Chair for Network Architectures and Services
Department of Informatics
TU München – Prof. Carle

Peer-to-Peer Systems and Security

Chapter 3
3.1 Anonymity

Dipl.-Inform. Heiko Niedermayer
Prof. Dr. Georg Carle

Vorlesung SS 2009

Overview

- Motivation
- Anonymity
- Adversery Models
- Anonymity Measures
- Basic concepts
  - Re-Routing
  - Mixing
  - Layered-encryption
  - Padding / Dummy Traffic
- System concepts
  - Infrastructure, Cascade, P2P
- Conclusion

Motivation

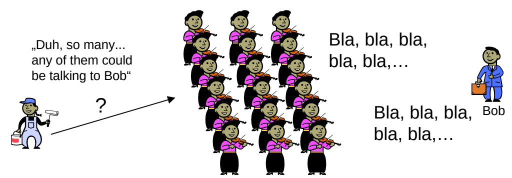
- Alice and Bob communicate using encryption.
  → Eve cannot read the data Alice and Bob are sending.
  But…
  → Eve knows that Alice and Bob are communicating.
  → Eve knows the amount of data Alice and Bob are sending. Alice observes the traffic patterns.
  - e.g. Bob as Webserver may sent the page which is fingerprinted in having 13kB of data, and 13 included objects with size from 2kB to 117kB.
    → Eve knows what Bob is sending to Alice
    → encryption not sufficient for static content

Anonymity

„Anonymity is the state of being not identifiable within a set of subjects, the anonymity set.“
Andreas Pfitzmann et. al.

Anonymity Set

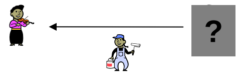
- The set of all possible suspects who might cause an action.
- The larger the anonymity set, the better the anonymity.
  - ... not completely true. Also, the more equal the probability for the suspects in the set, the better.



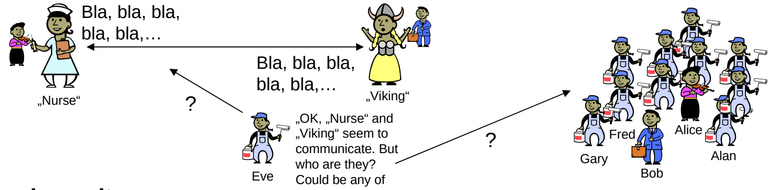
Anonymity

Terminology

- Sender Anonymity
  - The initiator of a message is anonymous. There may be a path back to the initiator.
  - „*???* to Bob"
- Receiver Anonymity
  - The receiver of a message is anonymous.
  - „Alice to *???*"
- Unlinkability
  - The observer cannot decide who is communicating with whom.
  - „*???* communicates with *???*"



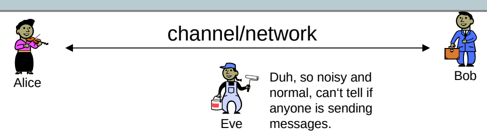
Pseudonymity



Pseudonymity

- A pseudonym is an identity for an entity in the system. It is a „false identity" (word origin of pseudonym) and not the true identity of the holder of the pseudonym. The holder hides the true identity behind the pseudonym.
  - e.g. a nickname in a forum, random string in an anonymity system
- Noone, but a trusted party may be able to link a pseudonym to the true identity of the holder of the pseudonym.
- A pseudonym can be tracked. We can observe its behaviour, but we do not know who it is.
  - „Nurse" is always „Nurse".
  - vs. anonymity: In anonymous systems, we cannot say if it is the same user „Nurse" again. An anonmyous entity is indistinguishable from all other anonymous entities.

Unobservability / Covert Channel



Unobservability

- „Unobservability is the state of items of interest being indistinguishable from any item of interest at all. " (according to Andreas Pfitzmann et. al)
- Eve will not see a different channel behaviour if Alice and Bob communicate or not.

Covert Channel

- An observer cannot tell from observing the network if there is communication or not.
- A covert channel is hidden within the noise of a system or in legitimate normal communication and its normal patterns.
- Methods
  - Spread Spetrum Methods in Noisy Channels
  - Steganography
  - Hide in normal (preferably encrypted) communication.
  - …
- Discussion
  - Either extremely slow or statistical patterns uncover the channel.
  - Connecting to an anonymous system and hiding traffic patterns is not a covert channel.
  - A normal HTTP/HTTPS connection from Alice to Bob is also not a covert channel.

