No assumptions on the data**
  - **Protection even when attacker knows all but one of the rows in the database**
  - **Robust to arbitrary background knowledge**

Computer Networks & Differential Privacy

REDJACK

Many “objects” in network data

Many relationships among the “objects”

Relationships specified by underlying protocols and hidden processes

Implicit independence assumption

REDJACK

Objects in Computer Networks

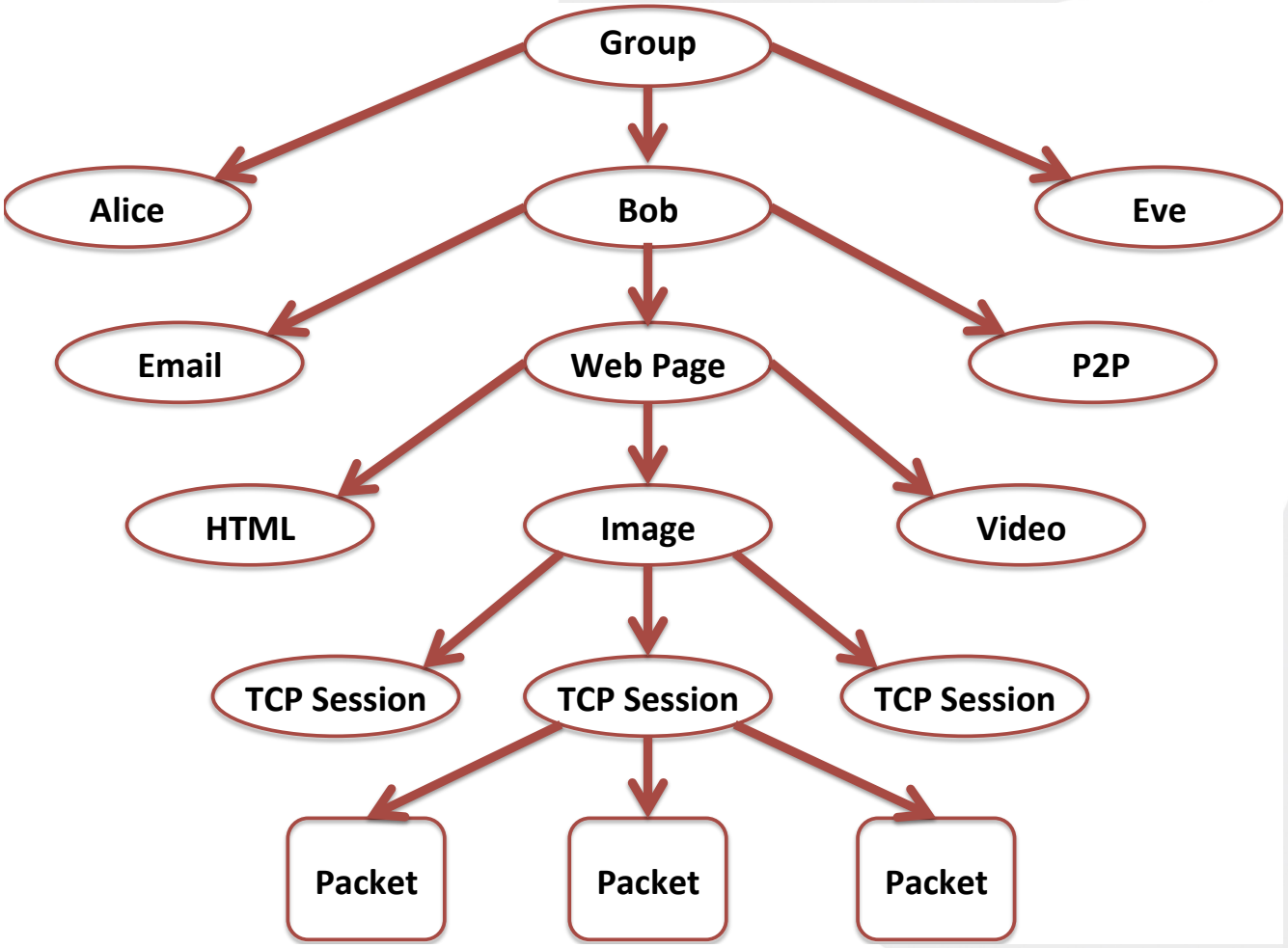
Objects include users, workstations, web pages, emails, flows, packets ...

