anonymous routing and mix nets (Tor)

Yongdae Kim

Significant fraction of these slides are borrowed from CS155 at Stanford

Anonymous web browsing

- Why?
  1. Discuss health issues or financial matters anonymously
  2. Bypass Internet censorship in parts of the world
  3. Conceal interaction with gambling sites
  4. Law enforcement

- Two goals:
  - Hide user identity from target web site: (1), (4)
  - Hide browsing pattern from employer or ISP: (2), (3)

- Stronger goal: mutual anonymity (e.g. remailers)

Current state of the world I

- ISPs tracking customer browsing habits:
  - Sell information to advertisers
  - Embed targeted ads in web pages (1.3%)
    - » Example: MetroFi (free wireless)

      [Web Tripwires: Reis et al. 2008]
- Several technologies used for tracking at ISP:
  - NebuAd, Phorm, Front Porch
  - Bring together advertisers, publishers, and ISPs
    - » At ISP: inject targeted ads into non-SSL pages
- Tracking technologies at enterprise networks:
  - Vontu (symantec), Tablus (RSA), Vericept

Current state of the world II

- EU directive 2006/24/EC: 3 year data retention
  - For ALL traffic, requires EU ISPs to record:
    - » Sufficient information to identify endpoints
      (both legal entities and natural persons)
    - » Session duration
      - … but not session contents
  - Make available to law enforcement
    - » … but penalties for transfer or other access to data

- For info on US privacy on the net:
  - "privacy on the line" by W. Diffie and S. Landau

