Two applications

(this is NOT an intro to provenance)

Dan Suciu
University of Washington

Two Applications

1. Provenance summaries for query answering using probabilistic views
    - With Chris Re
    - Status: ongoing

2. Provenance for privacy in RFID applications
    - With Vibhor Rastogi
    - Status: preliminary

Query Answering Using Views

V(x) = R(x,y),S(x,y,z),T(x,z)

Materialize:

V =

x
a
c
b
f

Query:

q = R(x,y),S(x,y,z),T(x,z),U(x,v),K(v,w)

Rewrite to:

q = V(x), U(x,v),K(v,w)

More efficient !

Using *Probabilistic* Views

R^p:

x	y	P
a	m	**0.3**
a	n	**0.2**
b	m	**0.4**
b	p	**0.1**

S^p:

x	y	z	P
a	m	s	**0.1**
a	n	s	**0.5**
b	m	t	**0.4**
b	p	t	**0.9**

T^p:

x	z	P
a	s	**0.3**
b	s	**0.2**
b	t	**0.4**

V^p:

$$V(x) = R^p(x,y),S^p(x,y,z),T^p(x,z)$$

Using *Probabilistic* Views

R^p:

x	y	P
a	m	**0.3**
a	n	**0.2**
b	m	**0.4**
b	p	**0.1**

S^p:

x	y	z	P
a	m	s	**0.1**
a	n	s	**0.5**
b	m	t	**0.4**
b	p	t	**0.9**

T^p:

x	z	P
a	s	**0.3**
b	s	**0.2**
b	t	**0.4**

V^p:

x	P
a	**0.1**
b	**0.5**

Marginal probabilities

$$\boxed{V(x) = R^p(x,y), S^p(x,y,z), T^p(x,z)}$$

Using *Probabilistic* Views

R^p:

x	y	P
a	m	0.3
a	n	0.2
b	m	0.4
b	p	0.1

S^p:

x	y	z	P
a	m	s	0.1
a	n	s	0.5
b	m	t	0.4
b	p	t	0.9

T^p:

x	z	P
a	s	0.3
b	s	0.2
b	t	0.4

V^p:

x	P
a	0.1
b	0.5

Marginal probabilities

$V(x) = R^p(x,y), S^p(x,y,z), T^p(x,z)$

$q = V(x),\ \ U(x,v), K(v,w)$

Marginal Prob in V^p insufficient

Enter Provenance

R^p:

x	y	E
a	m	**E1**
a	n	**E2**
b	m	**E3**
b	p	**E4**

S^p:

x	y	z	E
a	m	s	**F1**
a	n	s	**F2**
b	m	t	**F3**
b	p	t	**F4**

T^p:

x	z	E
a	s	**G1**
b	s	**G2**
b	t	**G3**

V^p:

$$V(x) = R^p(x,y), S^p(x,y,z), T^p(x,z)$$

Enter Provenance

R^p:

x	y	E
a	m	**E1**
a	n	**E2**
b	m	**E3**
b	p	**E4**

S^p:

x	y	z	E
a	m	s	**F1**
a	n	s	**F2**
b	m	t	**F3**
b	p	t	**F4**

T^p:

x	z	E
a	s	**G1**
b	s	**G2**
b	t	**G3**

V^p:

x	E
a	$\mathbf{E1 \wedge F1 \wedge G1 \vee E2 \wedge F2 \wedge G1}$
b	$\mathbf{E3 \wedge F3 \wedge G3 \vee E4 \wedge F4 \wedge G3}$

Provenance
[Trio: “lineage”]

$V(x) = R^p(x,y), S^p(x,y,z), T^p(x,z)$

Enter Provenance

R^p:

x	y	E
a	m	**E1**
a	n	**E2**
b	m	**E3**
b	p	**E4**

S^p:

x	y	z	E
a	m	s	**F1**
a	n	s	**F2**
b	m	t	**F3**
b	p	t	**F4**

T^p:

x	z	E
a	s	**G1**
b	s	**G2**
b	t	**G3**

V^p:

x	E
a	**E1∧F1∧G1∨E2∧F2∧G1**
b	**E3∧F3∧G3∨E4∧F4∧G3**

Provenance
[Trio: “lineage”]

$$V(x) = R^p(x,y),S^p(x,y,z),T^p(x,z)$$

$$q = V(x),\ \ U(x,v),K(v,w)$$

Can compute now but inefficient

“Provenance Summary”

V^p:

x	E
a	E1∧F1∧G1∨E2∧F2∧G1
b	E3∧F3∧G3∨E4∧F4∧G3

➔

x	E
a	H1
b	H2

A very concise summary of the provenance

“Provenance Summary”

V^p:

x	E
a	$E1 \wedge F1 \wedge G1 \vee E2 \wedge F2 \wedge G1$
b	$E3 \wedge F3 \wedge G3 \vee E4 \wedge F4 \wedge G3$

➔

x	E
a	H1
b	H2

Now we *know* we can use the marginals

A very concise summary of the provenance

“Provenance Summary”