Adversary Models

Basic adversary characteristics

- Position
  - External: „sits" on the wire
  - Internal: participates in the anonymous system
- Geographic
  - Global: sits on all wires
  - Local: sits on some local wires
  - Partial: controls parts of the network
- Participation
  - Passive: only observes traffic
  - Active: may send, modify, and drop messages.

Adversary Models

Typical adversary models

- Global Passive Adversary (GPA)
  - Observes and efficiently analyses the complete network.
  - No active participation in the network.
  - External attacker.
- Global Active Adversary (GAA)
  - Also performs active attacks.
- Partial Passive Adversary (PPA)
  - Observes only parts (<< 50 %) of the network.
  - External attacker.
- PPA or GPA with some active nodes
  - Add some internal nodes that may also perform active attacks.
- Local observer
  - An observer that locally observes the endpoints of a communication.

→ *All of these attacker models are too strong for current realtime low-latency anonymous networks.*

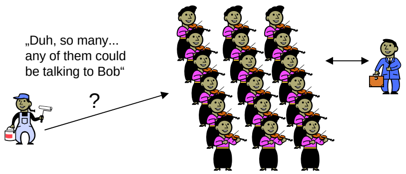
Measuring Anonymity

How anonymous is a systems?

- Number of known attacks?
- Lowest Complexity of successful attacks?
- Information leaked though messages and maintenance procedures?

Examples

- Anonymity set
  - Anonymity Set = |{suspects}|
  - Suspects are all entities that could have sent / received / participated.
  - In the example, the anonymity set is 18.
  - Limitations
    - No way to include meta knowledge.
      - An attacker could know that Alice is more likely to communicate with Bob than others because she is an attacker in a security lecture ;).

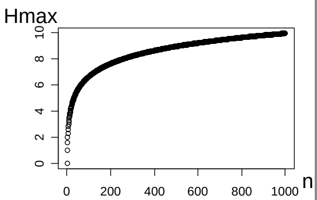


Measuring Anonymity

So, we are an attacker in a security lecture. For talking with Bob, we use this knowledge to conclude Alice 0.9 and other 100 suspect 0.001.

Any metric for that?

- Entropy
  - Combines the number of suspects and their probabilities in one metric.
  - Let p_i be the probability for suspect i.
  - Entropy $H = -\sum_i p_i ld(p_i)$
  - Entropy is maximized for a fixed number of suspects if all are equally likely (p_i=1/n for all i) → Hmax=ld(n)
  - e.g. 101 nodes as above Hmax = 6.7, if we use meta knowledge with probability p_alice=0.9 then H=1.1.

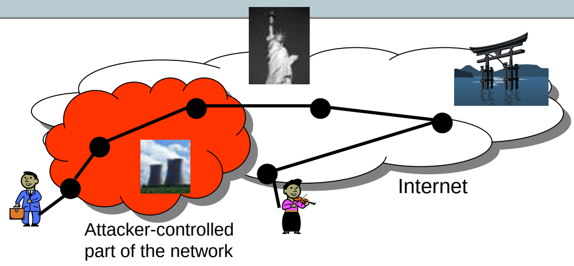


Basic concepts for anonymous systems

Basic concepts for anonymous systems

- Escape geographically (→ Re-Routing)
- Confuse packet flows at re-routers (→ Mixing)
- Hide content (→ Layered Encryption and Hop-by-Hop encryption)
- Hide message properties (→ Padding)
- Hide communication / flow properties (→ Dummy Traffic)

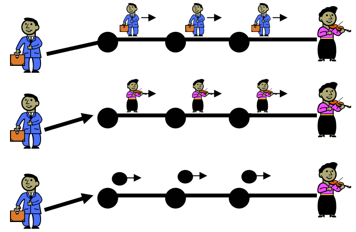
Re-Routing



Re-Routing

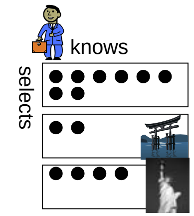
- Anonymity requires to hide sender/receiver relationships. As a direct message would be such a relationship, anonymity requires to route message via other intermediate nodes (*re-routers*).
- With respect to fighting an attacker, re-routing tries to get the message out of the area controlled by the attacker. The idea is to globally espace a partial attacker (*„escape geographically“*).
- Messages need to be encrypted.
  - Otherwise, attacker can simply read source/target locator.
  - Usually, re-encryption hop-by-hop. → Packet looks different on each path section.

