Privacy and Anonymity

Josh Wilcox (jw14g24)

March 18, 2025

Table of Contents

Confidentiality, Privacy, and Secrecy

- Keeping data **confidential** means determine the access control of who can access data
  - Actually more focussed on the aunauthorised **learning of information** - which is harder to prevent
- Privacy - Confidentiality for individuals
  - Used in the sense of anonymity
- Secrecy - Confidentiality for organisations

Solove's Taxonomy of Privacy

Solove's Taxonomy of privacy defines four basic groups of **harmful activities**

- **Information Collection**: Activities that gather information about individuals
  - *Surveillance*: Watching, listening to, or recording individuals
  - *Interrogation*: Questioning or probing for information
- **Information Processing**: Activities that use, store, and manipulate data
  - *Aggregation*: Combining various pieces of data about a person
  - *Identification*: Linking information to particular individuals
    - Involves aggregation
    - Alters what others learn about people
  - *Insecurity*: Careless protection of stored information
  - *Secondary use*: Using information for purposes other than collected for
  - *Exclusion*: Failing to allow people to know about or control their data
- **Information Dissemination**: Activities that spread or transfer information
  - *Breach of confidentiality*: Breaking promises to keep personal information confidential
  - *Disclosure*: Revealing truthful information that impacts reputation
  - *Exposure*: Revealing nudity, grief, or bodily functions
  - *Appropriation*: Using identity for the purposes and benefits of another
  - *Distortion*: Disseminating false or misleading information
- **Invasion**: Direct interferences with the individual
  - *Intrusion*: Invasive acts that disturb one's solitude
  - *Decisional interference*: Incursion into people's decisions regarding private affairs

Privacy Enhancing Technologies

Privacy Enhancing Technologies (PETs) are tools, mechanisms, or architectures that aim to mitigate privacy concerns.

- **Three main privacy research paradigms**:
  - **Privacy as Confidentiality**
    - Data anonymisation/minimisation techniques
    - Secure and anonymous communication protocols
  - **Privacy as Control**
    - Anonymous credentials and authentication systems
    - Privacy policy enforcement mechanisms
    - Purpose-based access control frameworks
    - Compliance verification tools
  - **Privacy as Practice**
    - Feedback and awareness tools for privacy management
    - Privacy nudges to guide better privacy decisions
    - Decision support systems for privacy choices
- **Key Privacy Enhancing Technologies**:
  - **Communication Anonymisers**
    - Replace real identifiers (email, IP) with non-traceable alternatives
    - Examples: Tor network, VPNs, anonymous remailers
  - **Enhanced Privacy ID (EPID)**
    - Digital signature algorithm supporting anonymity
    - Uses group public verification keys with unique private signature keys
    - Provides device authentication without revealing identity
  - **Zero-knowledge Proofs**
    - Allows proving knowledge of information without revealing it
    - Applications in authentication, blockchain privacy, and credential systems
  - **Homomorphic Encryption**
    - Enables computation on encrypted data without decryption
    - Supports privacy-preserving data analysis and outsourced computation
  - **Secure Multi-Party Computation**
    - Joint computation across multiple parties without sharing inputs
    - Enables privacy-preserving analytics and collaborative computation
  - **Differential Privacy**
    - Mathematical framework for sharing dataset insights while protecting individuals
    - Adds calibrated noise to query results for privacy guarantees
  - **Federated Learning**
    - Trains machine learning models across distributed devices without centralizing data
    - Keeps sensitive data on local devices while improving global models

Virtual Private Networks (VPN)

Proxies

- A proxy server acts as an intermediary between a client and destination server
- **How it works**:
  - Client sends requests to the proxy instead of directly to destination
  - Proxy forwards requests to destination and returns responses to client
  - Destination server sees the proxy's IP address, not the client's
- **Privacy implications**:
  - Hides client's IP address from destination servers
  - The proxy server has visibility into all traffic passing through it
  - Your ISP/network can still see that you're connecting to the proxy
  - Does not provide encryption by default (unlike VPNs)
- **Types of proxies**:
  - HTTP proxies: Web traffic only
  - SOCKS proxies: Support various protocols
  - Transparent, anonymous, and elite proxies (varying levels of anonymity)

