

Spycy Privacy and Anonymity

Patrick Lincoln

Privacy

“The right to be let alone” [Warren & Brandeis 1890]

More precisely:

“The ability of an individual or organization to decide whether, when, and to whom information is released”

Example Motivation



- ◆ Airline passenger databases
 - Anti-terrorism, intelligence, law enforcement
- ◆ Financial transaction records
 - Fraud detection
- ◆ Medical research databases
 - Research queries for interactions
- ◆ Computer network monitoring
 - Intrusion detection



Need to protect personal and organizational privacy while enabling security capabilities

Privacy and Anonymity in Education

To: Instructors and TA's

Re: Posting of grades

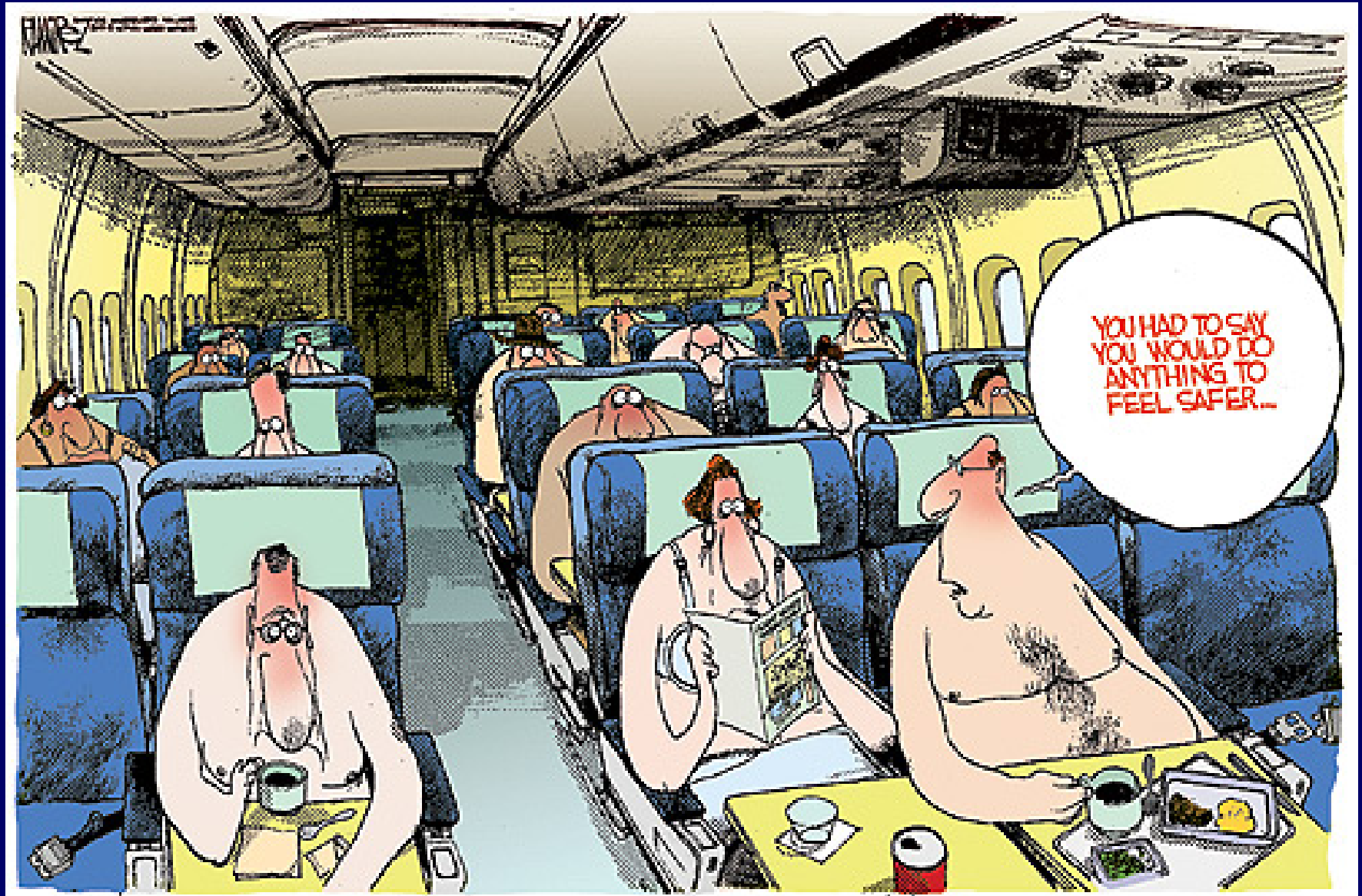
If you want to post grades on the web you should use the grade posting program or use a program that posts only the student's grade and the mean, max, etc. using the student's ID number. You cannot use Social Security Numbers. You should not post grades in a list by student ID because this does invade the students' privacy (e.g. if two or three grades are very low it may become known who the students are.) Also, from my own experience it is bad policy to post grades in such a fashion because students will use it against you in arguing for a grade change (i.e. students on the borderline will know it and argue accordingly.)

If people can figure out a student's grade from what is posted we are in **LEGAL TROUBLE.**

Organizational Privacy and Anonymity

- ◆ Corporate and agency reporting
 - To regulators, shareholders, law enforcement
- ◆ Whistleblowing
 - Enable limited anonymous disclosure
- ◆ Cashlike electronic transactions
 - Business intelligence through credit card or bank purchase records
- ◆ Battlefield communications
 - Make it difficult for attacker to determine who is sending messages
- ◆ Devices
 - Unmanned sensors, processors, and actuators

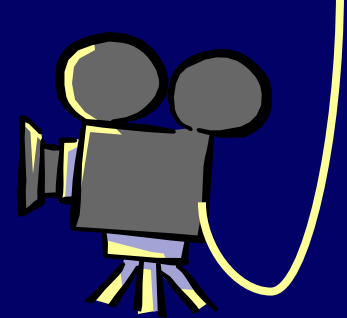
How Exposed Must We Be To Achieve Better Security?



Ubiquitous Monitoring and Recording is Inevitable



- ◆ Echelon, Carnivore, DCS1000, TIPS, DARPA-TIA, Zyuumin Kihon Daityou, Palladium (MS), and TPCA (Intel/AMD)
- ◆ Massive databases
 - Terabytes, PetaBytes and beyond
- ◆ Combinations of databases
 - Car rental, web use, immigration, health records ...
- ◆ Correlation of disparate events
 - Fingerprint, traffic stop, license plate of serial sniper



Useful capabilities for business optimization, law enforcement, and public policy, but pose unprecedented threats to personal and organizational privacy

Privacy Interests



◆ Personal privacy

- Medical, financial, other detailed information
- Implied by U.S. Constitution

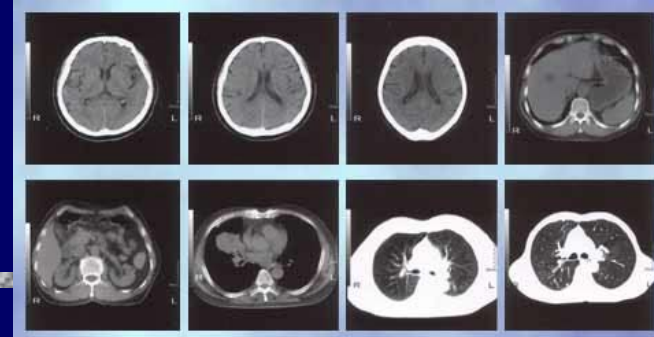
Pursuit of individual interests away from the prying eyes of commercial or societal oversight

