15-853:Algorithms in the Real World

Cryptography 3, 4 and 5

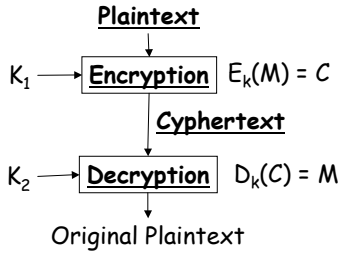
Cryptography Outline

Introduction: terminology, cryptanalysis, security
Primitives: one-way functions, trapdoors, …
Protocols: digital signatures, key exchange, ..
Number Theory: groups, fields, …
Private-Key Algorithms: Rijndael, DES
Public-Key Algorithms:
- Diffie-Hellman Key Exchange
- RSA, El-Gamal, Blum-Goldwasser
- Quantum Cryptography

Case Studies: Kerberos, Digital Cash

Public Key Cryptosystems

Introduced by Diffie and Hellman in 1976.



Public Key systems
K_1 = public key
K_2 = private key

Digital signatures
K_1 = private key
K_2 = public key

Typically used as part of a more complicated protocol.

One-way trapdoor functions

Both Public-Key and Digital signatures make use of one-way trapdoor functions.

Public Key:
- Encode: $c = f(m)$
- Decode: $m = f^{-1}(c)$ using trapdoor

Digital Signatures:
- Sign: $c = f^{-1}(m)$ using trapdoor
- Verify: $m = f(c)$

Example of SSL (3.0)

SSL (**Secure Socket Layer**) is the standard for the web (**https**).

Protocol (somewhat simplified): Bob -> amazon.com

B->A: **client hello**: protocol version, acceptable ciphers
A->B: **server hello**: cipher, session ID, $|\text{amazon.com}|_{\text{verisign}}$
B->A: **key exchange**, $\{\text{masterkey}\}_{\text{amazon's public key}}$
A->B: **server finish**: $([\text{amazon,prev-messages,masterkey}])_{\text{key1}}$
B->A: **client finish**: $([\text{bob,prev-messages,masterkey}])_{\text{key2}}$

} hand-shake

A->B: **server message**: $(\text{message1},[\text{message1}])_{\text{key1}}$
B->A: **client message**: $(\text{message2},[\text{message2}])_{\text{key2}}$

} data

$|h|_{\text{issuer}}$ = Certificate
= Issuer, $\langle h, h\text{'s public key, time stamp}\rangle_{\text{issuer's private key}}$

$\langle\ldots\rangle_{\text{private key}}$ = Digital signature $\{\ldots\}_{\text{public key}}$ = Public-key encryption
$[..]$ = Secure Hash $(\ldots)_{\text{key}}$ = Private-key encryption

key1 and key2 are derived from masterkey and session ID

Public Key History

Some algorithms

- **Merkle-Hellman**, 1978, based on "knapsack problem"
- **McEliece**, 1978, based on algebraic coding theory
- **RSA**, 1978, based on factoring
- **Rabin**, 1979, security can be reduced to factoring
- **ElGamal**, 1985, based on Discrete logs
- **Blum-Goldwasser,** 1985, based on quadratic residues
- **Elliptic curves**, 1985, discrete logs over Elliptic curves
- **Chor-Rivest,** 1988, based on knapsack problem
- **NTRU**, 1996, based on Lattices
- **XTR,** 2000, based on discrete logs of a particular field

Diffie-Hellman Key Exchange

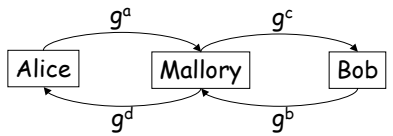
A group $(G,*)$ and a primitive element (generator) g is made public.

- Alice picks a, and sends g^a (publicly) to Bob
- Bob picks b and sends g^b (publicly) to Alice
- Alice computes $(g^b)^a = g^{ab}$
- Bob computes $(g^a)^b = g^{ab}$
- The shared key is g^{ab}

Note this is easy for Alice or Bob to compute, but assuming discrete logs are hard, is hard for anyone with only g^a and g^b.

Can someone see a problem with this protocol?

Person-in-the-middle attack



$\text{Key}_1 = g^{ad}$ $\text{Key}_1 = g^{cb}$

Mallory gets to listen to everything.

Merkle-Hellman

Gets "security" from the **Subet Sum** (also called **knapsack**) problem which is NP-hard to solve in general.

