Privacy Enhancing Technologies
CSE 701 Fall 2017

Lecture 2: Anonymity Applications

Department of Computer Science and Engineering
University at Buffalo

Lecture Outline

- Anonymous communication
  - mixes, anonymizing proxies, onion routing
- Anonymous authentication
- Applications requiring anonymity
  - digital money
  - voting

Anonymous Communication

- Are we anonymous on the internet?
  - if we read a web page or connect to a chat channel, the server knows from what address we are coming
  - if you send an encrypted email, the endpoints still can be recovered
- But does it really matter?
  - internet surveillance techniques are known as traffic analysis
    - knowing the source and destination of our traffic allows others to track your behavior and interests
  - an e-commerce website can use price discrimination based on your origin
  - this can threaten your job and physical safety by revealing who and where you are (e.g., when traveling abroad)

Anonymous Communication

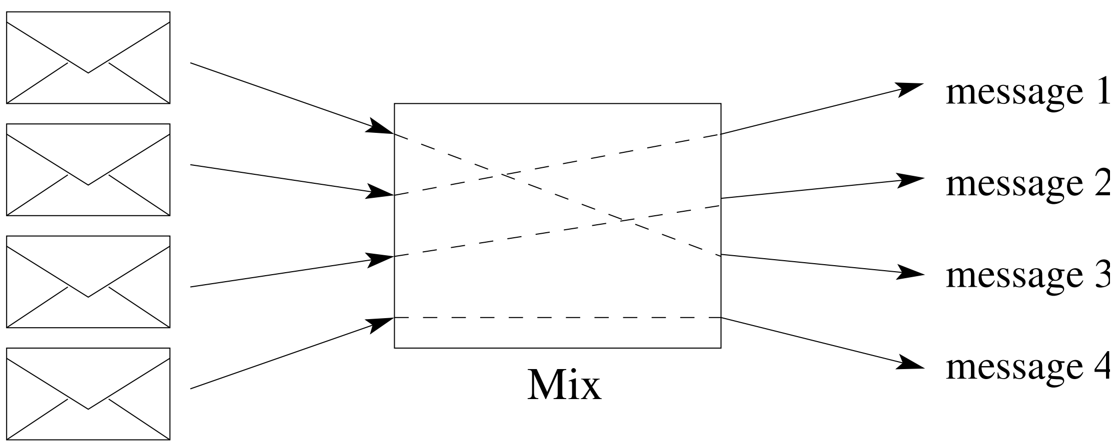
- Traffic analysis focuses on the header that discloses source, destination, size, timing, etc.
  - this is seen by the recipient of your communications, authorized (i.e., Internet service providers) and even unauthorized intermediaries
  - a simple form of traffic analysis can involve someone sitting between the sender and recipient looking at headers
  - governments can spy on multiple parts of the Internet and using sophisticated statistical techniques to track communication patterns
- "We kill people based on metadata" – M. Hayden, former director of NSA

Benefits of Anonymous Communication

- If we build anonymous communication channels, what are we able to do?
  - organizations and individuals can share information over public networks without compromising privacy
  - individuals can keep websites from tracking them
  - individuals can connect to news sites and other services when these are blocked by their local Internet providers
  - individuals can conduct socially sensitive communication (e.g., use chat rooms forums for rape and abuse survivors or people with illnesses)
  - journalists can communicate more safely with whistleblowers and dissidents
  - law enforcement can visit and surveil websites without leaving government IP addresses in their logs

Anonymous Communication

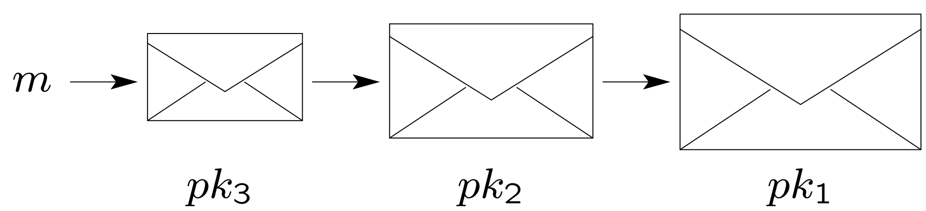
- Anonymity likes company: you cannot be anonymous by yourself
- There are several technical approaches to achieve anonymity such as mixes and proxies
- A mix receives encrypted messages, randomly permutes and decrypts them