V^p:

x	E
a	**E1∧F1∧G1∨E2∧F2∧G1**
b	**E3∧F3∧G3∨E4∧F4∧G3**

➔

x	E
a	**H1**
b	**H2**

A very concise summary of the provenance

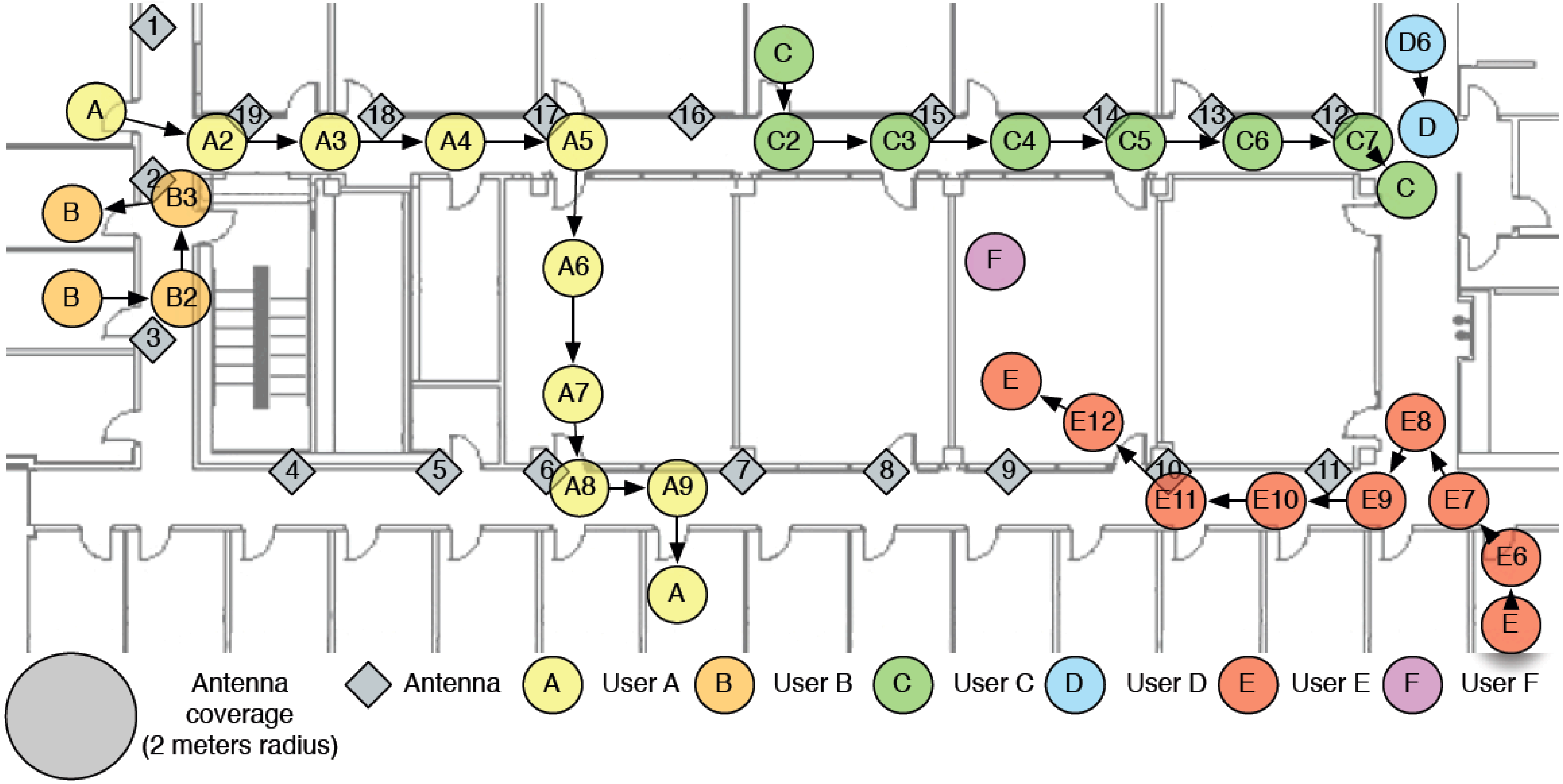
Now we *know* we can use the marginals

Status: deciding if a view V has independent tuples is Π^p_2 complete

Open: find a *minimal provenance summary*

Application 2: RFID Security

RFID Ecosystem at UW



[Welbourne'2007]

RFID Data

Base table

SIGHTINGS(tagID, antennaID, time)

EnteredRoom(personTagID, room, time)
CarriesObject(personTagID, objectTagID, time)
Meeting(personTagID1, personTagID2, time)
……….

Derived tables (views)

Application 2: RFID Security

Privacy w.Authorization Views

Alice's query

q(x) = EnteredRoom(x,"Rm552",t), Yesterday(t)

v1(x,l,t) = LocatedAt(x,l,t), LocatedAt("Alice",l,t)
v2(x,r) = EnteredRoom(x,r,t),EnteredRoom("Alice",r,t'),|t-t'|<10
v3(x,r,t) = Friend(x,"Alice"), EnteredRoom(x,r,t)
. . . . .

Authorization view

System answers the query if it can be rewritten in terms of views; else deny

[Rizvi'2004]

Privacy and Provenance

- Issue 1: the data *itself* is a materialized view. How can we make access control decisions based on how the data was derived ?

- Issue 2: the *authorization views* are probabilistic. How can we grant access with probability, say, 75% ?

Questions ?