How VPNs are better

- **VPN Architecture**:
  - Creates an encrypted tunnel between client device and VPN server
  - All internet traffic is routed through this secure tunnel
  - Provides both anonymity AND encryption (unlike basic proxies)
- **Encapsulation Process**:
  - Original data packet is encrypted using VPN protocols
  - Encrypted packet is encapsulated within another packet (tunneling)
  - Outer packet contains routing information to VPN server
  - VPN server decapsulates and forwards to destination
- **Key VPN Protocols**:
  - OpenVPN: Open-source, highly secure, uses SSL/TLS
  - WireGuard: Modern, faster, simpler codebase
  - IPsec: Works at network layer, commonly used with L2TP
  - SSTP: Microsoft protocol using SSL over TCP port 443
- **Advantages over Proxies**:
  - **Full traffic encryption**: All application traffic is protected
  - **OS-level integration**: Routes all traffic, not just browser/specific apps
  - **Authentication**: Uses strong authentication methods
  - **Persistent connection**: Maintains consistent secure tunnel
  - **Protection from ISP surveillance**: ISP can't see content or destinations
- **Limitations**:
  - VPN provider can still monitor your traffic
  - Speed reduction due to encryption overhead
  - Some services block known VPN IP addresses
  - Not all VPNs have strong no-logging policies

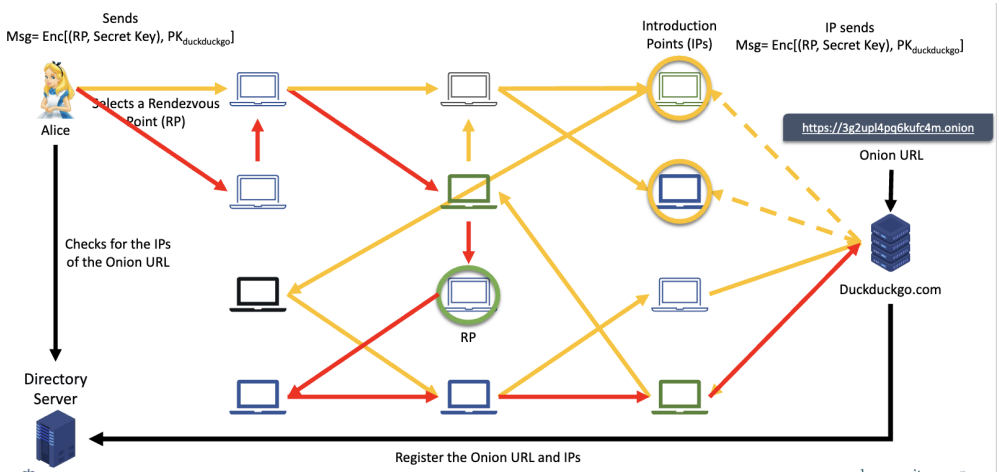
Onion Routing Method

- Forms a cascade of anonymous proxies - relay servers
- Between each proxy, traffic is encrypted each time - making **layers** of encryption (hence the onion)
- Traffic passes through 3+ nodes
- Types of nodes:
  - Entry Nodes
    - Know identity of sender
  - Relay Nodes
    - Routes and adds layer of encryption to message
  - Exit Nodes
    - Knows receiver identity and can see unencrypted traffic
- **Directory Servers** Maintain the type of each node to ensure sufficient encryption

Two Way Anonymity in Tor

5 Onion Routing Method

- Two Way Anonymity in Tor



- **Hidden Services Overview**:
  - Hidden services allow both clients and servers to remain anonymous
  - Uses special .onion addresses (not standard DNS)
  - Neither party knows the other's real IP address
  - Creates complete end-to-end anonymity in communications
- **Key Components**:
  - **Introduction Points (IPs)**: Relay nodes that know the hidden service
  - **Rendezvous Points (RPs)**: Meeting points for anonymous communication
  - **Directory Servers**: Maintain information about hidden services
- **Hidden Service Setup Process**:
  - Service generates a public-private key pair
  - Service selects Introduction Points and builds circuits to them
  - Service registers its .onion address and Introduction Points on Directory Servers
  - .onion address is derived from the service's public key
- **Connection Process**:
  1. Client obtains .onion address from Directory Server
  2. Client selects a Rendezvous Point in the Tor network
  3. Client creates a circuit to an Introduction Point with an encrypted message containing:
     - The Rendezvous Point address
     - A one-time secret key
  4. Hidden service retrieves the client request via its Introduction Points
  5. Hidden service establishes a separate circuit to the designated Rendezvous Point
  6. Both parties communicate through the Rendezvous Point using encrypted channels
- **Security Properties**:
  - **Client Anonymity**: Client's IP hidden behind multiple relays
  - **Service Anonymity**: Server's location concealed by Introduction Points
  - **Mutual Authentication**: Both parties can verify they're communicating with intended recipient
  - **End-to-end Encryption**: Complete privacy of communication content
- **Practical Applications**:
  - Secure communication platforms (e.g., SecureDrop for whistleblowers)
  - Private search engines (e.g., DuckDuckGo's .onion service)
  - Anonymous publishing and content hosting
  - Protection for individuals in high-risk situations (journalists, activists)