CS 194-1 (CS 161)
Computer Security

Midterm 3
Review

Part 1

Database Inference Control

Census bureau problem

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

Complex Queries Only

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 $\wedge$ DRUGS=1
- List NAME where (SEX=M $\wedge$ DRUGS=1) $\vee$ (SEX$\neq$M $\wedge$ SEX $\neq$ F) $\vee$ (DORM=AYRES)

Approaches that don't work

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

More generally any linear combination

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

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

Watermarking

Watermarking

We want to protect data:

- Video, sound, music (Digimarc, Intertrust, etc)
- Programs (Collberg, Thomborson)
- Statistical data

Examples of "traditional" protection methods:

- False entries in biographical dictionaries
- Copyright notices
- Licensing agreements
- Secure coprocessors

Watermarking

Watermarking:

- include low level bit data that marks information
- Either on a per-copy basis or a per-provider basis

Example: temperature database

- slightly adjust temps to mark uniquely

- Store copies of info released
  - If reused, prove using similarities

- But what if adversary changes low-level info?

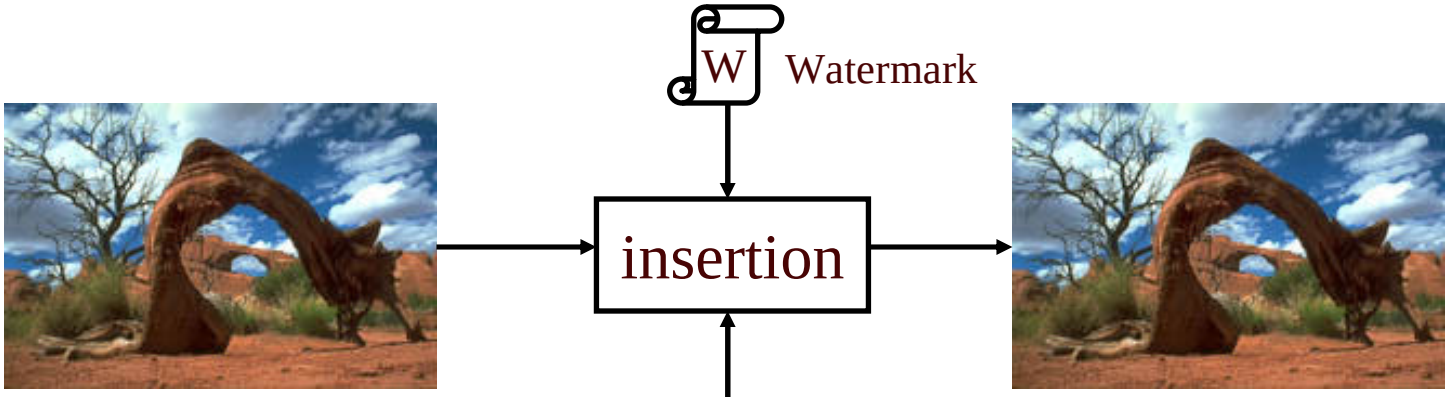
Motivation

- Intellectual property is important for the Internet
- IP (images) are valuable
  - Costly to create high quality images
  - Users are attracted by good design
- Binary data is trivial to copy
- The web is a headache for copyright protection
- Many methods for free data exchange