Protocols impose structure on network data and govern interaction of objects

One way to represent our understanding of the structure via an *ontology*

REDJACK

Objects in Computer Networks



REDJACK

Objects in Computer Networks

Many such ontologies may exist

May not even be a hierarchy

Objects carry information about both ancestors and descendants

Adversary may have a more complete (complex) “view” than we do

Independence Assumptions

REDJACK

- Differential privacy developed in the context of databases of individuals
- What happens when records imply each other's presence?
  - Nothing good

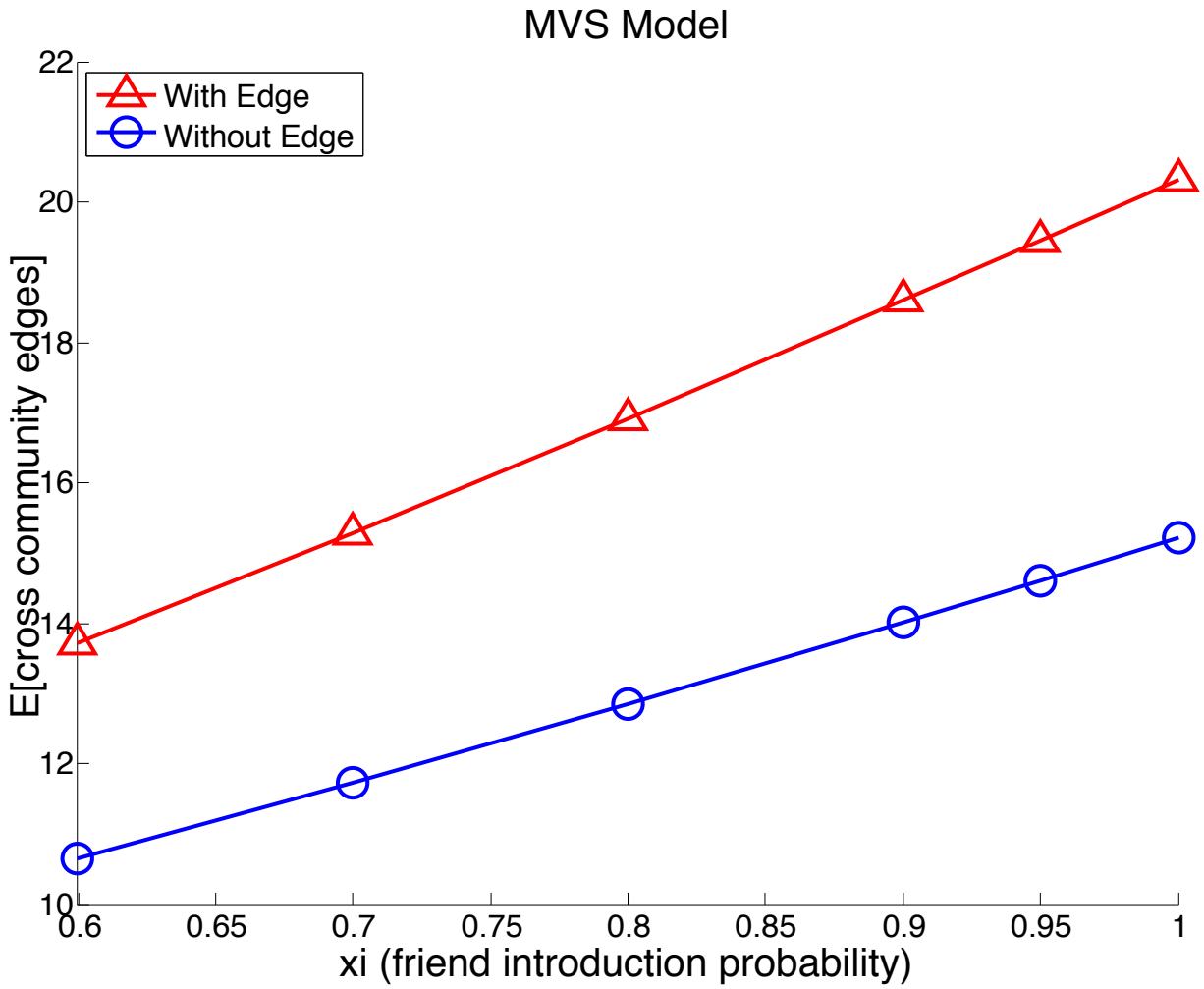
Independence Assumptions

REDJACK

- **No Free Lunch in Data Privacy[1]**
  - **Knowledge of correlations leads to failure**
  - **Example: Remove edge from a social network graph; growth pattern changes significantly**

[1] Daniel Kifer and Ashwin Machanavajjhala. No Free Lunch in Data Privacy. In ACM SIGMOD, pages 193–204, 2011.

