Anonymous Communication

COS 518: Advanced Computer Systems
Lecture 22
Michael Freedman

Slides based heavily on Christo Wilson’s CS4700/5700 at Northeastern

Definition

• Hiding identities of parties involved in communications from each other, or from third-parties
  – “Who you are” from the communicating party
  – “Who you are talking to” from everyone else

Quantifying Anonymity

• How can we calculate how anonymous we are?

- Larger anonymity set = stronger anonymity
Anonymity Systems

Crypto (SSL)
- Content is unobservable
  - Due to encryption
- Source and destination are trivially linkable
  - No anonymity!

Anonymizing Proxies
- Source is known
- Destination anonymity

Anonymizing VPNs
- Source is known
- Destination anonymity

HTTPS Proxy

VPN Gateway
Crowds

- Key idea
  - Users’ traffic blends into a crowd of users
  - Eavesdroppers and end-hosts don’t know which user originated what traffic

- High-level implementation
  - Every user runs a proxy on their system
  - When a message is received, select $x \in [0, 1]$
    - If $x > p$: forward the message to a random proxy
    - Else: deliver the message to the actual receiver

Crowds Example

- Links between users use public key crypto
- Users may appear on the path multiple times

Anonymity in Crowds

- No source anonymity
  - Target receives $m \geq 0$ msgs, sends $m+1$ msgs
  - Thus, target is sending something
- Destination anonymity is maintained
  - If the source isn’t sending directly to the receiver
Anonymity in Crowds

- Source and destination are anonymous
  - Source and destination are proxies
  - Destination is hidden by encryption

Anonymity in Crowds

- Destination known
  - Source is anonymous
    - $O(n)$ possible sources, where $n$ is the number of proxies

Anonymity in Crowds

- Destination is known
  - Evil proxy able to decrypt the message
- Source is somewhat anonymous
  - Suppose $f$ evil in system and if $p_f > 0.5$ and $n > 3(f + 1)$, source cannot be inferred with prob > 0.5

Summary of Crowds

- The good:
  - Crowds has excellent scalability
    - Each user helps forward messages and handle load
    - More users = better anonymity for everyone
  - Strong source anonymity guarantees

- The bad:
  - Very weak destination anonymity
    - Evil proxies can always see the destination
  - Weak unlinkability guarantees
**Mix Networks**

- A different approach to anonymity than Crowds
- Originally designed for anonymous email
  - David Chaum, 1981
  - Concept has since been generalized for TCP traffic
- Hugely influential ideas
  - Onion routing
  - Traffic mixing
  - Dummy traffic (a.k.a. cover traffic)

**Onion Routing**

- Mixes form a cascade of anonymous proxies
- All traffic is protected with layers of encryption

\[ E(K_P, E(K_P, E(K_P, M))) = C \]

**Another View of Encrypted Paths**
Return Traffic

- In a mix network, how can the destination respond to the sender?
- During path establishment, the sender places keys at each mix along the path
  - Data is re-encrypted as it travels the reverse path

Traffic Mixing

- Hinders timing attacks
  - Messages may be artificially delayed
  - Temporal correlation is warped
- Problems:
  - Requires lots of traffic
  - Adds latency to network flows

Applied to cryptographic voting

- Server collects votes
- Computes random shuffle of votes
- Outputs votes in randomized order
- Includes “proof” that correctly shuffled

Chain multiple MIXes for security

- Synchronously collects and shuffles messages (votes)
- Secure as long as at least 1 honest
Dummy / Cover Traffic

- Simple idea:
  - Send useless traffic to help obfuscate real traffic

In practice
Hard to be anonymous
Information leaked at many layers

Using Content to Deanonymize

- Login to email account
- Information sent in cookies
- Accessing Facebook pages

It’s Hard to be Anonymous!

- **Network location** (IP address) can be linked directly to you
  - ISPs store communications records (legally required for several years)
  - Law enforcement can subpoena these records
- **Application** is being tracked
  - Cookies, Flash cookies, E-Tags, HTML5 Storage, browser fingerprinting
  - Centralized services like Skype, Google voice
- **Activities** can be used to identify you
  - Unique websites and apps that you use, types of clicked links
  - Types of links that you click
You Have to Protect at All Layers!

Challenges:

- Maintain performance
- Provide functionality!

Wednesday’s reading

- Tor: 2nd generation onion routing (2004)