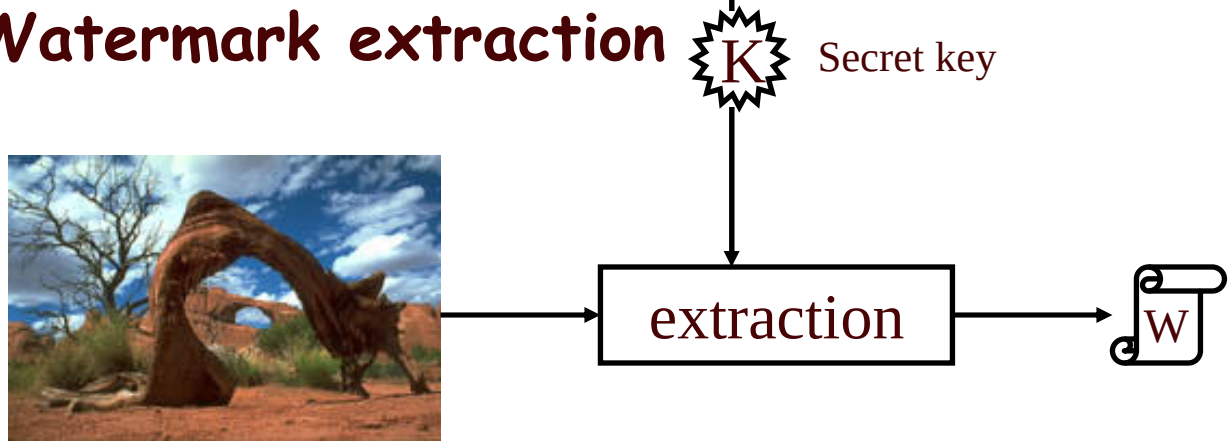
- Watermarking is seen as the white knight of copyright protection

The watermarking process (private wm)

- **Embed a watermark**



- **Watermark extraction**



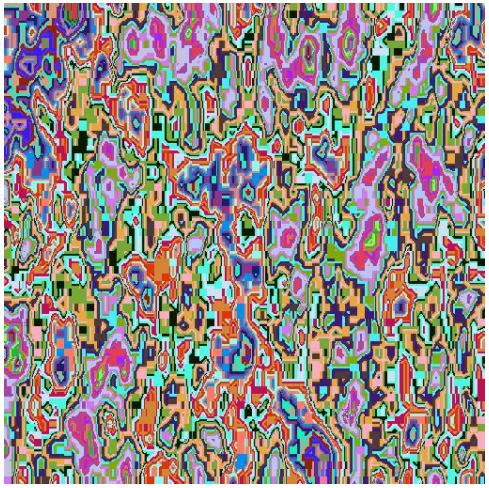
Requirements of invisible watermarks

- Robust against tampering (un- & intentional)
  - Various image transformations (RST)
  - Image compression
  - Color requantization
  - Non-linear transformations (print and scan)
- Non-perceptible, hard to detect
- Easy to use, exportable, etc.

- How can watermarking be possible?
  - The visual system has very strong "error correction"
  - An images contains a lot of redundancies
  - Small changes are undetected
  - People are used to low image quality (TV, newspaper images)

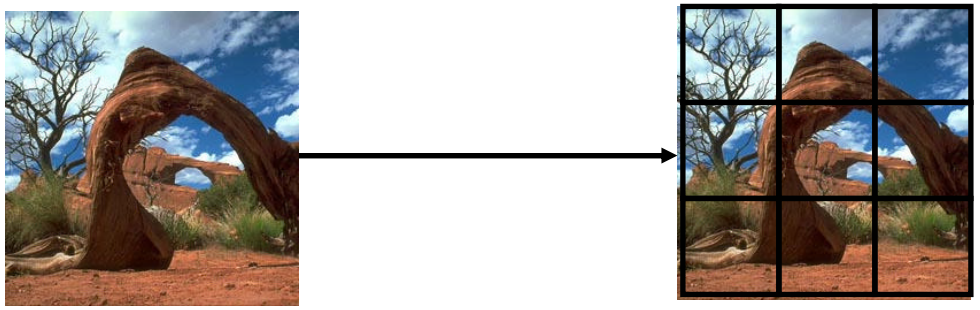
Example: The NEC watermark

- There is no perceptible difference between the original and watermarked image
- But the difference image looks interesting
- The watermark is present everywhere!