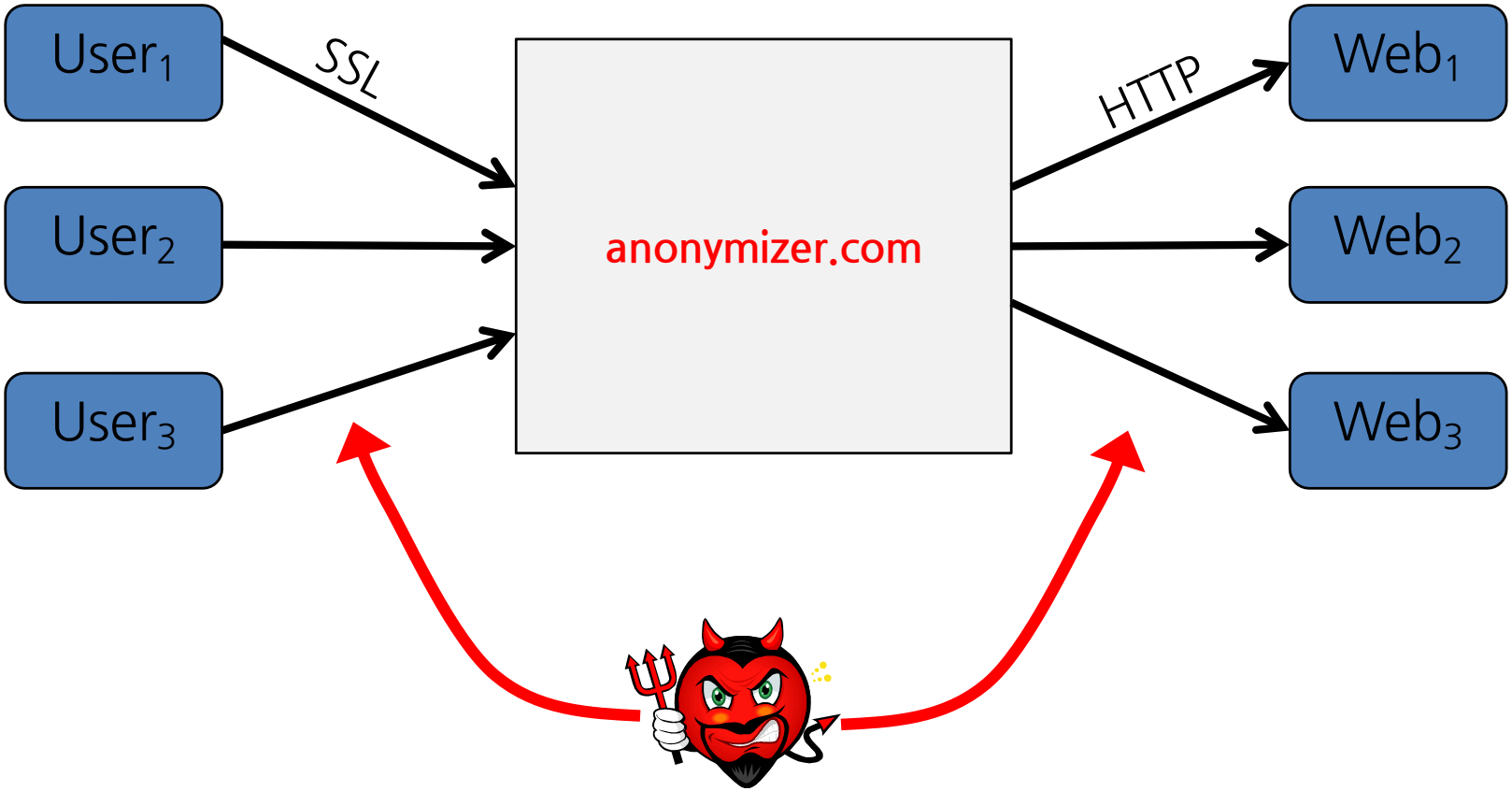
Part 1: network-layer privacy

Goals:

- Hide user's **IP address** from target web site
- Hide browsing destinations from network

1st attempt: anonymizing proxy

HTTPS:// anonymizer.com ? URL=target



Anonymizing proxy: security

- Monitoring ONE link: eavesdropper gets nothing
- Monitoring TWO links:
  - Eavesdropper can do traffic analysis
  - More difficult if lots of traffic through proxy

- Trust: proxy is a single point of failure
  - Can be corrupt or subpoenaed
    - » Example: The Church of Scientology vs. anon.penet.fi

- Protocol issues:
  - Long-lived cookies make connections to site linkable

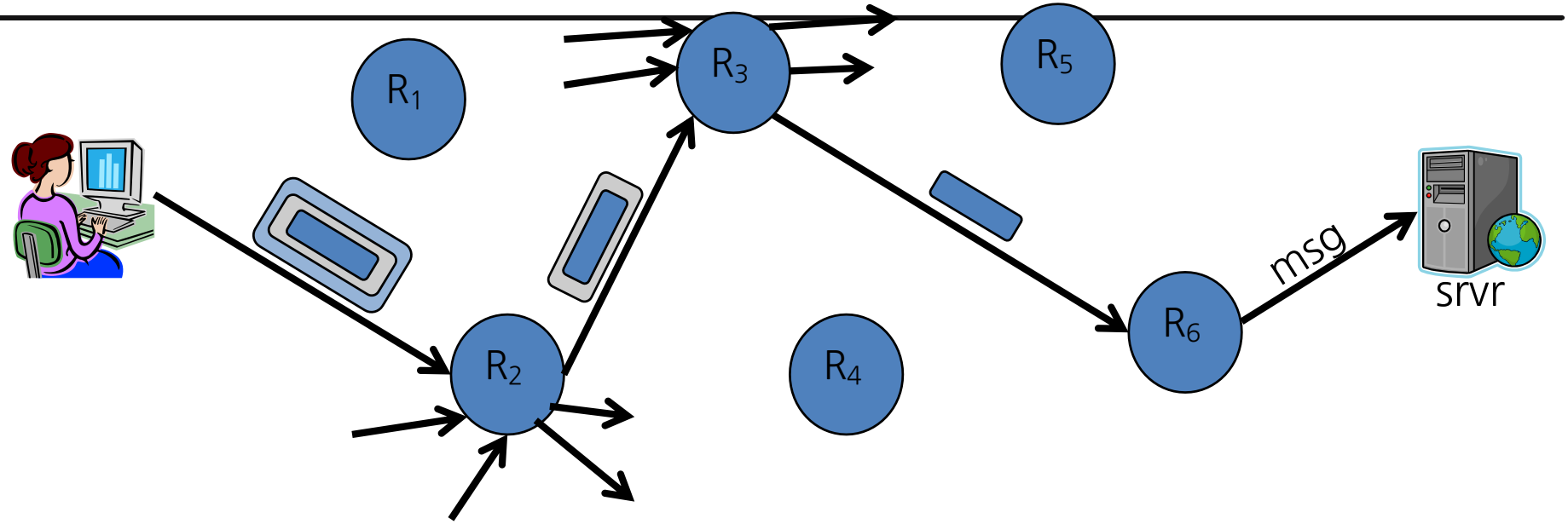
How proxy works

- Proxy rewrites all links in response from web site
  - Updated links point to anonymizer.com
    - » Ensures all subsequent clicks are anonymized
- Proxy rewrites/removes cookies and some HTTP headers
- Proxy IP address:
  - if a single address, could be blocked by site or ISP
  - anonymizer.com consists of >20,000 addresses
    - » Globally distributed, registered to multiple domains
    - » Note: chinese firewall blocks ALL anonymizer.com addresses
- Other issues: attacks (click fraud) through proxy