Path Selection Strategies



Who selects?

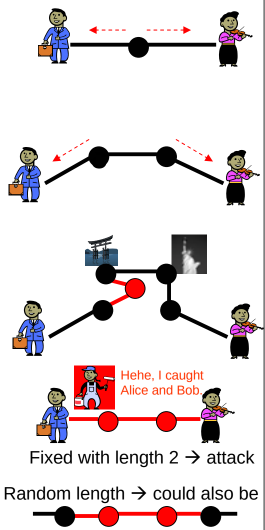
- Sender
  - The sender initiates a path hop-by-hop.
    → „Sender controls her anonymity“
- Receiver
  - The receiver initiates a path from some rendezvous point to herself hop-by-hop.
    → „Receiver controls her anonymity“
- Re-router
  - Each re-router selects the next hop for a path.
  - Problem: An internal attacker may select other attackers.
- Network design
  - The route is fixed by the system itself.



Selection

- Selection requires knowledge of large set of re-routers.
- Random selection provides most entropy.
- Biased selection strategies
  - Geographic diversity of used re-routers (→ Optimize trust, escape attacker geographically).
  - Organizational diversity of used re-routers (→ Optimize trust).

Path Length



1 Hop (simply proxy)

- Trust problem as proxy knows everything.
- Trusted proxy may leak meta-information about those who trust it.

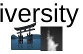
e.g. trust-proxy-tuebingen may imply „someone in Tübingen“ … hmm… only Bob is from Tübingen → Bob

2 Hops

- No hop knows sender and receiver.
- But each hop likely to know its position on path.

More hops

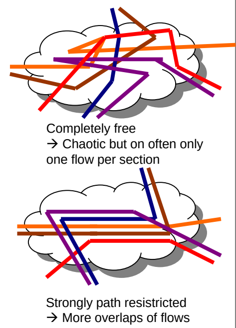
- Position on path for a re-router less clear.
- Better diversity / but more likely to select attacker.



Fixed length vs. random length

- Random length makes attacks based on positions in the path harder.

Re-routing

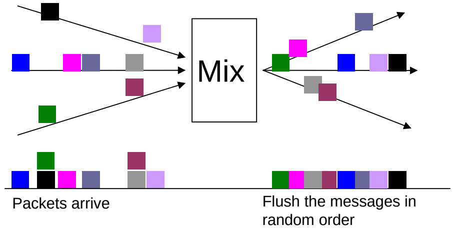


Other aspects

- Degree of freedom for path selection (Topology)
  - A high degree has advantages with respect to trust.
  - A low degree better hides communication properties as many flows follow identical paths.
- Lifetime of a path – fixed path vs dynamic path
  - Fixed path
    - Use same path for entire session.
    - \+ performance, overhead, no need to change good path
    - \- easier to observe for an attacker
  - Dynamic path
    - Change path frequently during session.
    - \+ makes (long-term) observations harder
    - \- with internal attackers, the more often a path is changed the more likely it is to hit a path solely consisting of attackers.

Mixing

How does a re-router operate?



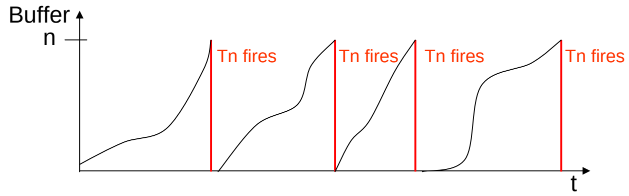
Assumption

- Packets change appearance -> re-encryption

Mix

- Concept by David Chaum (1981)
- A mix is a re-router that does not directly forward messages. A mix first collects a number of messages and then sends them out in random order.
- An attacker observing a mix cannot tell which incoming messages is which outgoing message („escape through re-ordering“).

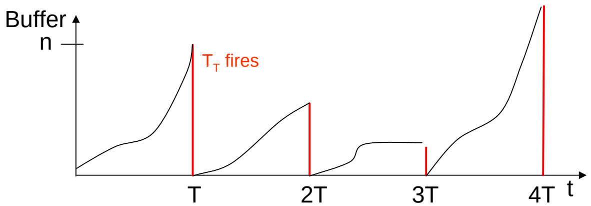
Threshold Mix



Threshold Mix

- A threshold mix T_n with threshold n.
- Operation
  - T_n collects messages until it buffers n messages.
  - Then it fires = T_n sends these n messages in random order.
- Anonymity Set = n.
- Performance depends on rate of incoming messages.