Subset Sum (Knapsack): Given a sequence W = {w_0,w_1, …,w_{n-1}}, $w_i \in Z$ of weights and a sum S, calculate a boolean vector B, such that:

$$\sum_{i=0}^{i<n} B_i W_i = S$$

Even deciding if there is a solution is NP-hard.

Merkle-Hellman

W is **superincreasing** if: $w_i \geq \sum_{j=0}^{j-1} w_j$

It is easy to solve the subset-sum problem for superincreasing W in O(n) time – give me a proof!

Main idea:

- Hide the easy case by multiplying each w_i by a constant **a** modulo a prime **p**

$$w_i' = a * w_i \bmod p$$

- Knowing a and p allows you to retrieve easy case

Merkle-Hellman

What we need

- $w_1, \cdots, w_n$ superincreasing integers
- $p > \sum_{i=1}^{n} w_i$ and prime
- a, $2 \leq a \leq p-1$
- $w'_i = a\, w_i \bmod p$

Public Key: w'_i
Private Key: w_i, p, a,

Encode:
$y = E(m) = \sum_{i=1}^{n} m_i\, w'_i$

Decode:
$z = a^{-1}\, y \bmod p$
$= a^{-1} \sum_{i=1}^{n} m_i\, w'_i \bmod p$
$= a^{-1} \sum_{i=1}^{n} m_i a w_i \bmod p$
$= \sum_{i=1}^{n} m_i\, w_i$
Solve subset sum prob:
$(w_1, \cdots, w_n, z)$
obtaining $m_1, \cdots m_n$

Merkle Hellman: Problem

Was broken by Shamir in 1984.

Shamir showed how to use integer programming to solve the particular class of Subset Sum problems in polynomial time.

Lesson: don't leave your trapdoor loose.

RSA

Invented by Rivest, Shamir and Adleman in 1978

Based on **difficulty of factoring**.

Used to **hide** the size of a group Z_n^* since:

. $$\left|Z_n^*\right| = \phi(n) = n\prod_{p|n}(1-1/p)$$

Factoring has not been reduced to RSA

- an algorithm that generates m from c does not give an efficient algorithm for factoring

On the other hand, factoring has been reduced to finding the private-key.

- there is an efficient algorithm for factoring given one that can find the private key.

RSA Public-key Cryptosystem

What we need:

- p and q, primes of approximately the same size
- n = pq
  $\phi(n) = (p-1)(q-1)$
- $e \in Z_{\phi(n)}^*$
- $d = e^{-1} \bmod \phi(n)$

Public Key: (e,n)
Private Key: d

Encode:
$m \in Z_n$
$E(m) = m^e \bmod n$

Decode:
$D(c) = c^d \bmod n$

RSA continued

Why it works:

$D(c) = c^d \bmod n = c^d \bmod pq$

$= m^{ed} \bmod pq$

$= m^{1 + k(p-1)(q-1)} \bmod pq$

$= m \bullet (m^{p-1})^{k(q-1)} \bmod pq = m \bullet (m^{q-1})^{k(p-1)} \bmod pq$

Chinese Remainder Theorem: If p and q are relatively prime, and a = b mod p and a = b mod q, then a = b mod pq.

$m \bullet (m^{p-1})^{k(q-1)} = m \bmod p$
$m \bullet (m^{q-1})^{k(p-1)} = m \bmod q$

$D(c) = m \bmod pq$

RSA computations

To **generate the keys**, we need to

- **Find two primes p and q**. Generate candidates and use primality testing to filter them.
- **Find e^{-1} mod (p-1)(q-1).** Use Euclid's algorithm. Takes time $\log^2(n)$

To **encode and decode**

- **Take m^e or c^d**. Use the power method. Takes time $\log(e)\log^2(n)$ and $\log(d)\log^2(n)$.

In practice e is selected to be small so that encoding is fast.

Security of RSA

Warning:

- Do not use this or any other algorithm naively!

Possible security holes:

- Need to use "safe" primes p and q. In particular p-1 and q-1 should have large prime factors.
- p and q should not have the same number of digits. Can use a middle attack starting at sqrt(n).
- e cannot be too small
- Don't use same n for different e's.
- You should always "pad"

Algorithm to factor given d and e

If an attacker has an algorithm that generates d from e, then he/she can factor n in PPT. Variant of the Rabin-Miller primality test.