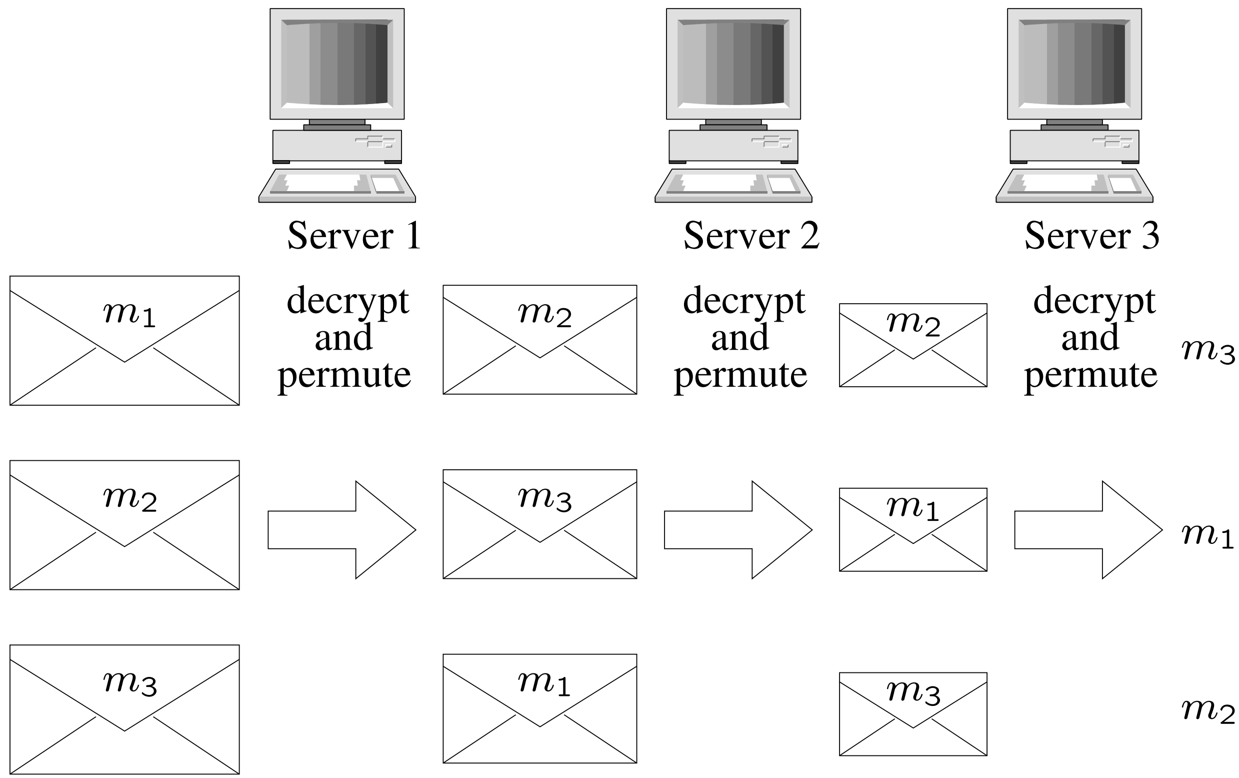
- The key property is that an adversary cannot tell which ciphertext corresponds to a given message

Mixes

- The basic mixnet was introduced by Chaum in 1981
  - it was introduced for email and other high latency applications because of its use of expensive public-key cryptography
  - there is a number of servers each with its own public key pk_i
  - to send a message m through servers 1, 2, and 3, envelope it using all of the servers' keys $c = \mathsf{Enc}_{pk_1}(\mathsf{Enc}_{pk_2}(\mathsf{Enc}_{pk_3}(m)))$