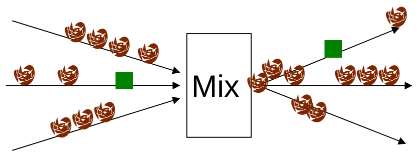
Timed Mix



Timed Mix

- A timed mix T_T with interval time T.
- Operation
  - T_T collects messages for time T.
  - Then it fires = T_T sends these messages in random order.
- Anonymity Set = number of messages that arrived in interval
  - Can be small (1 = no anonymity) or large („buffer capacity of mix“). → Anonymity depends on rate of incoming messages

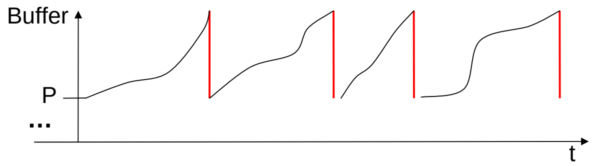
n-1 attack on mixes



n-1 attack on a mix

- An n-1 attack is an active attack.
- Basic idea
  - The attacker inserts messages and degrades the anonymity set.
- Attack situation
  - n messages arrived at mix
  - n-1 messages are from the attacker
- The mix fires.
  - Attacker knows its n-1 messages, can identify the other one.
- Basic form is against threshold mix, but a strong attacker could also delay messages towards a timed mix.

Pool Mix / Exponential Mix



Pool Mix

- Basic idea
  - To increase anonymity set and to make the n-1 attack more difficult, ensure that always a pool of P old messages is in the mix.
- Operation
  - Collect messages and fire at some point in time (threshold/timed/…).
  - With S messages in the buffer, randomly select S-P and send them in random order.

Exponential Mix

- Mix messages by randomly-delaying. No firing.
- Operation
  - Message Mt arrives at time t.
  - Add a random delay D (exponential distribution / geometric distribution) and schedule message for time t+D.
  - Send Mt at scheduled time t+D.

Discussion of Mixes

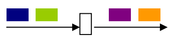
Discussion

- *When a message passes a set of mixes, one honest mix is enough to provide anonymity! (for the message)*
- Mixes protect single messages.
  - Flows with several messages may be identified due to their traffic volume.
- To ensure performance or a good anonymity set, a mix needs a lot of traffic.
  - Not suitable for decentralized approaches that opt for low-latency.
- The operation of a mix is targeted against a strong observer that controls all interfaces of a mix or all mixes in a mix network.
  - Maybe an overkill for overcoming realistic attackers in combination with the use of re-routing.
  - Most low-latency anonymity systems only re-route and do not mix.
- Re-routers with lots of traffic also slightly randomize order due to internal processing and queuing (despite FIFO and Round Robin).

Layered Encryption and Hop-by-Hop encryption

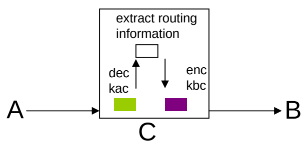
Goals

- Hide the content from observers.
- The outgoing message from a re-router should look different than the corresponding incoming message.



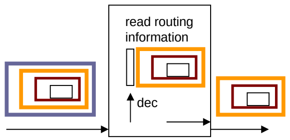
Hop-by-Hop encryption

- Each hop decrypts (key with predecessor) and re-encrypts (key with successor) message.
- End-to-end message confidentiality can be achieved by adding end-to-end encryption.
- Discussion
  - Re-routers see identical packets → internal attacker
  - Difficult to implement unless re-routers select paths.

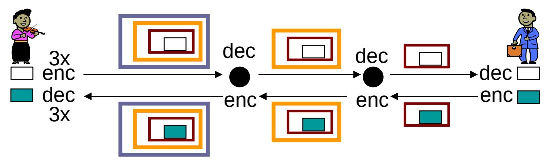


Layered encryption

- Sender encrypts message several times with keys for all hops. It adds a layer of encryption over the message for each hop.
- Either public key of re-router or an established shared key between sender and re-router.
- Re-routers decrypt the message to determine next hop and send the decrypted message.