2nd Attempt: MIX nets

Goal: no single point of failure

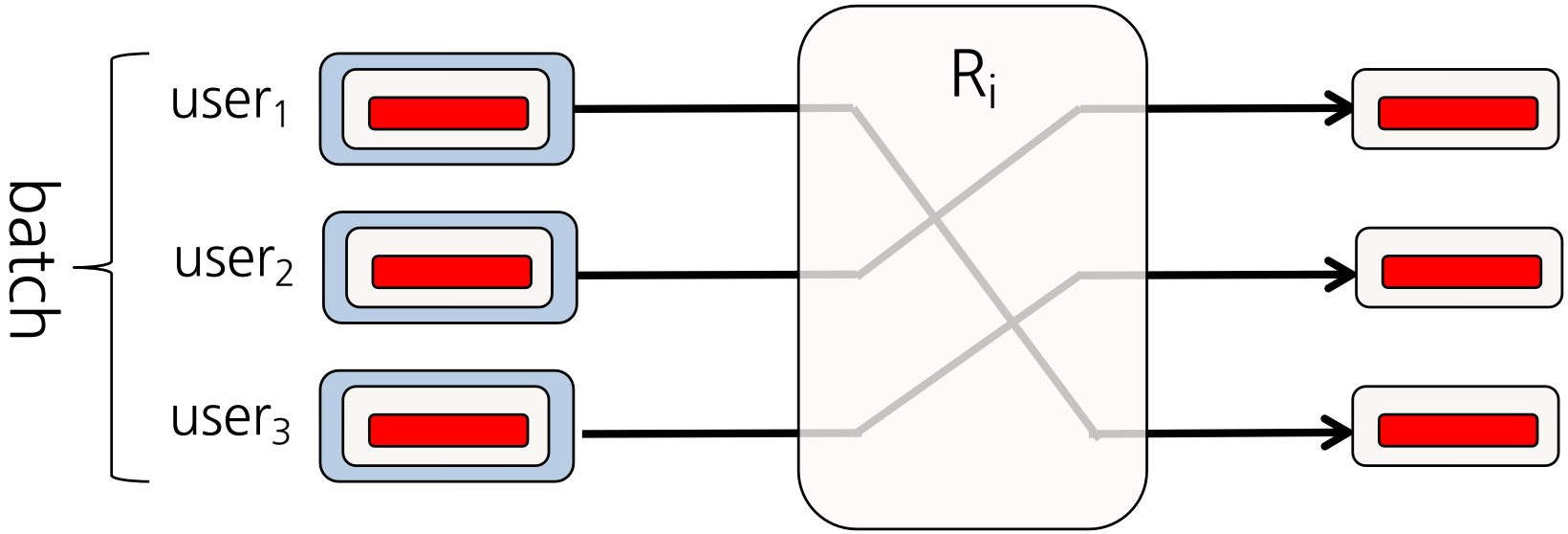
MIX nets [Chaum'81]



- Every router has public/private key pair
  - Sender knows all public keys
- To send packet:
  - Pick random route: R2 ➔ R3 ➔ R6 ➔ srvr
  - Onion packet:

$E_{pk_2}(R_3, E_{pk_3}(R_6, E_{pk_6}(srvr, msg)))$

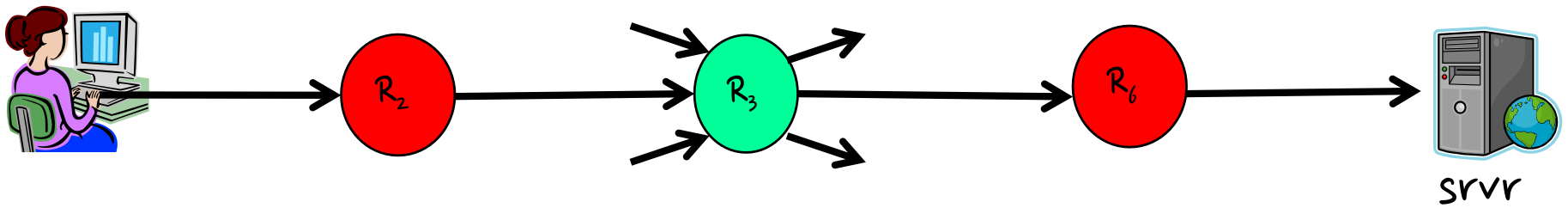
Eavesdropper's view at a single MIX



- Eavesdropper observes incoming and outgoing traffic
- Crypto prevents linking input/output pairs
  - Assuming enough packets in incoming batch
  - If variable length packets then must pad all to max len
- Note: router is stateless