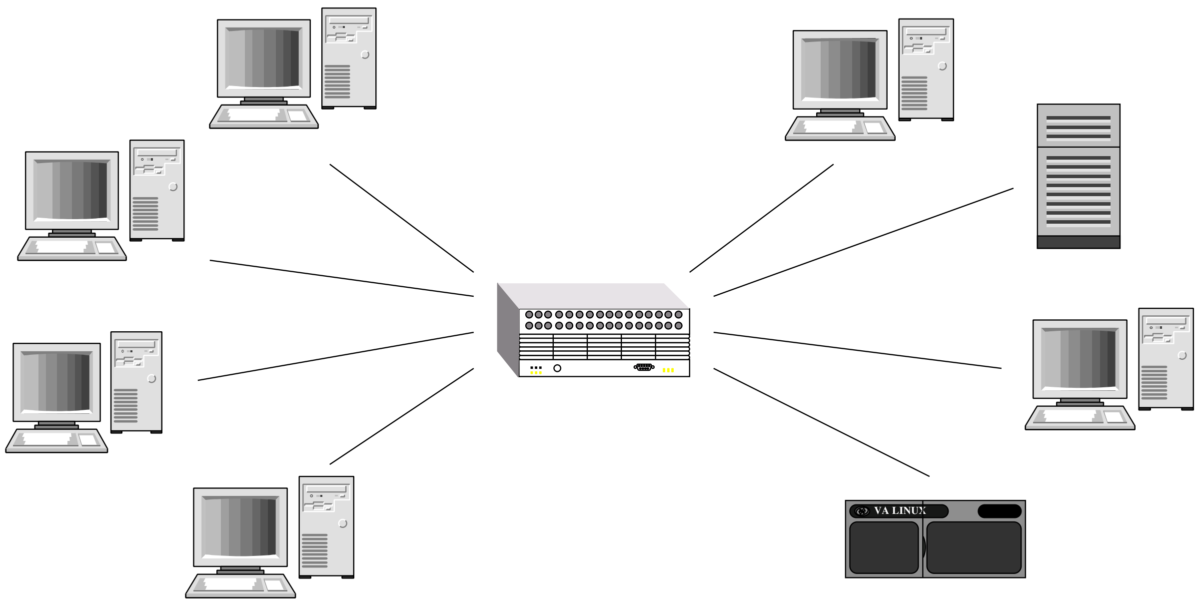
Mixes



- Each server on the way knows only which server gave it data and which server it is giving data to
- One honest server preserves privacy!

Proxies

- Anonymizing proxy



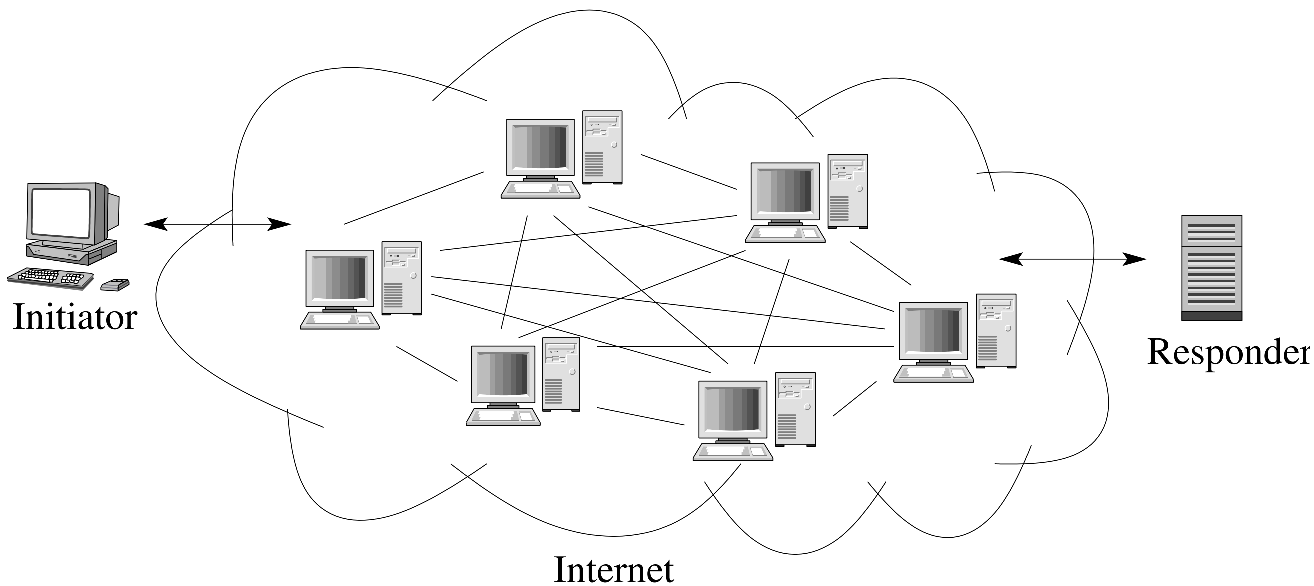
  - communications appear to come from the proxy, not true senders
  - it can use low-cost symmetric encryption (or no encryption)
    - it thus is appropriate for web connections, SSL/TLS, ssh, etc.