◆ Organizational privacy

- Trade secrets
- Legal but secret business practices
- Competitive advantage

Preservation of competitive advantage

Specific Examples



- ◆ CAT scans now accurate enough to reconstruct recognizable faces
 - Modern CT scan data inherently contains PII: Personally Identifiable Information
- ◆ “Anonymized” patient records include zip code, birthdate, sex
 - Combined with driver records, one can “de-anonymize” putting names back onto 90% of records
- ◆ U.S. SEC filings require social security numbers
 - Then filings made available on web
 - ...including Bill Gates’ social security number

Security Needs

- ◆ Intelligence and law enforcement needs to mine databases for evidence of terrorist and criminal activities
- ◆ Medical researchers need to search databases for correlations indicating predictive factors for adverse drug reactions
- ◆ Financial institutions need to detect and respond to fraudulent transactions
- ◆ Internet service providers need to identify and respond to denial-of-service attacks

Major Current Approaches

- ◆ Trust the investigators
 - Government listens to all calls
 - Visa knows all your transactions
 - HMO knows your entire health history
- ◆ Trust a third party
 - Key escrow (Clipper)
 - Centralized anonymous remailer
 - Patient-record privacy monitor
- ◆ Trust multiple parties not to collude
 - Institutional or individual collusion
- ◆ Security by obscurity
 - Hope what you care about is never noticed

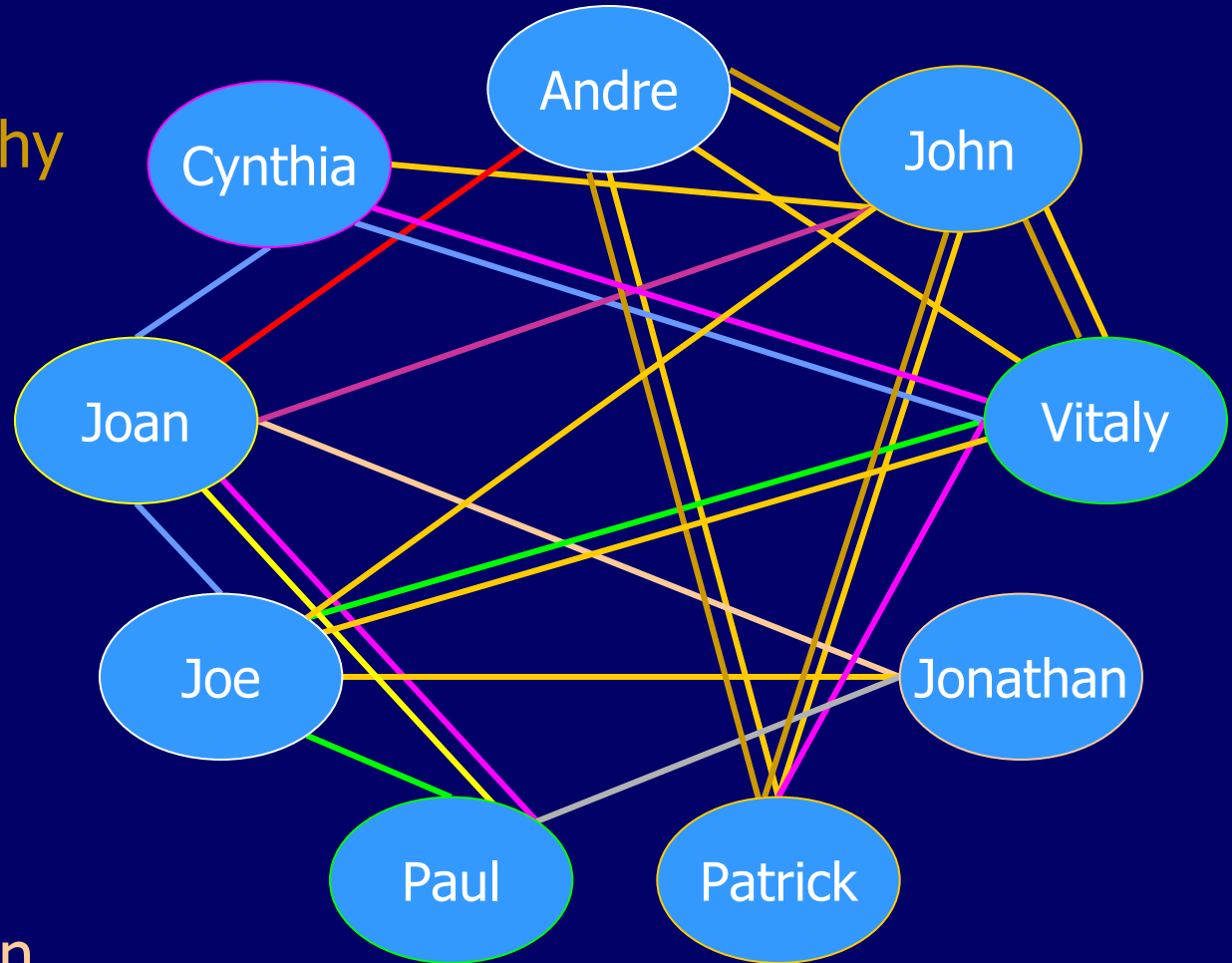
Privacy and Anonymity in Diffuse Computing

- ◆ Authorization mechanisms
- ◆ Information retrieval
- ◆ A good use of frameworks and approaches developed for diffuse computing

People won't buy into diffuse computing infrastructures without some guarantees of limited privacy and anonymity

Spyce Interaction Graph

- Protocol Analysis
- Formal Methods
for Cryptography
- Anonymity
- Privacy
- Algorithmic
Mech Design
- Authorization
- Decision Theory
- BGP
- Digital Rights
- Network Congestion



Privacy and Anonymity Papers

- ◆ Richard E. Newman, Ira S. Moskowitz, Paul Syverson, Andrei Serjantov. *Metrics for Traffic Analysis Prevention*
In [2003 Workshop on Privacy Enhancing Technologies \(PETS\)](#)
- ◆ Alessandro Acquisti, Roger Dingledine, and Paul Syverson. *On the Economics of Anonymity*
In [Financial Cryptography 2003](#).
- ◆ Andrei Serjantov, Roger Dingledine, and Paul Syverson. *From a Trickle to a Flood: Active Attacks on Several Mix Types*. In [Information Hiding](#), Oct 2002.
- ◆ Roger Dingledine, Nick Mathewson, and Paul Syverson. *Reputation in Privacy Enhancing Technologies*.
In [Computers, Freedom, and Privacy](#), Apr 2002.

Privacy and Anonymity Papers

- ◆ Roger Dingledine and Paul Syverson. *Reliable MIX Cascade Networks through Reputation*. In [Financial Cryptography 2002](#).
- ◆ Cynthia Dwork and Moni Naor, working title *SPAM reduction*
- ◆ P. Golle, M. Jakobsson, and Paul Syverson *Universal re-encryption*
- ◆ Joan Feigenbaum, Michael J. Freedman, Tomas Sander, and Adam Shostack. *Privacy Engineering for Digital Rights Management*, in Proceedings of the [2001 ACM Workshop on Security and Privacy in Digital Rights Management](#). vol. 2320, Lecture Notes in Computer Science, Springer, Berlin, 2002, pages 76-105.