Performance

- Main benefit:
  - Privacy as long as at least one honest router on path



- Problems:
  - High latency (lots of public key ops)
    - » Inappropriate for interactive sessions
    - » May be OK for email (e.g. Babel system)
  - No forward security

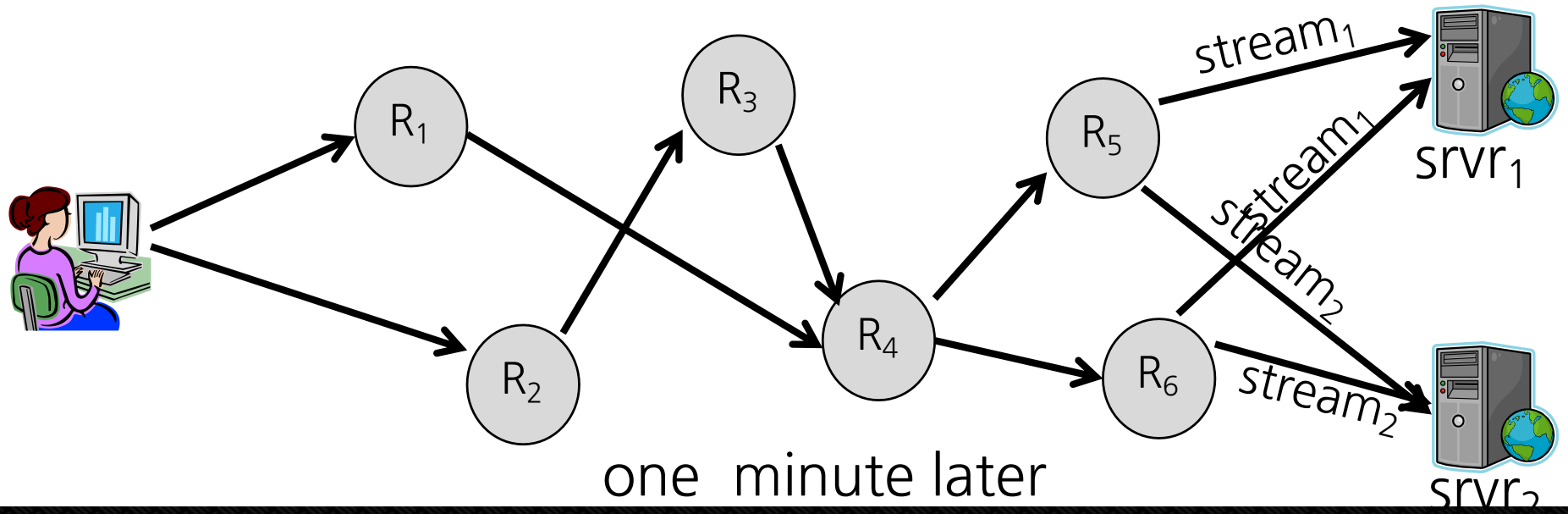
3rd Attempt: Tor MIX
circuit-based method