Independence Assumptions



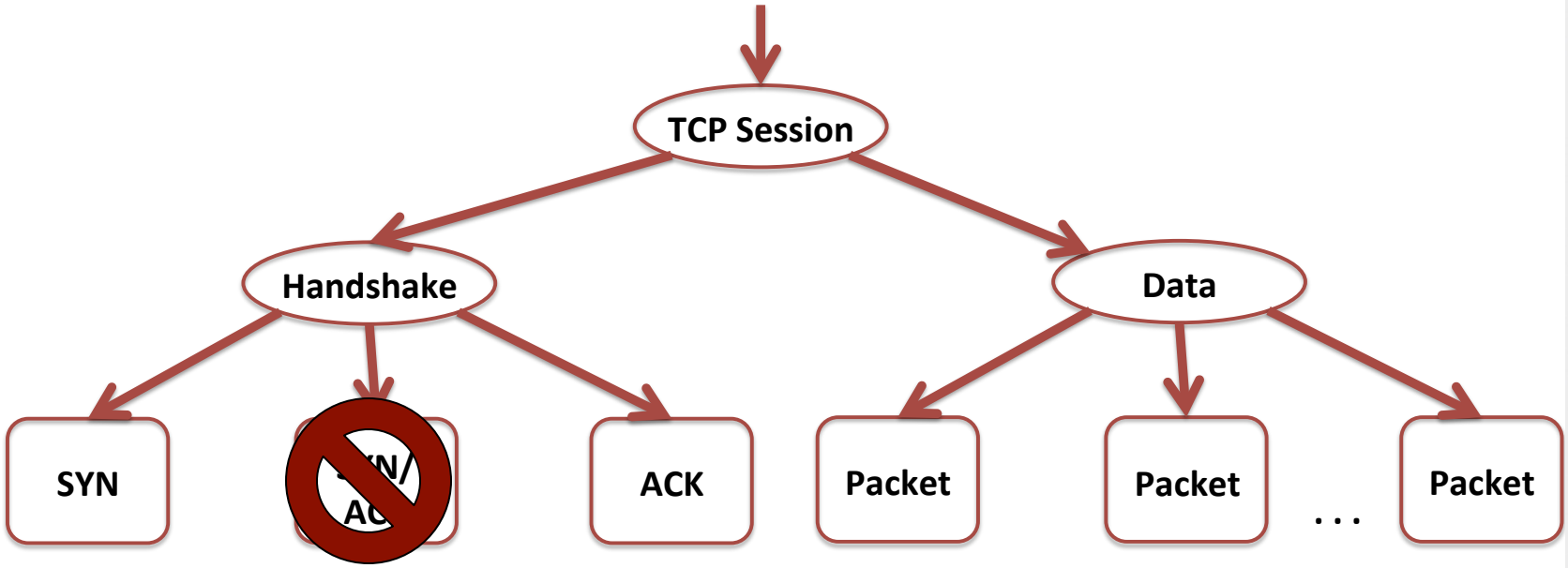
Independence Assumptions

REDJACK

- Same problems with network data…
  - Ontology describes exactly the correlations we can use to infer whether object is present
  - Example: Remove a TCP handshake packet; total traffic volume should change
- If we ignore these semantics the *best* we get is effectively *packet privacy*