Privacy and Anonymity Papers

- ◆ Jarecki, S., Lincoln, P., and Shmatikov, V. *Negotiated Privacy* (extended abstract). *LNCS Proc. International Symposium on Software Security (ISSS)*, 2002.
- ◆ Hughes, D., and Shmatikov, V. *Information Hiding, Anonymity and Privacy: A Modular Approach*. 19th Annual Conference on Mathematical Foundations of Programming Semantics (MFPS XIX), revised version to appear in *Journal of Computer Security*, 2003.
- ◆ Shmatikov, V., and Hughes, D. *Defining Anonymity and Privacy*. In *Workshop on Issues in the Theory of Security (WITS '02)*, 2002.
- ◆ Shmatikov, V. *Probabilistic Analysis of Anonymity*. In *Proc. 15th IEEE Computer Security Foundations Workshop (CSFW-15)*, pages 119-128, 2002.

Privacy and Anonymity Papers

- ◆ Joseph Halpern and Kevin O'Neill. *Anonymity and Information Hiding in Multiagent Systems*
To appear, Proceedings of the 16th IEEE Computer Security Foundations Workshop, 2003
- ◆ Joseph Halpern and Kevin O'Neill. *Secrecy in Multiagent Systems* Proceedings of the 15th IEEE Computer Security Foundations Workshop, 2002, pp. 32-46

What Does It Mean to Be Anonymous?

A knowledge-based perspective

Dominic Hughes

Stanford University

Vitaly Shmatikov

SRI International

What Is "Anonymity?"



*FBI intercepted three emails,
and learned that ...*

- ◆ Two of the emails came from the same account
- ◆ Emails are *not* in English
- ◆ The recipients are Bob386@hotmail.com, Dick Tracy and Osama Bin Laden, but it's not known who received which email
- ◆ Emails were routed via Anonymizer.com

Wrong question: has "anonymity" been violated?

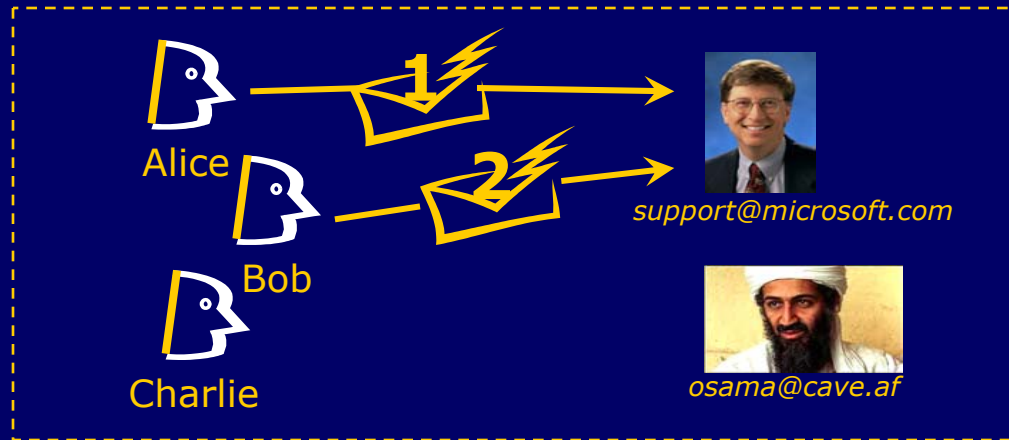
Right question: what does FBI actually **know**?

Anonymity and Knowledge

- ◆ Anonymity deals with hiding information
 - Agent's identity is hidden
 - Relationship between agents is hidden
 - Agent cannot be identified within a set of suspects
- ◆ Natural way to express anonymity is to state what the attacker should not know
 - Typically requires logic of knowledge
 - Not supported by conventional formalisms for security (process calculi, I/O automata, ...)
- ◆ To determine whether anonymity holds, need some representation of knowledge

2-Anonymity

What actually happened



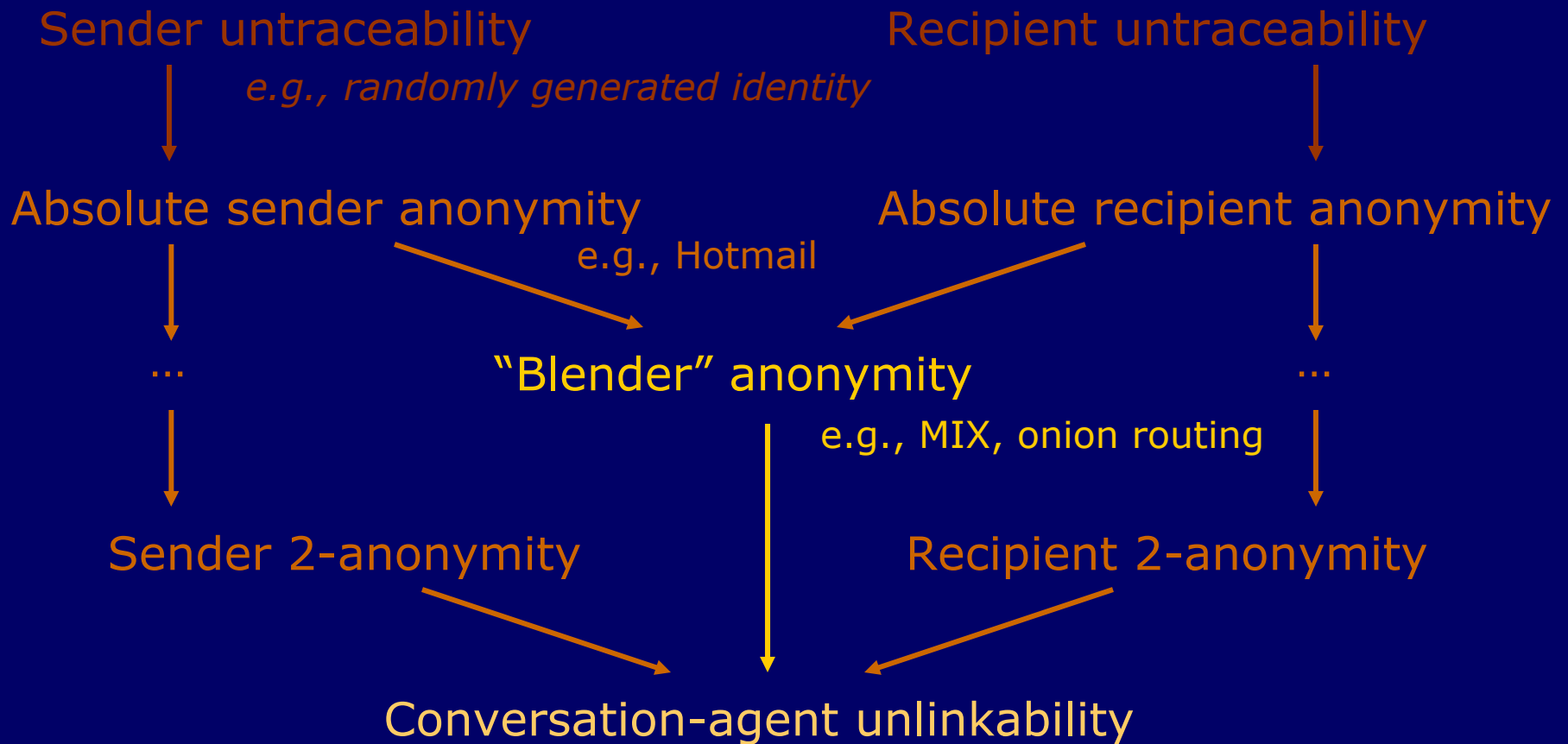
What attacker knows

Sender suspects(1) = Alice or Charlie

Sender suspects(2) = Bob or Charlie

2-anonymity for senders:
2 plausible senders for each message

Hierarchy of Common Properties



General Framework

◆ Arbitrary information hiding properties

- Entire protocol sessions instead of single messages
- Arbitrary relations between agents (modeled as types)
- Higher-order properties