Proxies

- Anonymizing proxy
  - advantages: simple, focuses a lot of traffic for more anonymity
  - disadvantages: a single point of failure, compromise, attack
  - risks of using anonymizing HTTP proxies
    - all data you send to the service must first go through the proxy
    - a malicious proxy server can record everything you send to it, including unencrypted logins and passwords
    - if you don't trust the proxy, don't send any sensitive information unencrypted

Onion Routing

- Onion Routing can be used to build traffic analysis-resistant infrastructure
  - the main idea is to combine advantages of mixes and proxies
    - use (expensive) public-key crypto to establish circuits and (cheaper) symmetric-key crypto to move data
- The Onion Routing (TOR) network



TOR

- Tor establishes routing connections called circuits
  - session keys are negotiated using servers' public keys
    - the client chooses a set of onion routers to tunnel packets through
    - the client's proxy establishes a session key and circuit with the first onion router on the list
    - proxy tunnels through that circuit to extend to the second router on the list, etc.
  - many client applications can connect and communicate over the established circuit
  - after some time session keys used in a circuit are refreshed to limit the impact of key compromise

TOR Hidden Services

- Tor also makes it possible for users to offer services while hiding their locations
  - they are called hidden services and can be used for web publishing, instant messaging servers, etc.
  - nobody is able to determine who is offering the site and nobody know who is posting to it
- Setting up a hidden service includes
  - selecting a few onion routers as introduction points
  - advertising these points on the lookup service
  - building a circuit from each introduction point to the service

TOR

- Directory servers maintain a list of onion routers (their location, current keys, etc.) and control which nodes can join the networks
- Tor properties
  - trust is distributed like in mixes
  - replay attacks are not effective
  - perfect forward secrecy is achieved
  - it can adapt to network dynamics
- Tor is an active research project and software is available for download
  - see http://www.torproject.org for more detail

Anonymous Authentication and Credentials

- Anonymous authentication allows one to prove her credentials and gain access to some resources without disclosing her identity
  - e.g., a user can prove current membership in a digital library and anonymously browse books and articles
- Anonymous authentication and other functionalities can be realized by using anonymous credentials
  - a user obtains authority's certification on some attributes, which can only be partially known by the authority
  - later the user can reveal only necessary information to prove validity of his credentials and gain access
  - the credentials need to changed on each use to prevent multiple showing to be linked together

Anonymous Credentials

- Examples
  - the user can prove that she is over 21 without revealing the birth date (or anything else)
  - the user can prove that she is a student member and the expiration date is some time in the future
  - the user can prove membership in a service and that the membership is current
- A common way of constructing anonymous credentials is by using signatures with special properties

Anonymous Credentials

- Signatures with protocols were originally proposed by Camenisch and Lysyanskaya
  - traditional signature schemes consist of three algorithms: key generation, signing, and signature verification
  - signatures with protocols come with interactive algorithms (or protocols):
    - signing can be an interactive process if the signer doesn't have all values it is signing in the clear
      - instead the user can supply commitments to some values (and possibly prove some properties about them), which the signer uses to form its signature
    - the signature can be proved to be valid and statements over signed values can be proved without revealing additional information

Commitments

- The above can be realized using commitments and zero-knowledge proofs of knowledge
- A commitment scheme allows one
  - to produce a commitment $\mathsf{com}(m)$ on message m
    - commitments are often randomized and use new randomness r to form each $\mathsf{com}(m)$
  - and later open $\mathsf{com}(m)$ to reveal m
- Each commitment must satisfy the following properties:
  - hiding: given $\mathsf{com}(m)$, it is not feasible to learn information about m
  - binding: given $\mathsf{com}(m)$, it is not feasible to open it to another value different from m

Zero-Knowledge Proofs of Knowledge

- Zero-knowledge proofs of knowledge (ZKPK) allow one to prove statements about private values without revealing anything else
- ZKPKs exist for many types of problems including all NP-languages
  - many general constructions are not efficient and of theoretical interest
  - efficient ZKPKs are available for statements based on discrete logarithms
  - for $y = g_1^{x_1} g_2^{x_2} \ldots g_n^{x_n}$, the integers $x_1, \ldots, x_n$ are called the discrete logarithm representation of y to the bases $g_1, \ldots, g_n$
  - ZKPKs are known to prove that x_i is equal to some value, that x_i lies in a specific range, etc.
  - it is required that no information about private values is revealed beyond validity of the statement