Function TryFactor(e,d,n)

1. write **ed** – 1 as $2^s r$, r odd
2. choose w at random < n
3. $v = w^r \bmod n$
4. if v = 1 then return(**fail**)
5. while $v \neq 1 \bmod n$
6. $v_0 = v$
7. $v = v^2 \bmod n$
8. if $v_0 = n - 1$ then return(**fail**)
9. return(**pass,** $\gcd(v_0 + 1, n)$)

LasVegas algorithm
Probability of pass is > .5.
Will return p or q if it passes.
Try until you pass.

$w^{2^s r} = w^{ed-1}$
$= w^{k\phi} = 1 \bmod n$

$v_0^2 = 1 \bmod n$
$(v_0 - 1)(v_0 + 1) = k'n$

RSA Performance

Performance: (600Mhz PIII) (from: ssh toolkit):

Algorithm	Bits/key		Mbits/sec
RSA Keygen	1024	.35sec/key	
	2048	2.83sec/key	
RSA Encrypt	1024	1786/sec	3.5
	2048	672/sec	1.2
RSA Decrypt	1024	74/sec	.074
	2048	12/sec	.024
ElGamal Enc.	1024	31/sec	.031
ElGamal Dec.	1024	61/sec	.061
DES-cbc	56		95
twofish-cbc	128		140
Rijndael	128		180

RSA in the "Real World"

Part of many standards: PKCS, ITU X.509, ANSI X9.31, IEEE P1363

Used by: SSL, PEM, PGP, Entrust, …

The standards specify many details on the implementation, e.g.

- e should be selected to be small, but not too small
- "multi prime" versions make use of n = pqr… this makes it cheaper to decode especially in parallel (uses Chinese remainder theorem).

Factoring in the Real World

Quadratic Sieve (QS):

$$T(n) = e^{(1+o(n))(\ln n)^{1/2}(\ln(\ln n))^{1/2}}$$

- Used in 1994 to factor a 129 digit (428-bit) number. 1600 Machines, 8 months.

Number field Sieve (NFS):

$$T(n) = e^{(1.923+o(1))(\ln n)^{1/3}(\ln(\ln n))^{2/3}}$$

- Used in 1999 to factor 155 digit (512-bit) number. 35 CPU years. At least 4x faster than QS

The RSA Challenge numbers

ElGamal

Based on the difficulty of the discrete log problem.

Invented in 1985

Digital signature and Key-exchange variants

- DSA based on ElGamal AES standard
- Incorporated in SSL (as is RSA)
- Public Key used by TRW (avoided RSA patent)

Works over various groups

- Z_p,
- Multiplicative group $GF(p^n)$,
- Elliptic Curves

ElGamal Public-key Cryptosystem

(G,*) is a group

- α a generator for G
- $a \in Z_{|G|}$
- $\beta = \alpha^a$

G is selected so that it is hard to solve the discrete log problem.

Public Key: (α, β) and some description of G
Private Key: a

Encode:
Pick random $k \in Z_{|G|}$
$E(m) = (y_1, y_2)$
$= (\alpha^k, m * \beta^k)$

Decode:
$D(y) = y_2 * (y_1^a)^{-1}$
$= (m * \beta^k) * (\alpha^{ka})^{-1}$
$= m * \beta^k * (\beta^k)^{-1}$
$= m$

You need to know a to easily decode y!

ElGamal: Example

G = Z_{11}^*

- $\alpha = 2$
- $a = 8$
- $\beta = 2^8 \pmod{11} = 3$

Public Key: (2, 3), Z_{11}^*
Private Key: a = 8

Encode: 7
Pick random k = 4
$E(m) = (2^4, 7 * 3^4)$
$= (5, 6)$

Decode: (5, 6)
$D(y) = 6 * (5^8)^{-1}$
$= 6 * 4^{-1}$
$= 6 * 3 \pmod{11}$
$= 7$

Probabilistic Encryption

For RSA one message goes to one cipher word. This means we might gain information by running $E_{public}(M)$.

Probabilistic encryption maps every M to many C randomly. Cryptanalysists can't tell whether $C = E_{public}(M)$.

ElGamal is an example (based on the random k), but it doubles the size of message.

BBS "secure" random bits

BBS (Blum, Blum and Shub, 1984)

- Based on difficulty of factoring, or finding square roots modulo n = pq.

Fixed

- p and q are primes such that p = q = 3 (mod 4)
- n = pq (is called a Blum integer)

For a particular bit seq.