◆ Subtle forms of partial knowledge

- “This message exchange is an instance of SSL, but client’s and server’s identities are hidden”
- “Messages between Alice and Bob are in English or Dutch, but not in French”
- “This email was sent by someone in China”

◆ Automatic translation into verification conditions for your favorite process algebra

Anonymity and Information Hiding in Multiagent Systems: A Knowledge-based Approach

Joseph Halpern

&

Kevin O'Neill

Motivation

Anonymity is important to people in a variety of situations:

- Browsing the web
- Sharing files with other Internet users
- Sending messages
- Real-life situations, such as:
 - making large donations
 - whistle-blowing

Often people will be reluctant to engage in some behavior unless they can receive guarantees that their anonymity will be protected

Achieving Anonymity

There are protocols and systems that guarantee anonymity in restricted situations

- DC-nets based on the “dining cryptographers” protocol
- Anonymizer
- Crowds
- Herbivore

These systems are all quite different, and offer different kinds of anonymity guarantees

We want to be able to compare the guarantees provided by these systems using a total framework

A Formal Framework for Anonymity

Ideally, a formal framework for anonymity should:

- let us define different kinds of anonymity guarantees in a precise, intuitive way.
- model real-world systems.
- provide a way to verify formally that a given system provides a desired anonymity guarantee.

An Example

◆ An anonymous message-passing system:

- Agents send messages (i.e., email) to other agents in the system
- When sending a message, agents may sometimes want to ensure that:
 - the message is sent anonymously
 - the message is received anonymously
 - the message is both sent and received anonymously

Defining Anonymity

- ◆ We define anonymity as an instance of “information hiding”, where we ask:
 - what information needs to be hidden?
 - who does it need to be hidden from?
 - how well does it need to be hidden?
- ◆ Anonymity is closely related to the *knowledge* of the agents interacting with the system
 - Our definitions of anonymity use knowledge in a formal way.
- ◆ We relate anonymity to our earlier work on secrecy and noninterference.

Representing Multiagent Systems

Our model lets us represent all possible behaviors of the system as well as the state of the agents who use the system.

- n agents, each in some local state s_i at a given point in time
- The whole system in some global state (s_1, \dots, s_n, s_e)
- A run r is a function from time to global states
- A point of the system is a pair (r, m) – a particular execution sequence at a particular point in time
- A system R is a set of runs

Local States and Knowledge

We write $r_i(m)=s_i$ if i has local state s_i at point (r,m)

At the point (r,m) , agent i considers possible all the points (r',m') such that $r_i(m)=r_i(m')$.

If a fact φ is true at all points that i considers possible, we say that “ i knows the fact φ ”.

- Denoted “ $K_i\varphi$ ”

If a fact φ is true at some point that i considers possible, we say that “ i considers φ possible”.

- Denoted “ $P_i\varphi$ ”
- $P_i\varphi$ iff $\neg K_i\neg\varphi$

Defining Anonymity

We define anonymity in terms of actions and the agents who perform them.

- Let $\delta(i,a)$ represent the fact that i has performed action a

Action a , performed by agent i , is minimally anonymous with respect to agent j if R if the formula " $\neg K_j[\delta(i,a)]$ " is always true.

- If an observer j knows that i sent a message, then i doesn't have any anonymity, at least with respect to j .

Minimal anonymity is a very weak condition:

- Minimal anonymity holds as long as j is not 100% sure that i performed action a .

A Stronger Version of Anonymity

An agent i might want to ensure that observers think it possible that many agents, perhaps all the agents in some “anonymizing set” A , could have performed the anonymous action.

Action a , performed by agent i , is *anonymous up to* A with respect to an agent j in R if the following formula is always true:

$$\delta(i,a) \rightarrow \bigwedge_{i' \in A} P_j[\delta(I',a)]$$

Anonymity up to A is clearly more restrictive than minimal anonymity

- *Total anonymity* is an even stronger condition

Probabilistic Definitions of Anonymity

Problems with “possibilistic” guarantees:

- Suppose an observer o thinks that any of 101 agents in a set A could have performed an action a .
- What if o has a probability of 0.99 that i performed a , and a probability of 0.0001 that any of the other 100 agents performed a ?
- Here anonymity up to A doesn't provide too much comfort to i ...

We describe how probability can be added to the multiagent systems framework, and we give examples of stronger guarantees of anonymity that use probability.

- Previous formalizations have not dealt with probability.

Conditional Anonymity

Consider an anonymous message-passing system:

- Even if the system makes it impossible to trace messages to my identity, the content of my messages may leak information.
- Observers will have prior probabilities on what various agents might do in a given system.
 - Kevin is unlikely to make a multimillion-dollar donation!
- This makes it fundamentally difficult to give strong probabilistic anonymity guarantees for a real-world system.
- We give a new definition of conditional anonymity.
 - $\Pr_j(\delta(i,a) \mid \text{what } j \text{ observed})$
= $\Pr_j(\delta(i,a) \mid \text{what } j \text{ is entitled to know})$
 - It's related to our (much stronger) definition of secrecy.

Related Work

Others have formalized anonymity:

- using epistemic logics [Syverson and Stubblebine, 1999];
- using CSP [Schneider and Sidiropoulos, 1996];
- using functions views [Hughes and Shmatikov, 2002].
 - Actually, our work was inspired by Vitaly's talk at the fall SPYCE meeting!
- Many of our definitions have been given before, but we show that these definitions can all be captured in one framework.

Analyses of real-world anonymity systems:

- Shmatikov [2002] analyzes the Crowds system using a probabilistic model checker.

One Application: CSP and Anonymity

Schneider and Sidiropoulos define anonymity in terms of CSP:

- Let A be a set of “anonymous events”
- A process P is strongly anonymous on A if $f_A^{-1}(f_A(P))=P$ (where f_A is a particular renaming function).
- This definition is not very intuitive, but can be used to verify real-world protocols using model checkers for CSP.

We show that this definition is a special case of our definitions

- A process P can be associated with a set of runs R_P , and the set A with a particular action a and set of agents I_A .
- Theorem: P is strongly anonymous on A if and only if actions in A are anonymous up to I_A .

For the Future

Verification is an eventual goal:

- Using the knowledge-based framework directly [van der Meyden, 1998],
- Or indirectly, using a related framework such as CSP or the pi-calculus.