Example: Robustness to cropping

- Let's use the Fourier transform to construct a scheme which is robust against cropping
- Tile the image with small blocks of watermarks
  - For each block, we compute the Fourier transform
  - The watermark is embedded in the Fourier domain (each block)
  - Then we compute the inverse transform



Each block is handled individually

Problems of Watermarking

- Copyright protection is big business - many attackers
- Internet spans continents and countries seamlessly
- Digital information is easy to copy
- Hackers are knowledgeable, creative, have lots of time, and are numerous
- Many attack opportunities
  - Few inventors, many attackers
  - Inventors despair after 3 years
- Human factors:
  - The default user does not understand watermarking
  - Human vision system is very robust to noise in images
  - Used to low quality in images (TV, strong JPEG compression)

Electronic Commerce

Information goods

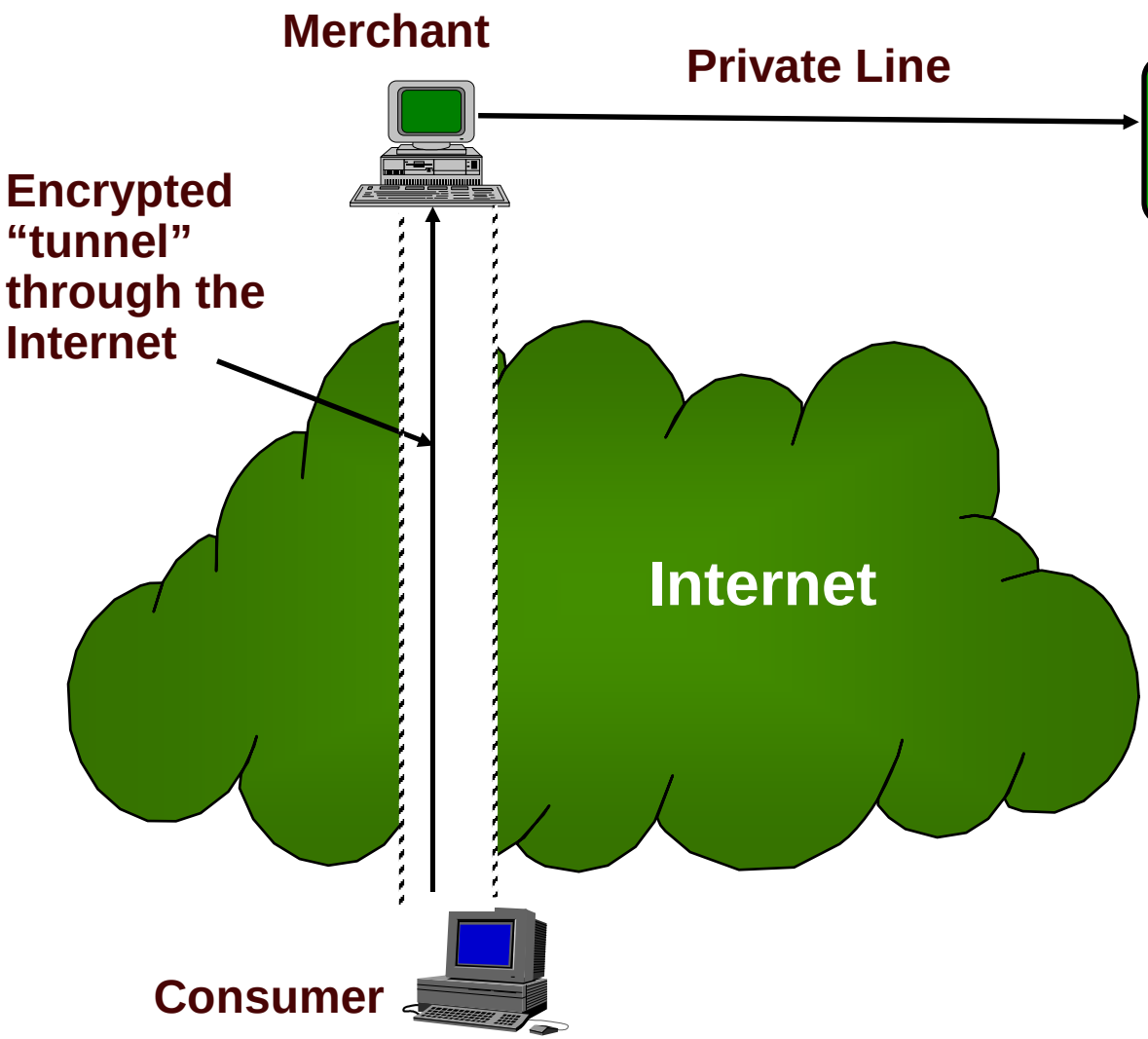
- Consider the purchase of an information good or service:
  - Library information
  - Search services
  - Software
  - Video clips
- These transactions may be large value or microtransactions
- In either case, atomicity is crucial

What is atomicity?