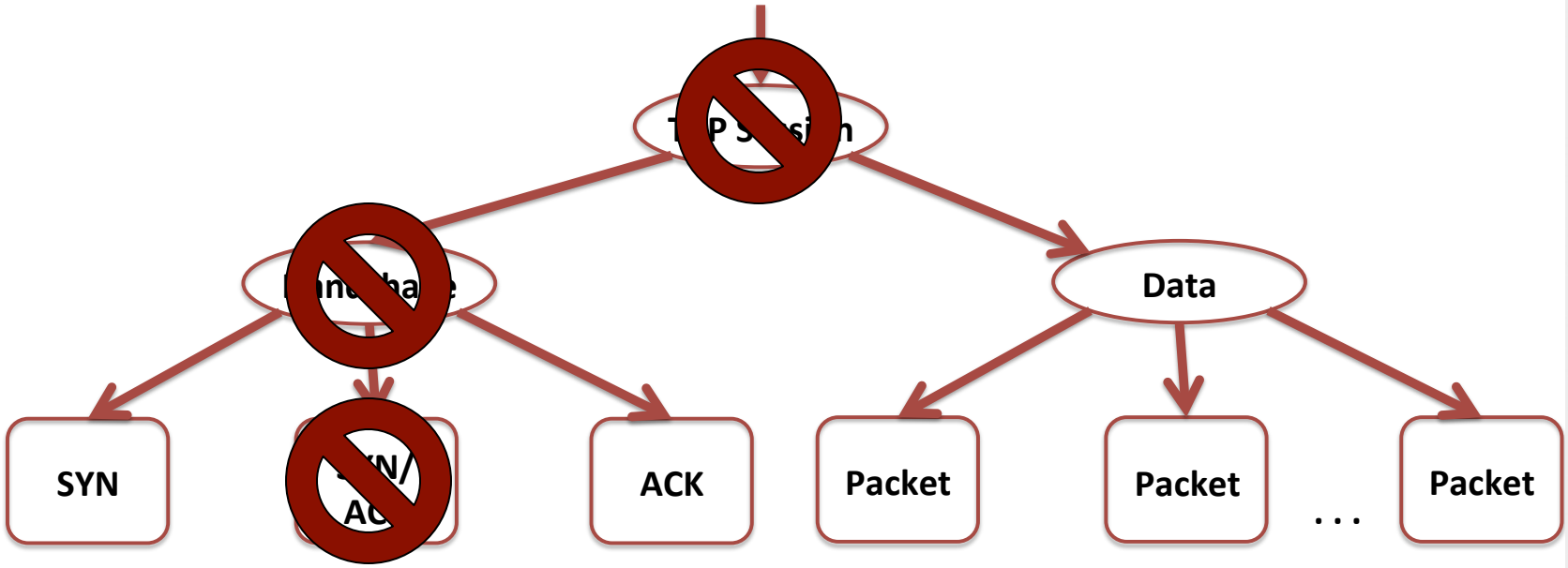
Independence Assumptions

REDJACK



Independence Assumptions

REDJACK



REDJACK

The Pufferfish Framework[2]

- **Generalization of differential privacy that explicitly states assumptions:**
  - **Objects we are trying to protect –** $\mathbb{S}$
  - **Mutually exclusive secret pairs**

$$\mathbb{S}_{pairs} \subset \mathbb{S} \times \mathbb{S}$$

  - **Set of data generating distributions –** $\mathbb{D}$

[2] Daniel Kifer and Ashwin Machanavajjhala. A Rigorous and Customizable Framework for Privacy. In PODS, pages 77–88, 2012.

The Pufferfish Framework

REDJACK

- **For all possible outputs** $\omega \in range(\mathcal{M})$
- **For all pairs of potential secrets**

$$(s_i, s_j) \in \mathbb{S}_{pairs}$$

- **For all distributions** $\theta \in \mathbb{D}$

$$P(\mathcal{M}(Data) = \omega | s_i, \theta) \leq e^{\epsilon} P(\mathcal{M}(Data) = \omega | s_j, \theta)$$
$$P(\mathcal{M}(Data) = \omega | s_j, \theta) \leq e^{\epsilon} P(\mathcal{M}(Data) = \omega | s_i, \theta)$$

The Pufferfish Framework

REDJACK

- **Also supports a nice semantic interpretation of the definition:**

$$e^{-\epsilon} \leq \frac{P(s_i|\mathcal{M}(Data) = \omega, \theta)}{P(s_j|\mathcal{M}(Data) = \omega, \theta)} \Big/ \frac{P(s_i|\theta)}{P(s_j|\theta)} \leq e^{\epsilon}$$

- **Adversary's belief in s_i changes to at most $e^{\epsilon}\alpha$ and at least $e^{-\epsilon}\alpha$**

Challenges

REDJACK

- How do we define the data generation distributions?
  - Network data is notoriously difficult to model
  - Is it possible to design the protocol to make this easier for us?
  - G. Danezis et al.'s MCMC sampling?

Challenges

REDJACK

- How do we define the private algorithm (M) for anonymity networks?
  - For general network data we are probably out of luck since there is just so much to measure
  - For encrypted traffic we have a more restrictive set of measurements – time, size

Final Thoughts

REDJACK

- Can we think about multiple attacker models at once, in a single metric?
  - Yes, and we probably have to if we want to move away from the break-fix cycle.
  - I think Pufferfish accommodates for this in the secrets and data generating distributions

Final Thoughts

REDJACK

- Are Bayesian approaches used for mixes extensible to Tor-like systems?
  - Yes…probably…maybe?
  - A Bayesian view on the problem is likely to yield semantically-meaningful guarantees.
  - Pufferfish definitions have a well-defined Bayesian interpretation that actually tells us something useful.

Final Thoughts

REDJACK

- **Should differential privacy be an inspiration for this space?**
  - **Most other commonly used definitions, like k-anonymity, are vulnerable to direct attack.**
  - **Differential privacy at least offers the possibility of a strong guarantee.**
  - **Applying it to anonymity networks is going to be tricky due to difficulty in modeling the data.**

Final Thoughts

REDJACK

- What is the role of user modeling in this space?
  - Seems to be very important.
  - Some model is going to be necessary to make reasonable assumptions for definitions.
  - It is not just about user modeling, but modeling general relationships in the data.