Onion Routing



Onion Routing

- Onion Routing is based on layered-encryption.
- The term is a metaphor for the operation of such routers as the packets is peeled like an onion.
- Onion routers (ORs) do not mix or delay packets. They usually operate with simple FIFO or round robin (between flows) queues.
- Pad message to constant length at each hop.

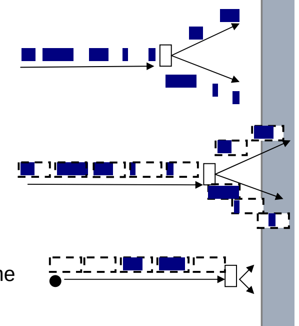
Keys

- Public keys of re-routers (not very efficient).
- Sender/Initiator uses public key of re-routers for path establishment and establish shared key with each re-router on the path.

Padding / Dummy Traffic

Padding

- Message size
  - can be used to fingerprint messages.
  - unveils information like positions in a path
- Message Padding
  - Add padding (random data) to smaller packets so that all packets are of identical size.
  - → Necessary and thus widely used in anonymity systems
- Link Padding
  - Use dummy messages to pad the link to a constant bandwidth.
  - → Necessary against global and local observers, used in some systems. Link padding is covering the existence of real traffic.



Dummy Traffic

- Send dummy traffic through the network to hide traffic volumes of flows and cover real traffic.
  - Link padding is a subclass of dummy traffic.
- Except for link padding, dummy traffic is hardly used in anonymity systems → usually considered too expensive for too little gain.

Basic structure of anonymity systems

Trust anchors

- Trust in software and at least some re-routers (at least 1 on path).
- Certificate Authorities or TTPs may certify or rate re-routers.
- Existance of several distinct authorities beneficial to avoid single points of trust.

Information

- Directory servers or discover service necessary.
- Anonymity set can be severely degraded when nodes only know small distinct fractions of re-routers.

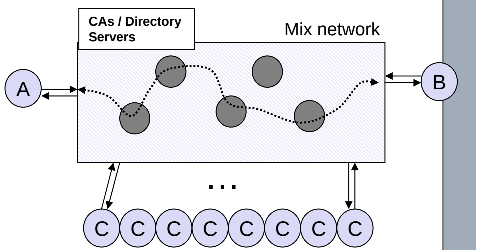
Services

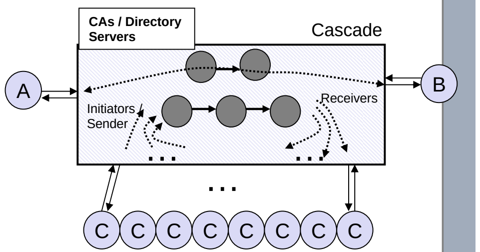
- Interval services
  - Some services may be provided within the system.
- Exit / Gateway nodes
  - Exit nodes are used to contact nodes outside the system, e.g. webservers.

Infrastructure-based (Mix net vs Cascade)

Infrastructure-based

- Distinction between clients consuming the server and re-routers.
- Re-routers are certified by one or more CAs (Certificate Authorities) for the system.
  → Trust
- Directory servers maintain lists of running re-routers.
- Mix network
  - Free or slightly restricted routes between re-routers. Path selected by clients.
- Mix cascade
  - Mixes form fixed cascades.
  - Client can only chose between cascades.
- Infrastructure can plausibly deny being responsible. Some approaches include revocation for prosecution.

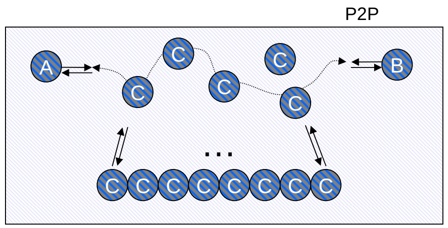




Peer-to-Peer-based

Peer-to-Peer-based anonymity system

- No distinction between re-router and client.
- Peers re-route traffic
  → need also for clients to plausibly deny actions of others.
- Path usually selected by clients.
- CA and directory server tasks either centralized or part of P2P algorithm.



Conclusion

Conclusion

- Encryption not always confidential….
- Anonymity, Pseudonymity, Covert Channel
- Adversary Models
- Anonymity Set, Entropy
- Concepts for anonymous communication
  - Escape geographically.
  - Confuse flows.
  - Hide properties of messages and flows.
- Distribute trust and information
- Mix cascade vs. Mix network vs. Peer-to-Peer