Anonymous Credentials

- We can combine commitments and ZKPK to realize anonymous credentials based on signatures with protocols as follows:
  - during credential issuance, if the signer not to have access to some attribute to be signed, the user sends a commitment to that value instead
  - the user can also prove in ZK that the committed value satisfied certain properties
  - during credential usage, the user typically has to randomize its credentials first and prove signature validity
  - then the user can reveal certain attributes or only prove in ZK that they satisfy the requirements
    - e.g., the user can prove that the expiration date is in the future without revealing its exact value

Anonymous Credentials

- Anonymous credentials have multiple applications from anonymous authentication to electronic cash
- One significant issue with using anonymous credentials in a commercial setting is prevention of duplicating user credentials
  - mechanisms for accomplishing this vary depending on the application

Electronic Cash

- As we perform many transactions in electronic form, there is a need for electronic money
  - check and credit cards leave trails
  - can we have an equivalent of anonymous cash?
- Properties of regular cash
  - it is anonymous and untraceable
  - it can be used off-line, not connected to a bank
  - it is transferable
  - it has different denominations, and one can make change with it
  - it can be used only once (or stolen)

Electronic Cash

- We might want to have similar properties for digital cash that can be sent through computer networks
  - in many constructions, a digital coin is implemented as a token which you obtain in exchange for bank account's debit
  - for a digital coin, we'll want to have:
    - anonymity: coin spending cannot be linked to its issuance
    - double spending prevention: a dishonest user cannot spend a coin at two merchants and a dishonest merchant cannot deposit a spent coin more than once
  - some constructions also achieve:
    - transferability: a coin can be transferred from one user to another
    - divisibility: a coin can be divided into coins of smaller denominations

Electronic Cash

- Early solutions go back to early 80s with constructions by Chaum and others
  - while they provided new ideas, their performance was undesirable
  - they required thousands of public-key operations per coin, large communication and maintenance of large databases
- More efficient constructions came along
  - due to Brands (90s)
  - due to Camenisch, Lysyanskaya, and others (00s)
- Bitcoin by Nakamoto in late 00s revolutionized the field

Electronic Cash

- E-cash based on Camenisch-Lysyanskaya signatures with protocols can be realized as follows (simplified version):
  - coin issuance:
    - the user chooses random serial number s, computes $\mathsf{com}(s, t, id)$, where id is her identity and t is random, and sends it to the bank
    - the user proves to the bank in ZK that id is the user's real identity and that she knows s and t
    - the bank produces signature $\sigma_B(s, t, id)$, gives it to the user, and deducts money from her account
  - coin spending at a merchant:
    - the user forms new commitments to s, t, and id and proves to the merchant possession of the bank's signature on the committed values

Camenisch-Lysyanskaya E-Cash

- E-cash based on Camenisch-Lysyanskaya signatures (cont.):
  - coin spending at a merchant:
    - the merchant computes R as a one-way function of its identity and the transaction and sends R to the user
    - the user gives s and $T = id + tR$ to the merchant and proves that she computed them correctly using the commitments
    - the merchant stores s, T, R, all commitments and proofs
  - spent coin deposit:
    - the merchant goes to the bank and submits s, T, R, and the proofs
    - the bank verifies R using the merchant's identity, that s hasn't be used before, and all proofs
    - the bank deposits money to he merchant's account

Camenisch-Lysyanskaya E-Cash

- What properties do we achieve?
  - double spending detection and prevention
    - if Alice spends her coin at two merchants, her identity is revealed using the double-spending equation $T = id + tR$
    - one such equation reveals nothing about Alice's identity, but given $T_1 = id + tR_1$ and $T_2 = id + tR_2$, her identity can be computed
    - the merchant cannot forge spent coins on this or other serial numbers
    - only a single merchant can claim a spent coin because R is a one-way function of a merchant's identity
  - anonymity
    - even if the bank and merchant conspire, Alice remains anonymous (assuming she doesn't double spend)