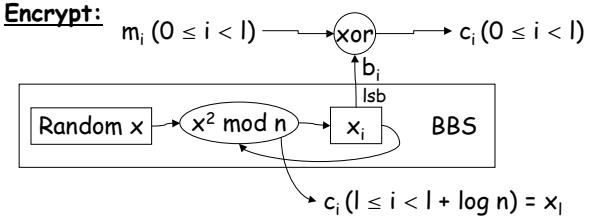
- **Seed:** random x relatively prime to n.
- **Initial state:** $x_0 = x^2$
- **i^{th} state:** $x_i = (x_{i-1})^2$
- **i^{th} bit:** lsb of x_i

Note that: $x_0 = x_i^{-2^i \bmod \phi(n)} (\bmod n)$
Therefore knowing p and q allows us to find x_0 from x_i

Blum-Goldwasser: A stream cypher

Public key: n (= pq) **Private key**: p or q

Encrypt:



Decrypt:

Using p and q, find $x_0 = x_i^{-2^i \bmod (p-1)(q-1)} (\bmod n)$

Use this to regenerate the $\mathbf{b_i}$ and hence $\mathbf{m_i}$

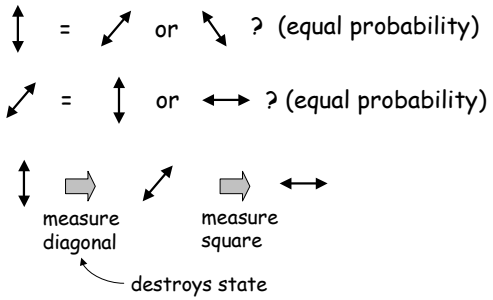
Quantum Cryptography

In quantum mechanics, there is no way to take a measurement without potentially changing the state. E.g.

- Measuring position, spreads out the momentum
- Measuring spin horizontally, "spreads out" the spin probability vertically

Related to Heisenberg's uncertainty principal

Using photon polarization



Quantum Key Exchange

1. Alice sends bob photon stream randomly polarized in one of 4 polarizations:
2. Bob measures photons in random orientations
   e.g.: X + + X X X + X (orientations used)
   \ | - \ / / - \ (measured polarizations)
   and tells Alice in the open what orientations he used, but not what he measured.
3. Alice tells Bob in the open which are correct
4. Bob and Alice keep the correct values

Susceptible to a **man-in-the-middle** attack

In the "real world"

Not yet used in practice, but experiments have verified that it works.

IBM has working system over 30cm at 10bits/sec.

More recently, up to 10km of fiber.

Cryptography Outline

Introduction: terminology, cryptanalysis, security
Primitives: one-way functions, trapdoors, …
Protocols: digital signatures, key exchange, ..
Number Theory: groups, fields, …
Private-Key Algorithms: Rijndael, DES
Public-Key Algorithms: Knapsack, RSA, El-Gamal, …
Case Studies:
- Kerberos
- Digital Cash

Kerberos

A key-serving system based on Private-Keys (DES).

Assumptions

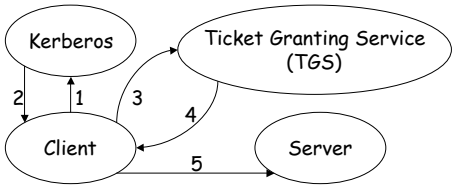
- Built on top of TCP/IP networks
- Many "**clients**" (typically users, but perhaps software)
- Many "**servers**" (e.g. file servers, compute servers, print servers, …)
- User machines and servers are potentially insecure without compromising the whole system
- A **kerberos server** must be secure.

At Carnegie Mellon

Single password (in SCS, ECE or ANDREW) gives you access to:

- Andrew file system
- Loging into andrew, ece, or scs machines
- POP and IMAP (mail servers)
- SSH, RSH, FTP and TELNET
- Electronic grades, HUB, …
- Root access

Kerberos V



1. Request *ticket-granting-ticket* (TGT)
2. <TGT>
3. Request *server-ticket* (ST)
4. <ST>
5. Request service

Tickets

Ticket: A message "signed" by a "higher authority" giving you certain rights at a particular server S.

$T_{C,S}$ = S, $\{C,A,V,K_{C,S}\}K_S$

C = client S = server

K_S = server key. A static key only known by the server and the "higher authority" (not by the client).

A = client's network address

V = time range for which the ticket is valid

$K_{C,S}$ = client-server key. A dynamic key specific to this ticket. Known by the server and client.

A ticket can be used **many times** with a single server.