Goals: privacy as long as one honest router on path,

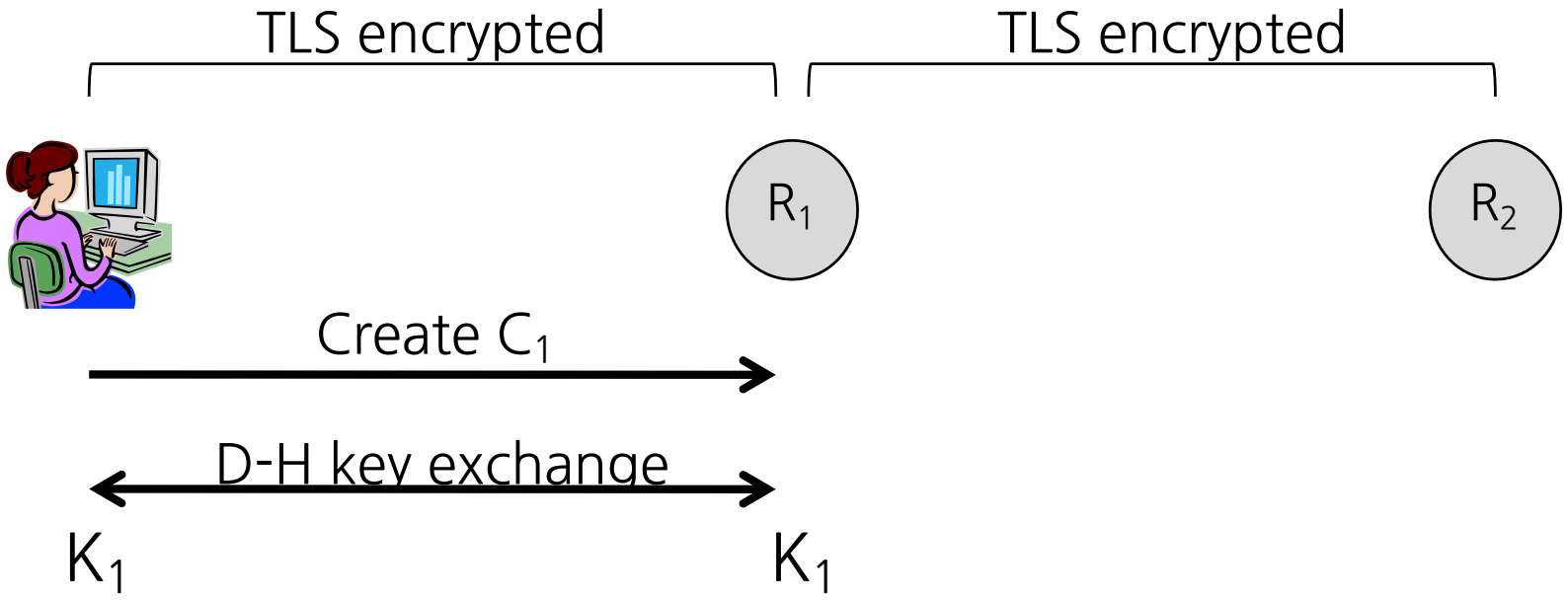
and reasonable performance

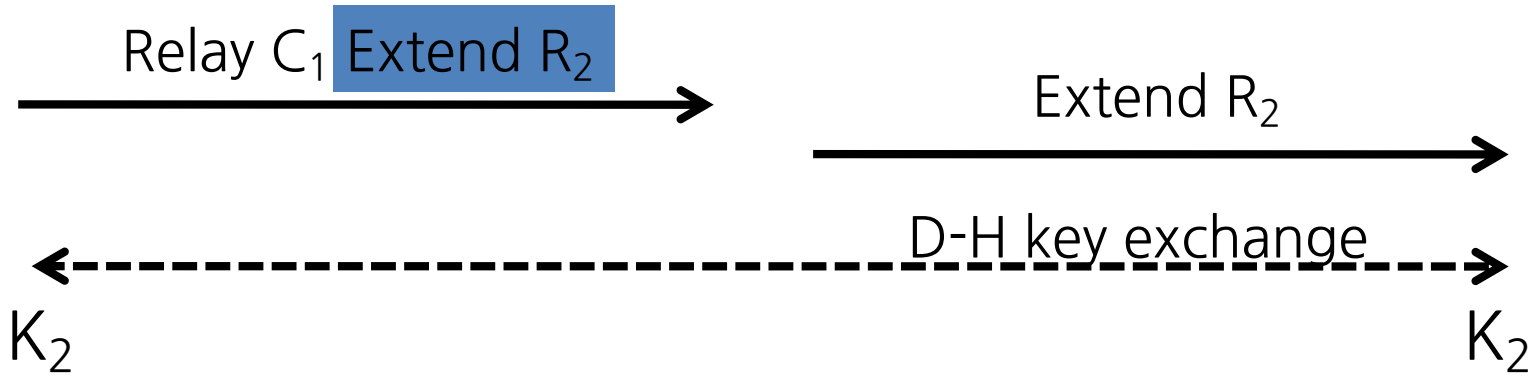
The Tor design

- Trusted directory contains list of Tor routers
- User's machine preemptively creates a circuit
  - Used for many TCP streams
  - New circuit is created once a minute