We would like to say more about the relationship between the knowledge-based system framework and the process algebra framework:

- We want a canonical translation from processes to multiagent systems so that information-hiding properties make sense for systems specified using process algebras.

Negotiated Privacy

Technical Provision of Security With Privacy

Stanislaw Jarecki, Patrick Lincoln, Vitaly Shmatikov

Stanford University and
Computer Science Laboratory
SRI International



Aspects of Anonymity and Privacy

- ◆ Ability for someone to release data open to some kinds of use, but not to other uses
 - Without having to trust a third party to hold data, restrict to appropriate use, and monitor users

- ◆ Main idea:
 - Encrypt database entries in a special way
 - Up-front negotiated set of supported queries
 - All other queries very difficult
 - Violating privacy in our system implies ability to break existing cryptographic systems

Key Question: Who Has The Keys?

- ◆ Who has access to keys that protect YOUR private data?
 - Analysts asking questions
 - Law enforcement
 - Judicial authorities
 - Database owner
 - Your doctor's outsourced MIS provider
 - That device you left on the battlefield
 - System administrator
 - System administrator's friend
- ◆ All of the above? Not good.
- ◆ Possible answer: None of the above.





Monitor

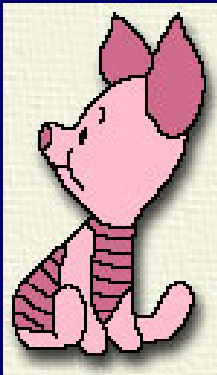
Verifiable Anonymous Encryption



Data

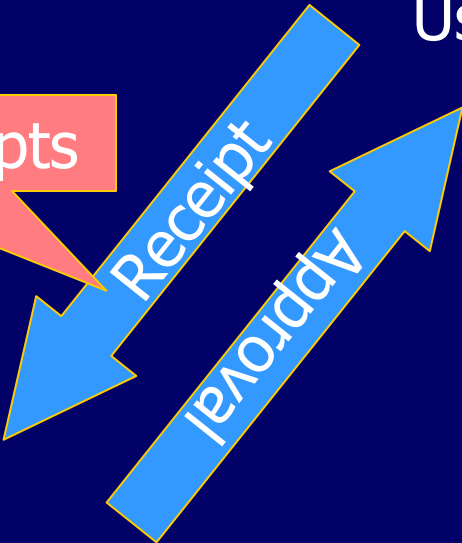


Receipt



User

Unlinkable Receipts



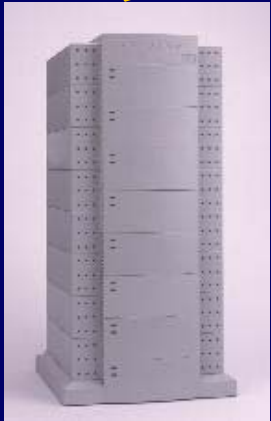
Receipt

Approval

Secure Tagging



Service



Storage

Related Work

- ◆ Search on encrypted databases [Song, Wagner, Perrig '00]
 - Doug's talk earlier today
- ◆ Untraceable electronic cash [Chaum, Fiat, Naor '88]
 - Anonymously spend coin once, but reveal name if spend coin twice
 - Prevent frequent flyer from using new coin per flight
- ◆ Group signature schemes [Ateniese, et al '00]
- ◆ Anonymous credentials [Camenisch, Lysyanskaya '01]
- ◆ Private Information Retrieval [Chor, Goldreich, Kushilevits, Sudan '98]
 - Protect privacy of analysts: keep queries private
 - Complementary to what we are concerned with here
- ◆ Symmetric Private Information Retrieval [Malkin, ...]

What Queries Can Be Supported in Negotiated Privacy?

◆ Threshold-type queries

- Has anyone flown to airport within 150 miles of Kabul, Afghanistan 3 times this year?
- Cryptographic protection below threshold

◆ Equality of subfields

◆ Linear arithmetic functions on subfields

◆ Can define subset of SQL for access

Note on Limitations

- ◆ Queries must be agreed up front
 - Standing queries vs Real-time queries
 - Before data is entered into database
- ◆ Service providers must enforce rigidly
 - If airline provides service without proper identification, no virtual mechanisms will enable information capture
- ◆ Presupposes Public Key Infrastructure (PKI)
- ◆ Legal and ethical issues
- ◆ International standards and cooperation is needed

Goals of Negotiated Privacy

- ◆ Enable collection of database entries that are
 - Reliable
 - Accurately describe user's activities
 - Despite misbehavior of user
 - Secure
 - Leak no information unless disclosure condition is satisfied, or analyst can break Decision Diffie-Hellman
 - Despite misbehavior by analyst, or collusion between analyst and service provider
- ◆ Claim: our described protocol achieves these goals, subject to certain limitations

Benefits of Negotiated Privacy

- ◆ Enables acceptable queries
 - Provision of partitioned answers to certain queries
- ◆ Disables unacceptable queries
 - Intractable to extract protected information
- ◆ Strongly protects privacy
 - Even stealing whole set of databases, can't mine for information beyond negotiated query set
- ◆ Encourage compliance
 - Less concerns about corruptible analysts
 - Less concerns about future policy changes allowing other uses of data

Additional Challenges

◆ Management

- Of keys, of infrastructure, of process

◆ Agreement on acceptable queries

- Law enforcement: “Need to see everything”
- Privacy advocates: just say “Nothing”

◆ Agreement on what happens on match

- Release only inherently opened information
- Open that one complete record
- Instantly open all that individuals records
- Go to a third party (go to judge for warrant)

How to Achieve High Assurance of Security and Privacy

Need to Engineer a Combination of Many Technical Approaches

- Restricted access
- Protected execution
- Selective revelation
- Data labeling
- Intrusion detection
- Tamper-evident audit trail
- Analysis of trusted platform
- Private information retrieval
- Negotiated privacy