- I won't try to give a formal definition
- 3 types of atomicity:

- Money atomicity
  - All money transfers complete with non-ambiguous results
  - Money is neither destroyed nor created
- Goods atomicity
  - One receives goods if and only if one pays
  - Example: Cash On Delivery parcels
- Certified delivery
  - Both buyer and seller can prove the delivered content
  - If you get bogus goods, you can prove it

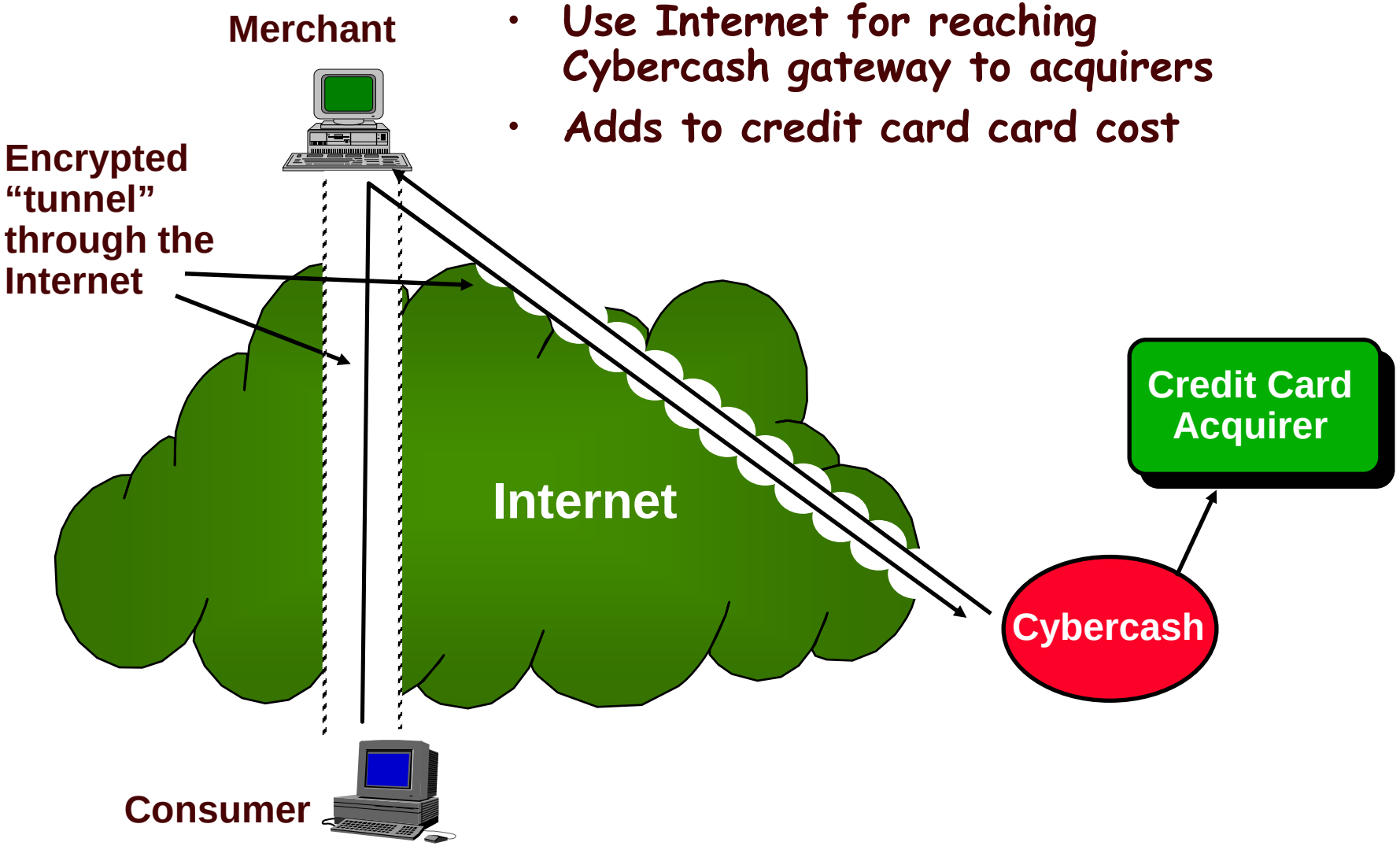
SSL model



- Consumer sends card # direct to merchant
- Similar to today's phone order
- Must trust merchant with card info
- Weak atomicity
- High transaction costs

Third party intermediary (Verisign)

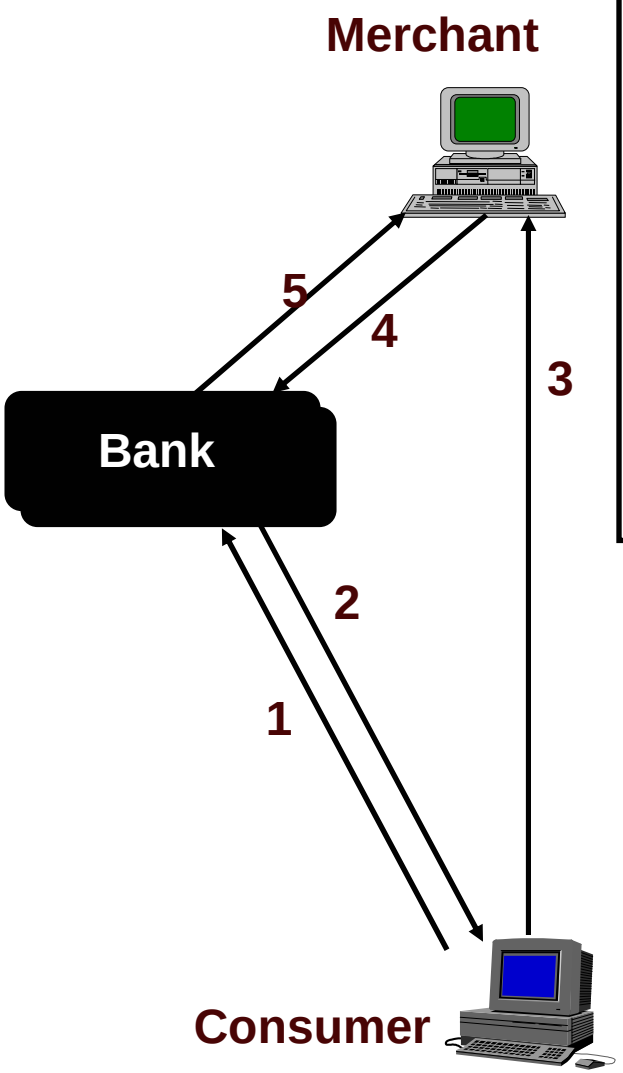
- Protects consumer's card info
- Use Internet for reaching Cybercash gateway to acquirers
- Adds to credit card card cost



Why study anonymity?