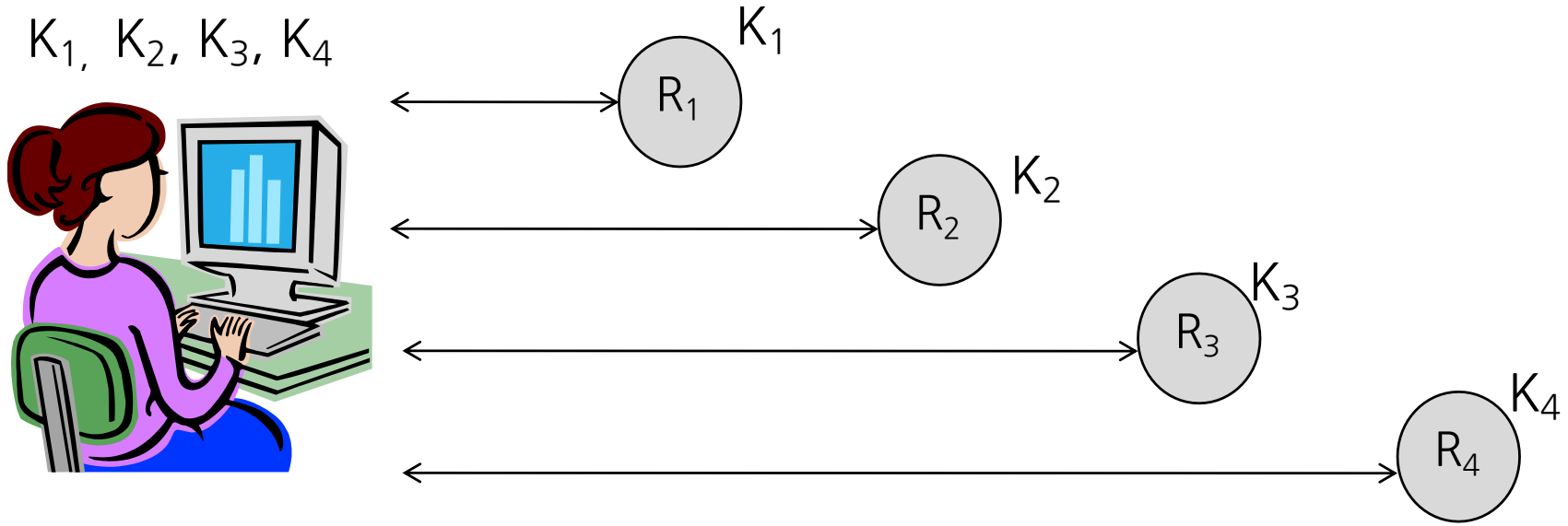


Creating circuits



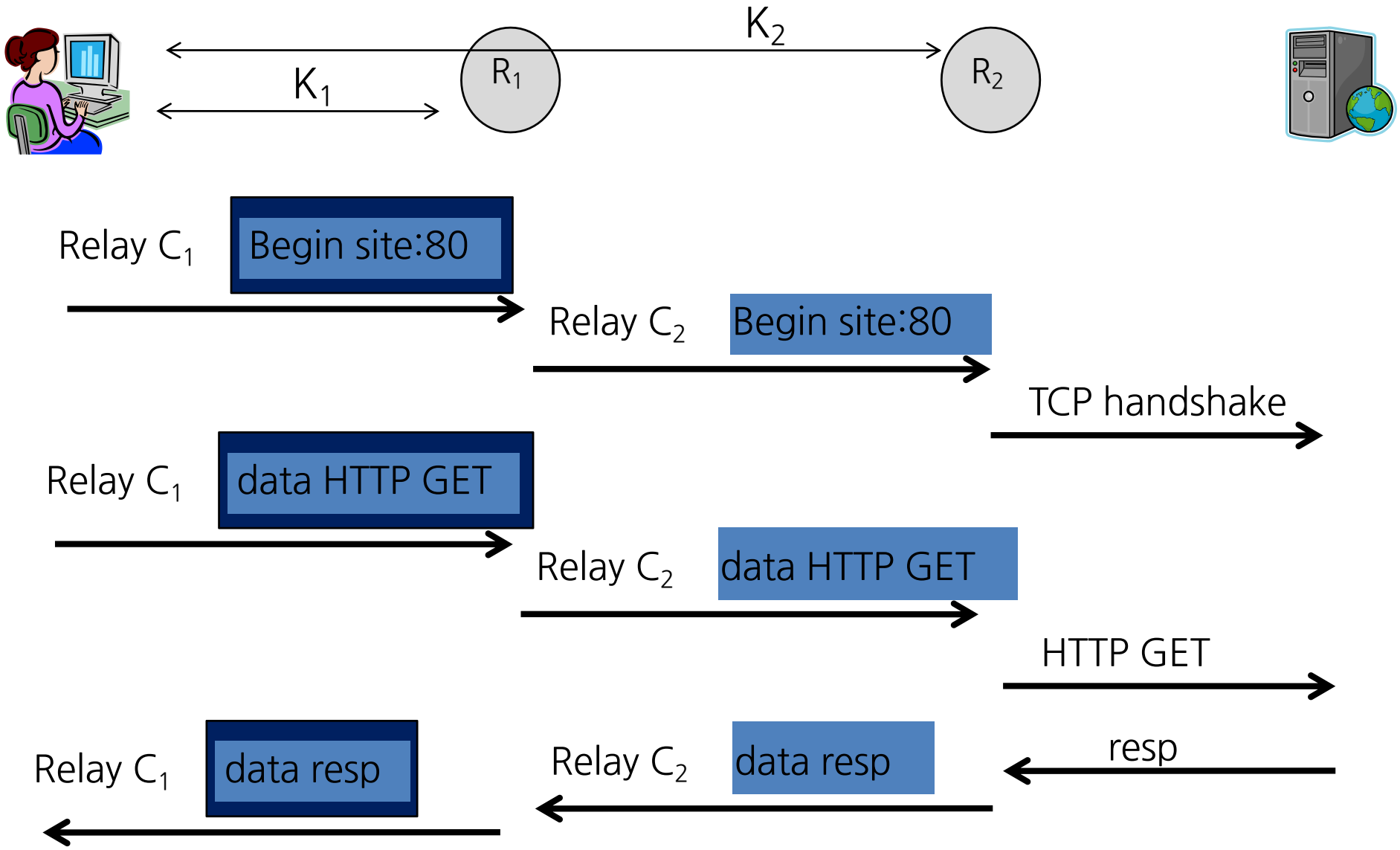


Once circuit is created