Discussion here just one aspect of larger problem

Universal Re-encryption: For Mix-Nets and Other Applications



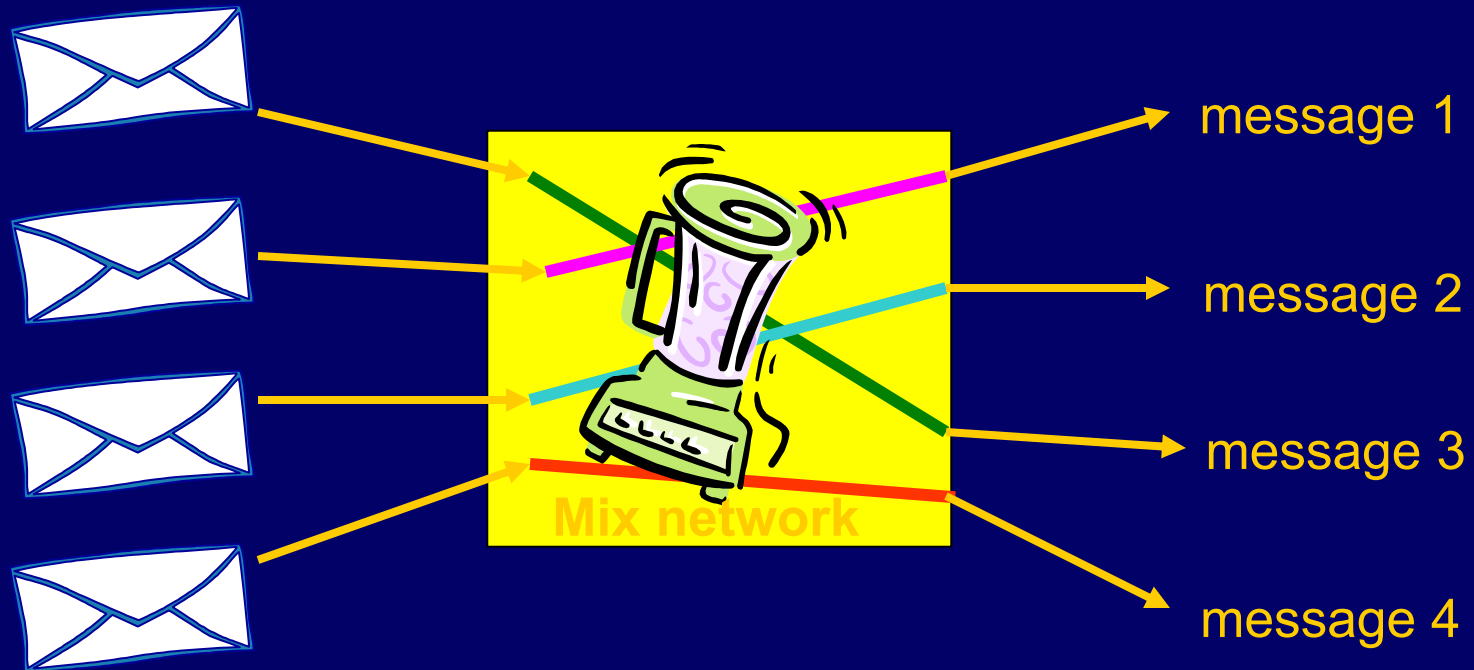
Philippe Golle
Stanford

Markus Jakobsson Ari Juels
RSA Labs

Paul Syverson
NRL

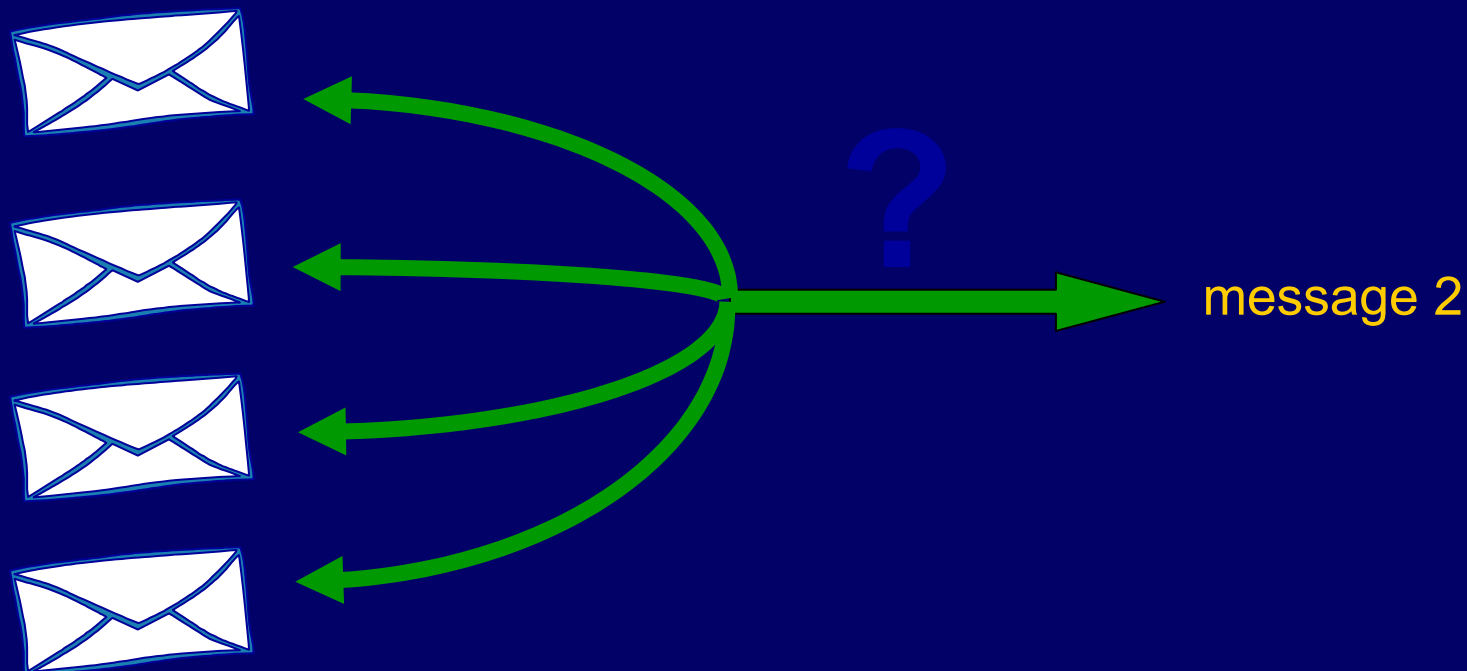


What does a mix network do?