- Privacy concerns
  - individual
  - corporate
  - national
- Technology for collecting private statistics
- Understand theoretical limits, countermeasures
- Understanding semi-anonymity
  - Allows government search in exceptional circumstances
- Insights
  - e-commerce
  - distributed protocols
  - cryptography
  - survivability

Digicash



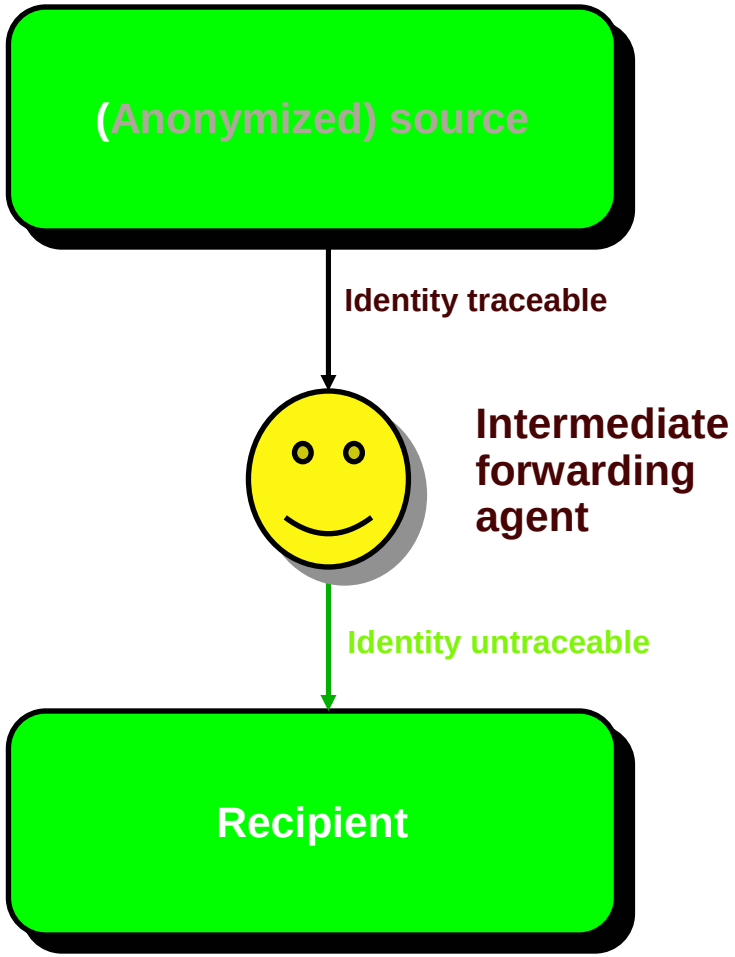
1. Consumer asks bank for anonymous digicash
2. Bank sends anonymous digicash bits to consumer
3. Consumer sends digicash to merchant in payment
4. Merchant checks that digicash has not been double spent
5. Bank verifies that digicash is valid

Problems

- No atomicity
- Anonymity restricted in US
- Interrupt transaction: ambiguous state
- Detecting double spending is expensive

Mixes

- Use intermediate forwarding agents
- Examples: onion routing, crowds, anonymizer.com, etc.
- Idea simultaneously thought of by several researchers
- Problems:
  - intermediary knows all
  - subject to traffic analysis and statistical analysis
  - can not link old messages to new messages



Pseudonymous identity

- Establish a consistent, but disguised identity
- Example: mail forwarders
- Can disguise basic facts about identity, but may be traceable from patterns of use
- Once identity is revealed, then all previous uses are traceable

Escrow

- Use pseudonym, but store real identity where law enforcement can find it.
  - Refinement: split identity into multiple parts
  - Store them in different locations
- Depends on procedural mechanisms (e.g. search warrants) for privacy
- Has drawbacks of pseudonym
- Government approach to cryptography

Auction types

- Auctions
  - Allocate scarce resources
  - Proposed to ration Internet bandwidth

- Three types of auctions

English auction (price goes up)
- advantages: encourages "honest" bids
- disadvantages: slow
  not private

Sealed bid auction
- advantages: constant time
- disadvantages: does not encourage "honest" bids,
  auctioneer knows all

Dutch auction (price goes down)
- advantages: protects privacy
- disadvantages: slow
  does not encourage "honest" bids

Vickrey auction

- Vickrey gave a way to combine best features of English auctions and sealed-bid auction

- Second-price auction
  - Highest bidder wins
  - Price is the value of the second highest bid
  - Example: Alice is highest bidder for $100; Bob is second highest bidder for $80; Alice wins the bid, but pays only $80