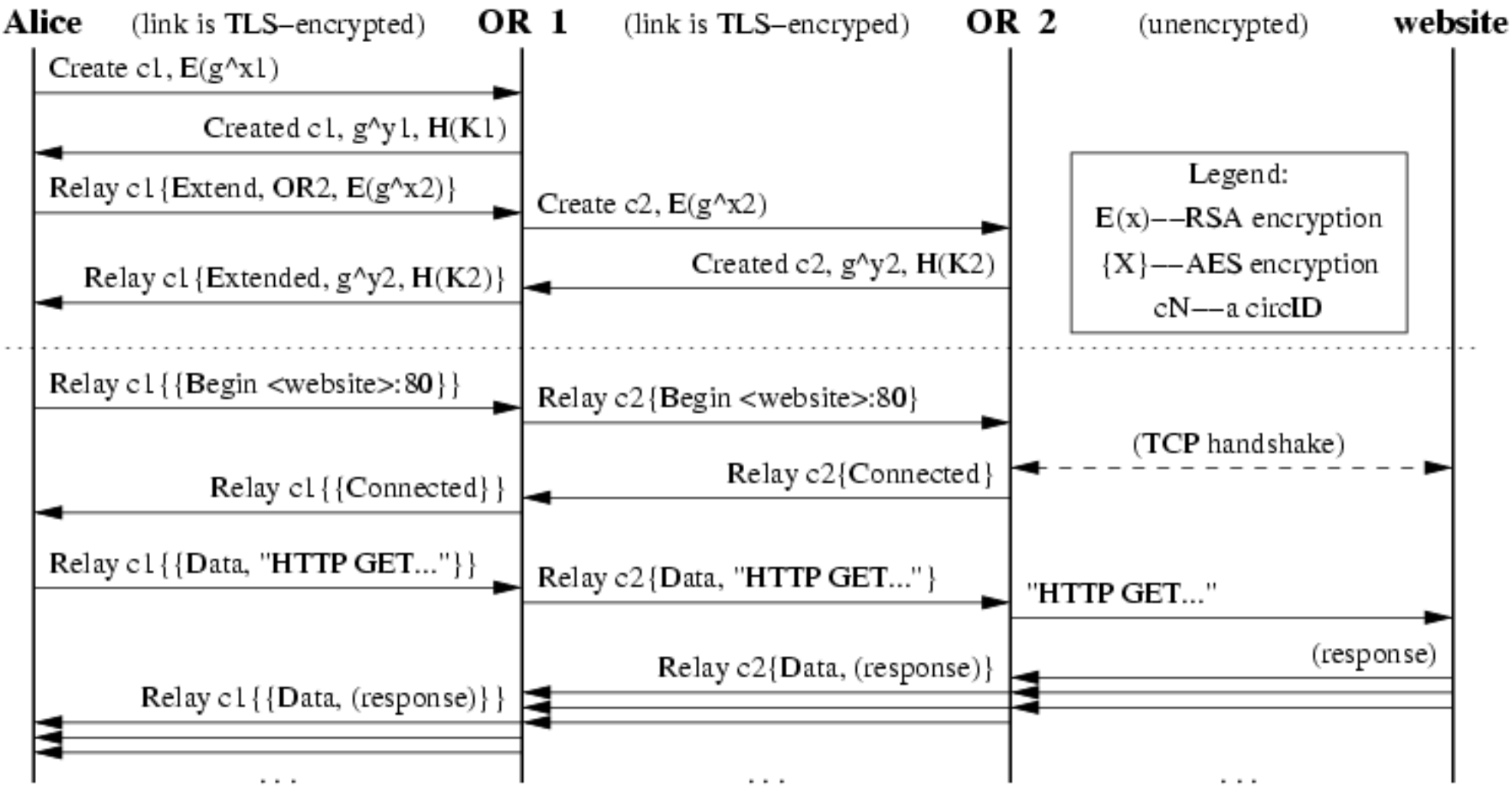


- User has shared key with each router in circuit
- Routers only know ID of successor and predecessor