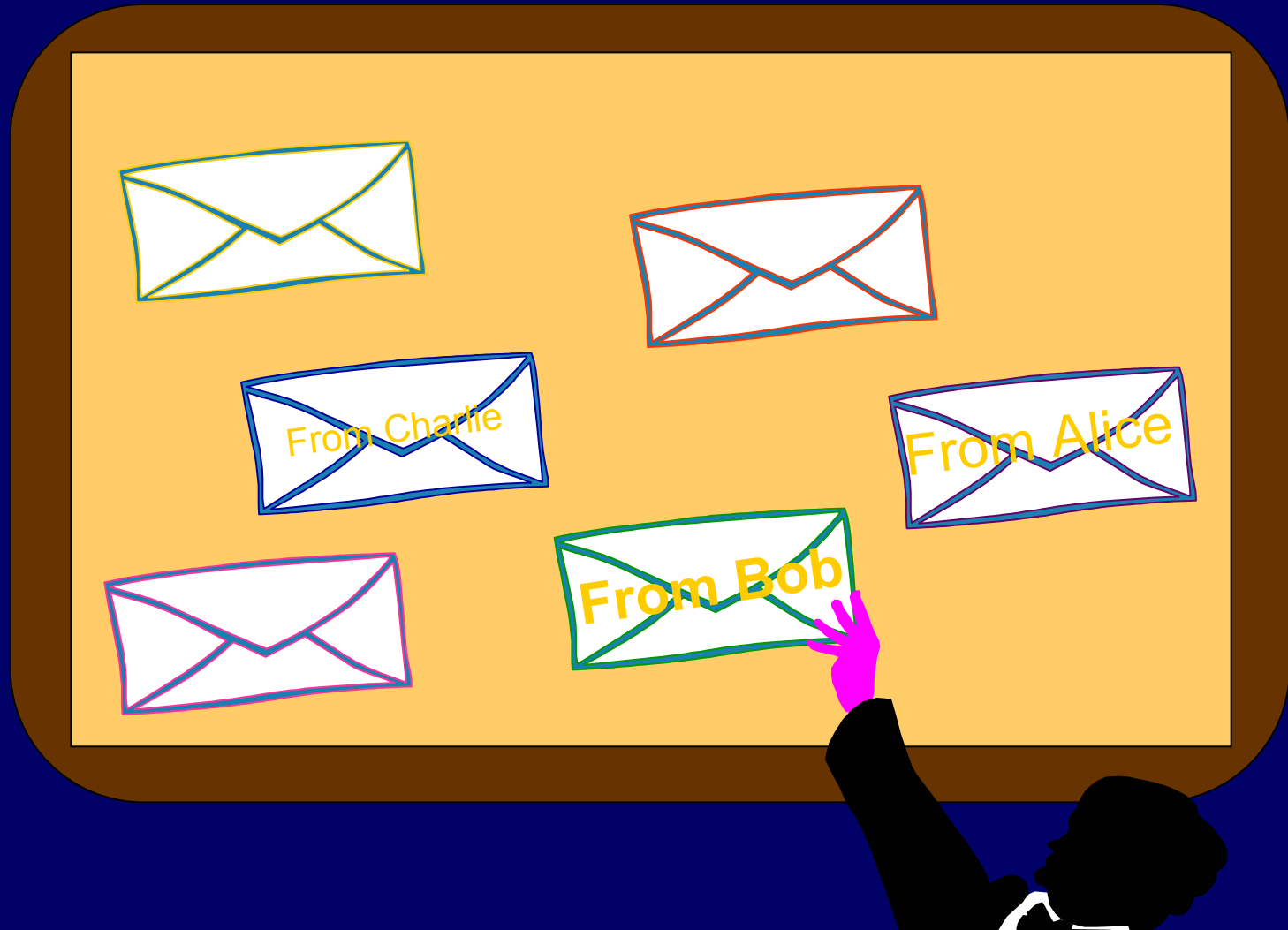
Randomly permutes and decrypts inputs

What does a mix network do?



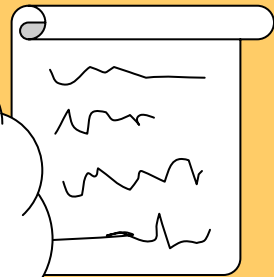
Key property: Adversary can't tell which ciphertext corresponds to a given message

Example application: Anonymizing bulletin board or e-mail

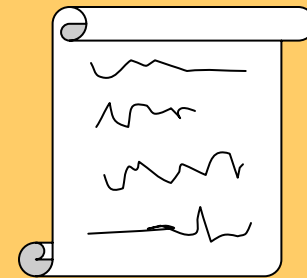


Example application: Anonymizing bulletin board or e-mail

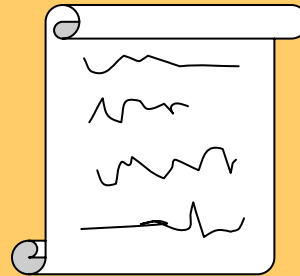
Is it Bob, Charlie,
self-love, or other?



"Nobody
loves Bob"

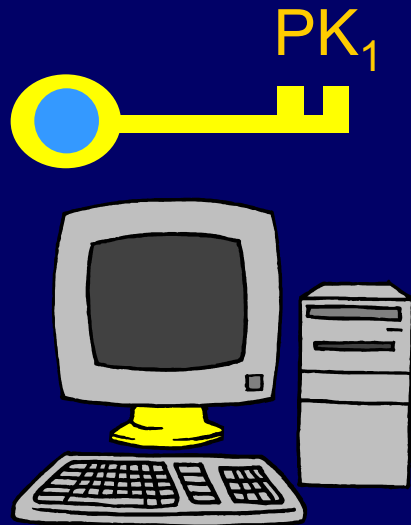


"I love
Alice"

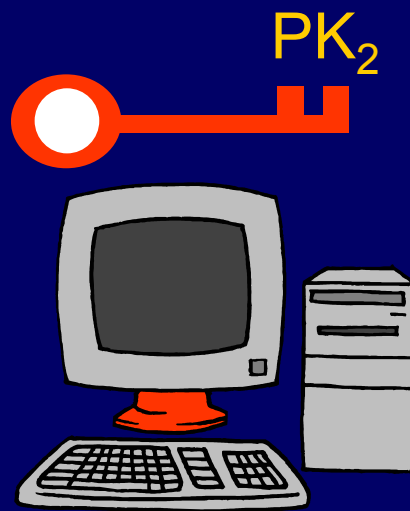


"I
love
Charlie"

Basic Mix (Chaum '81)