Authenticators

Authenticator: a message "signed" by the client identifying herself. It must be accompanied by a ticket. It says "I have the right to use this ticket"

$A_{C,S} = \{C,T,[K]\}K_{C,S}$

C = client S = server

$K_{C,S}$ = client-server key. A dynamic key specific to the associated ticket.

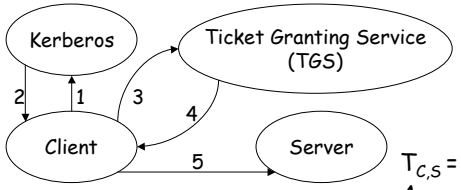
T = timestamp (must be in range of associated ticket)

K = session key (used for data transfer, if needed)

An authenticator can only be used **once**.

A single **ticket** can use many **authenticators**

Kerberos V Messages



$T_{C,S} = S, \{C,A,V,K_{C,S}\}K_S$
$A_{C,S} = \{C,T,[K]\}K_{C,S}$

1. Client to Kerberos: $\{C,TGS\}K_C$
2. Kerberos to Client: $\{K_{C,TGS}\}K_C$, $T_{C,TGS}$
3. Client to TGS: $A_{C,TGS}$, $T_{C,TGS}$
4. TGS to Client: $\{K_{C,S}\}K_{C,TGS}$, $T_{C,S}$
5. Client to Server: $A_{C,S}$, $T_{C,S}$

Possibly repeat

Kerberos Notes

All machines have to have synchronized clocks
- Must not be able to reuse authenticators

Servers should store all previous and valid tickets
- Help prevent replays

Client keys are typically a one-way hash of the password. Clients do not keep these keys.

Kerberos 5 uses CBC mode for encryption Kerberos 4 was insecure because it used a nonstandard mode.

Electronic Payments

Privacy
- Identified
- Anonymous

Involvement
- Offline (just buyer and seller) more practical for "micropayments"
- Online
  - Notational fund transfer (e.g. Visa, CyberCash)
  - Trusted 3rd party (e.g. FirstVirtual)

Today: "**Digital Cash**" (anonymous and possibly offline)

Some more protocols

1. Secret splitting (and sharing)
2. Bit commitment
3. Blind signatures

Secret Splitting

Take a secret (e.g. a bit-string **B**) and split it among multiple parties such that all parties have to cooperate to regenerate any part of the secret.

An implementation:

- Trent picks a random bit-string **R** of same length as **B**
- Sends Alice **R**
- Sends Bob **R xor B**

Generalizes to k parties by picking k-1 random bit-strings.

Secret Sharing

m out of n (m < n) parties can recreate the secret.

Also called an **(m,n)-threshold scheme**

An implementation (Shamir):

- Write secret as coefficients of a polynomial $GF(p^l)[x]$ of degree m-1 ($n \leq p^l$).
  $p(x) = c_{m-1}x^{m-1} + \ldots + c\_1\, x + c\_0$
- Evaluate p(x) at n distinct points in $GF(p^l)$
- Give each party one of the results
- Any m results can be used to reconstruct the polynomial.

Bit Commitment

Alice commits a bit to Bob without revealing the bit (until Bob asks her to prove it later)

An implementation:

- **Commit**
  - Alice picks random r, and uses a one-way hash function to generate **y = f(r,b)** f(r,b) must be "unbiased" on b (y by itself tells you nothing about b).
  - Alice **sends** Bob **y**.
- **Open** (expose bit and prove it was commited)
  - Alice **sends** Bob **b** and **r**.

Example: $y = \text{Rijndael}_r(000\ldots b)$, perhaps

Blind Signatures

Sign a message **m** without knowing anything about **m**

Sounds dangerous, but can be used to give "value" to an anonymous message

- Each signature has meaning:
  $5 signature, $20 signature, …

Blind Signatures

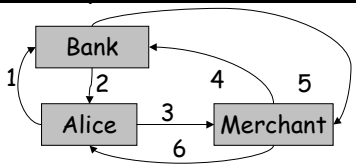
An implementation: based on RSA

Trent blindly signs a message m from Alice

- Trent has public key (e,n) and private key d
- Alice selects random r < n and generates
  $m' = m\ r^e \bmod n$
  and sends it to Trent.
  This is called **blinding** m
- Trent signs it: $s(m') = (m\ r^e)^d \bmod n$
- Alice calculates:
  $s(m) = s(m')\ r^{-1} = m^d\ r^{ed-1} = m^d \bmod n$

Patented by Chaum in 1990.