Bitcoin

- Bitcoin takes a significantly different approach
  - it is a decentralized system that works without a central repository, administrator or traditional bank
  - transactions take place between users directly
  - they are recorded in a distributed public ledger called a blockchain
  - transactions stating that user A sends to user B x bitcoins are broadcast to the network
  - network nodes can validate transactions, add them to their copy of the ledger, and broadcast the changes to others
  - to perform independent verification of coin ownership, each node stores its own copy of the blockchain

Bitcoin

- Bitcoin
  - with a fixed frequency, newly accepted transactions are placed in a block, which is quickly published to all nodes
    - this indicates that a particular coin has been spent and prevents double-spending
  - every transaction uses an unspent output (coin) of a previous transaction as its input and must include a digital signature
    - keys used with bitcoins are typically anonymous
  - coin mining is a record keeping service
    - miners verify and collect new transactions in a block
    - for a block to be accepted, the miner has to submit a proof-of-work
    - inability to easily forge or alter blocks keeps the system consistent

Bitcoin in Practice

- Unlike other e-cash proposals, bitcoin is a widely used cryptocurrency
  - bitcoin can be exchanged for other currencies, products, and services
  - there are currently millions of bitcoin and other cryptocurrencies users
  - hundreds of thousands merchants and vendors accept bitcoin payments
  - 1 bitcoin is currently equal to over $4,000
    - it is significantly more volatile than other currencies or assets (such as gold)
- The concept of blockchain is also finding a wide use in different areas of our life

Voting and Elections

- Voting traditionally has been based on paper ballots and mechanical machines
- Now it is common to see electronic voting machine, but the process of voting largely remains unchanged
  - they are termed Direct Recording Electronic (DRE) voting machines
  - they run proprietary code that cannot be verified
  - there is no good way to tell that votes were recorded and counted correctly
  - additionally, they are known to have their own security flaws
- By using voting protocols based on cryptographic techniques, we can achieve stronger security properties

Electronic Voting

- Desirable properties of a voting process
  - only registered voters can cast their vote
  - a voter can cast her vote anonymously
  - a voter can cast her vote at most once
  - a voter can verify that the votes were counted correctly
  - a voter can verify that her vote was included in the count
  - it is not possible to construct a proof of how one voted
    - this prevents coercion and paid votes
- With traditional systems, full transparency is not present and individual users are unable to verify correctness of counts

Electronic Voting

- Cryptographic voting can be easily constructed using blind signing
  - this refers to the ability to sign a message without knowing its content
- A high-level overview of the solution
  - a registered voter obtains a voting authority's signature $\sigma(s)$ on a randomly chosen serial number s of its choice without revealing it
  - the voter submits $\sigma(s)$ together with its vote v over a secure channel
  - the authority checks whether s has already been recorded and if not, publishes it together with v
  - each voter can verify correctness of her data and that the votes were added correctly
  - coercion-resistance is achieved as the voter could claim a different (s, v) pair as her own

Electronic Voting

- Purely cryptographic constructions developed early are deemed to be cumbersome to use and understand by an average voter
- Thus, efforts have been directed toward developing more user-friendly solutions and those that incorporate paper ballots or receipts
- Some examples are:
  - Chaum's (2004) and Neff's (2005) cryptographic voting protocols for use in DRE voting machines
  - Adida-Rivest (2006) scratch and vote paper-based cryptographic voting
  - all of them offer public verifiability

Electronic Voting

- Adida and Rivest Scratch & Vote construction has the following features:
  - ballot casting is entirely paper- and pen-based
    - candidates appear on a ballot in a randomly permuted order
    - a voter detaches portions that allow one to see her vote
    - the retained portion serves the purpose of the receipt
  - ballots contain all necessary information for auditing
    - a scratch surface on an empty ballot can be used to audit the process
  - each voter can audit the process on election day, prior to casting her ballot
  - the votes are stored and added together inside homomorphic encryption
    - only one decryption by election officials is needed to retrieve the result

Summary

- The ability to be truly anonymous has advantages in a number of settings
  - it enables communication on sensitive or controversial issues
  - it prevents others from learning a lot of information about our lives, habits, interests, etc.
  - it protects interest of businesses and government
  - it enables applications where anonymity is one of foremost requirements
- Work on anonymity started in early 80s and continues to date
  - efficient constructions for e-cash and anonymous credentials exist
  - there is room for further improvements