Server 1

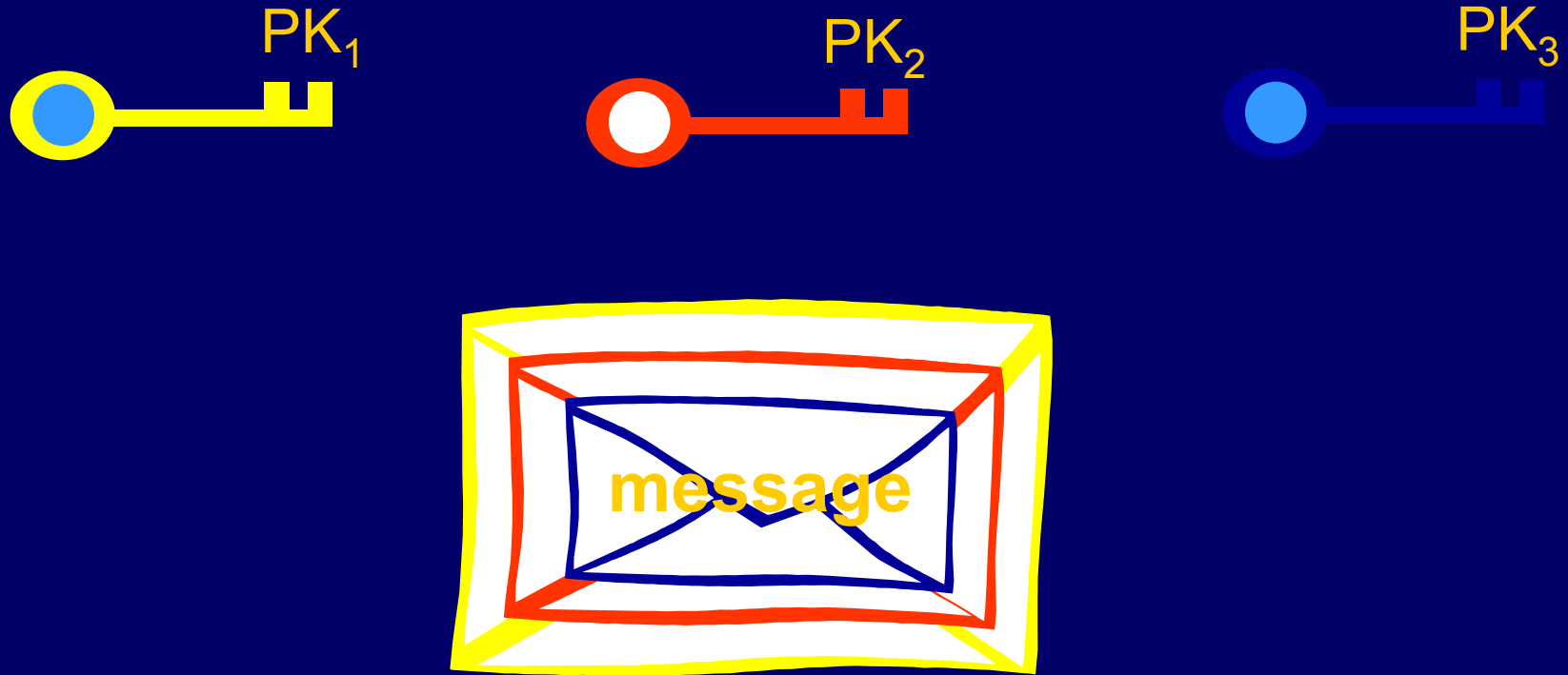


Server 2



Server 3

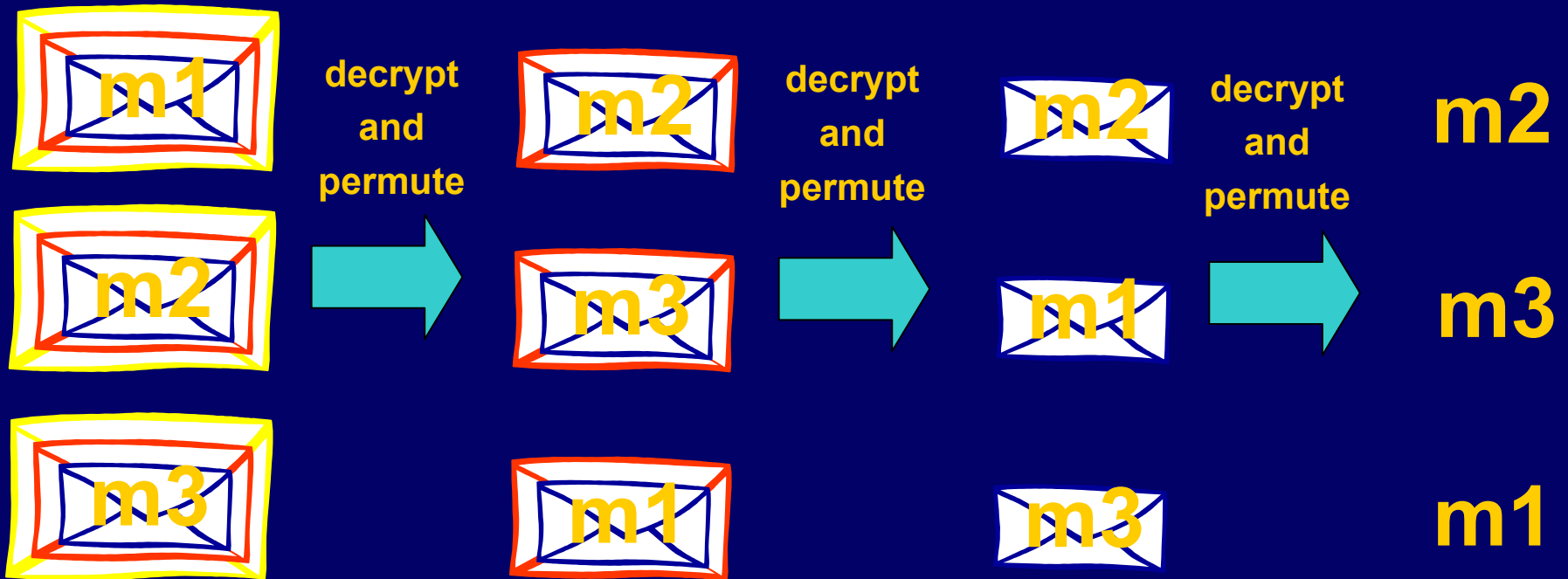
Encryption of Message



$$\text{Ciphertext} = E_{PK_1}[E_{PK_2}[E_{PK_3}[\text{message}]]]$$

Basic Chaumian Mix

Observe: As long as one server is honest, privacy is preserved



Basic Re-encryption Mixnet



- Inputs are ciphertexts
- Outputs are a re-encryption of the inputs.
- El Gamal public key encryption:
 - Anyone can encrypt with the public key e
 - Those who know the secret key d can also decrypt
 - Malleable: can produce $E_2(m)$ from $E_1(m)$ without knowing d
 - Verifiable
 - Multiplicative homomorphism: given $E(m)$ and $E(m')$ I can produce $E(mm')$

Universal Re-encryption Mixnet



- Inputs are ciphertexts
- Outputs are a re-encryption of the inputs.
- El Gamal public key encryption:
 - Anyone can encrypt **without** the public key e
 - Those who know the secret key d can also decrypt
 - Messages encrypted with different keys are indistinguishable

Universal Mixnets

- Any node can mix any message
- Nodes can be dynamic
- Network topology not significance
- No PKI and less trust of each node
- No robustness/reliability issues with node failure
- No overhead or threats from replay (universal semantic security)
- Can have free route re-encryption mixnets
 - With large anonymity sets

Conclusions

- Universal Re-encryption: New primitive
- Applications
 - Reduced trust in mixes
 - Less complex mixnets (no PKI)
 - Better anonymous connections
 - Privacy preserving RFID tags
- Open
 - Properties: Universal Semantic Security, Existential Construction Resistance
 - More Applications

Fighting Spam May Be Easier Than You Think

Cynthia Dwork
Microsoft Research SVC

Why?

◆ Huge problem

- Industry: costs in worker attention, infrastructure
- Individuals: increased ISP fees
- Hotmail: huge storage costs, 65-85%
- FTC: fraud, confidence crimes
- Ruining e-mail, devaluing the Internet

Computational Approach [DN'92]

◆ If I don't know you:

- Prove you spent ten seconds CPU time,
- just for me, and just for this message

◆ User Experience:

- Automatically and in the background
- Checking proof extremely easy

◆ All unsolicited mail treated equally

Principal Techniques

◆ Filtering

- Everyone: text-based
- Brightmail: decoys; rules updates
- Microsoft Research: (seeded) trainable filters
- SpamCop, Osirusoft, etc: IP addresses, proxies, ...

◆ Make Sender Pay

- Computation [Dwork-Naor'92; Back'97]
- Human Attention [Naor'96, DEC patent]
- Money [eg, Gates'96, Hayes, McCurley]

Summary

- ◆ Discussed computational approach, Turing tests, economics, cycle-stealing
- ◆ Briefly mentioned two architectures
- ◆ Examined difficulties of constructing memory-bound pricing functions and proposed a new one designed to avoid these difficulties (no proofs yet)

Future Work and Open Questions

- ◆ Implement and test Outlook, Pine plug-ins (at Stanford); add signatures
- ◆ Further study of DGN algorithm
- ◆ Distribution Lists
- ◆ Good MB functions with short descriptions (will subset product work)?

The Future of Privacy and Anonymity

- ◆ Stronger guarantees for negotiated privacy
 - Verifiable anonymous cryptography
 - Secure tagging
 - Unlinkable receipts
- ◆ Cryptographic signal processing
- ◆ Reputation-based incentives and trust
- ◆ Formal analysis of real-world protocols
 - Privacy
 - Anonymity
- ◆ Anonymity without keys

We will make a name for ourselves in anonymity