An anonymous online scheme



1. Blinded Unique Random large ID (no collisions). Sig_{alice}(request for $100).
2. $Sig_{bank\_\$100}$(blinded(ID)): signed by bank
3. $Sig_{bank\_\$100}$(ID)
4. $Sig_{bank\_\$100}$(ID)
5. OK from bank
6. OK from merchant

Minting: 1. and 2.
Spending: 3.-6.
Left out encryption

eCash

Uses the protocol

Bought assets and patents from Digicash
Founded by Chaum, went into Chapter 11 in 1998

Has not picked up as fast as hoped

- Credit card companies are putting up fight and transactions are becoming more efficient
- Government is afraid of abuse

Currently mostly used for Gift Certificates, but also used by Deutsche Bank in Europe.

The Perfect Crime

- Kidnapper takes hostage
- Ransom demand is a series of blinded coins (IDs) and a request to publish the signed blinded IDs in a newspaper (they're just strings)
- Banks signs the coins to pay ransom and publishes them
- Only the kidnapper can unblind the coins (only she knows the blinding factor)
- Kidnapper can now use the coins and is completely anonymous

Offline Anonymous Cash

A paradox: Digital cash is just a sequence of bits. By their very nature they are trivial to counterfeit.
Without a middleperson, how do you make sure that the user is not spending them twice?

I go to Amazon and present them a $20 "coin".

I then go to Ebay and use the same $20 "coin".

In the **offline** scheme they can't talk to each other or a bank during the transaction.

In an **anonymous** scheme they can't know who I am.

Any ideas?

Chaum's protocol for offline anonymous cash

Properties:

- If used properly, Alice stays anonymous
- If Alice spends a coin twice, she is revealed
- If Merchant remits twice, this is detected and Alice remains anonymous
- Must be secure against Alice and Merchant colluding
- Must be secure against one framing the other.

An amazing protocol

Basic Idea

Use blinded coins

Include Alice's ID in the coin

Alice uses interactive proof with merchant to prove that her ID is in the coin, without revealing ID.

If she does a second interactive proof on same coin it will reveal her ID.

"Questions" merchant asks as part of the proof are chosen at random, so it is unlikely the same ones will be asked twice.

Similar to "zero knowledge" ideas.

Chaum's protocol: **money orders**

u = Alice's account number (identifies her)

$r_0, r_1, \ldots, r_{n-1}$ = n random numbers

(ul_i, ur_i) = a secret split of u using r_i $(0 \leq i < n)$
e.g. using $(r_i, r_i \text{ xor } u)$

vl_i = a bit commitment of all bits of ul_i

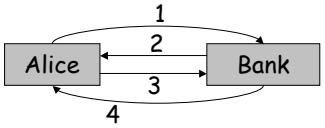
vr_i = a bit commitment of all bits of ur_i

Money order (created by Alice from u):

- Amount
- Unique ID
- $(vl_0,vr_0), (vl_1,vr_1), \ldots, (vl_{n-1},vr_{n-1})$

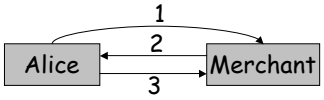
Alice keeps $r_0, \ldots, r_{n-1}$ and commitment keys.

Chaum's protocol: **Minting**



1. Two blinded money orders and Alice's account #
2. A request to unblind and prove all bit commitments for one of the two orders (chosen at random)
3. The blinding factor and proof of commitment for that order
4. Assuming step 3. passes, the other blinded order signed

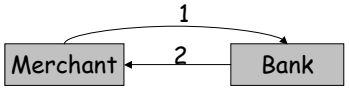
Chaum's protocol: **Spending**



1. The signed money order C (unblinded)
2. A random bit vector B of length n
3. For each i if $B_i = 0$ return bit values for ul_i else return bit values for ur_i
   Include all "proofs" that the ul or ur match vl or vr

Now the merchant checks that the money order is properly signed by the bank, and that the ul or ur match the vl or vr

Chaum's protocol: **Returning**



1. The signed money order
   The vector B along with the values of ul_i or ur_i that it received from Alice.
2. An OK, or fail

If fail, i.e., already returned:

1. If B matches previous order, the Merchant is guilty
2. Otherwise Alice is guilty and can be identified since for some i (where Bs don't match) the bank will have (ul_i, ur_i), which reveals her secret u (her identity).