Sending Data



Complete View



Properties

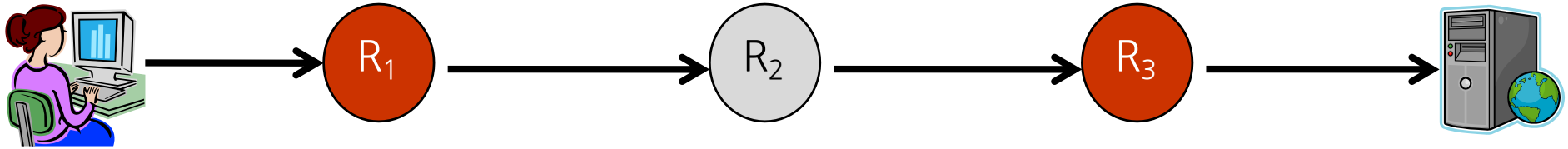
- Performance:
  - Fast connection time: circuit is pre-established
  - Traffic encrypted with AES: no pub-key on traffic
- Tor crypto:
  - provides end-to-end integrity for traffic
  - Forward secrecy via TLS
- Downside:
  - Routers must maintain state per circuit
  - Each router can link multiple streams via CircuitID
    - » all steams in one minute interval share same CircuitID

Privoxy

- Tor only provides network level privacy
  - No application-level privacy
    - » e.g. mail progs add “From: email-addr” to outgoing mail

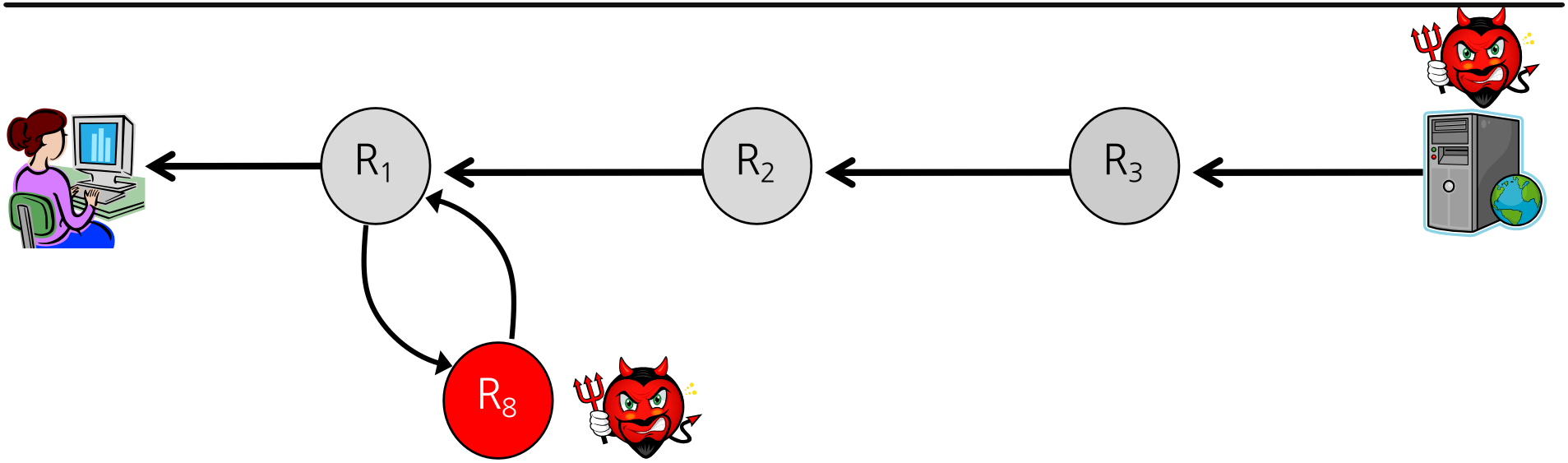
- Privoxy:
  - Web proxy for browser-level privacy
  - Removes/modifies cookies
  - Other web page filtering

Anonymity attacks: watermarking



- Goal: R_1 and R_3 want to test if user is communicating with server
- Basic idea:
  - R_1 and R_3 share sequence: $\Delta_1, \Delta_2, \cdots, \Delta_n \in \{-10, \cdots, 10\}$
  - R_1: introduce inter-packet delay to packets leaving R_1 and bound for R_2. Packet i delayed by Δ_i (ms)
  - Detect signal at R_3

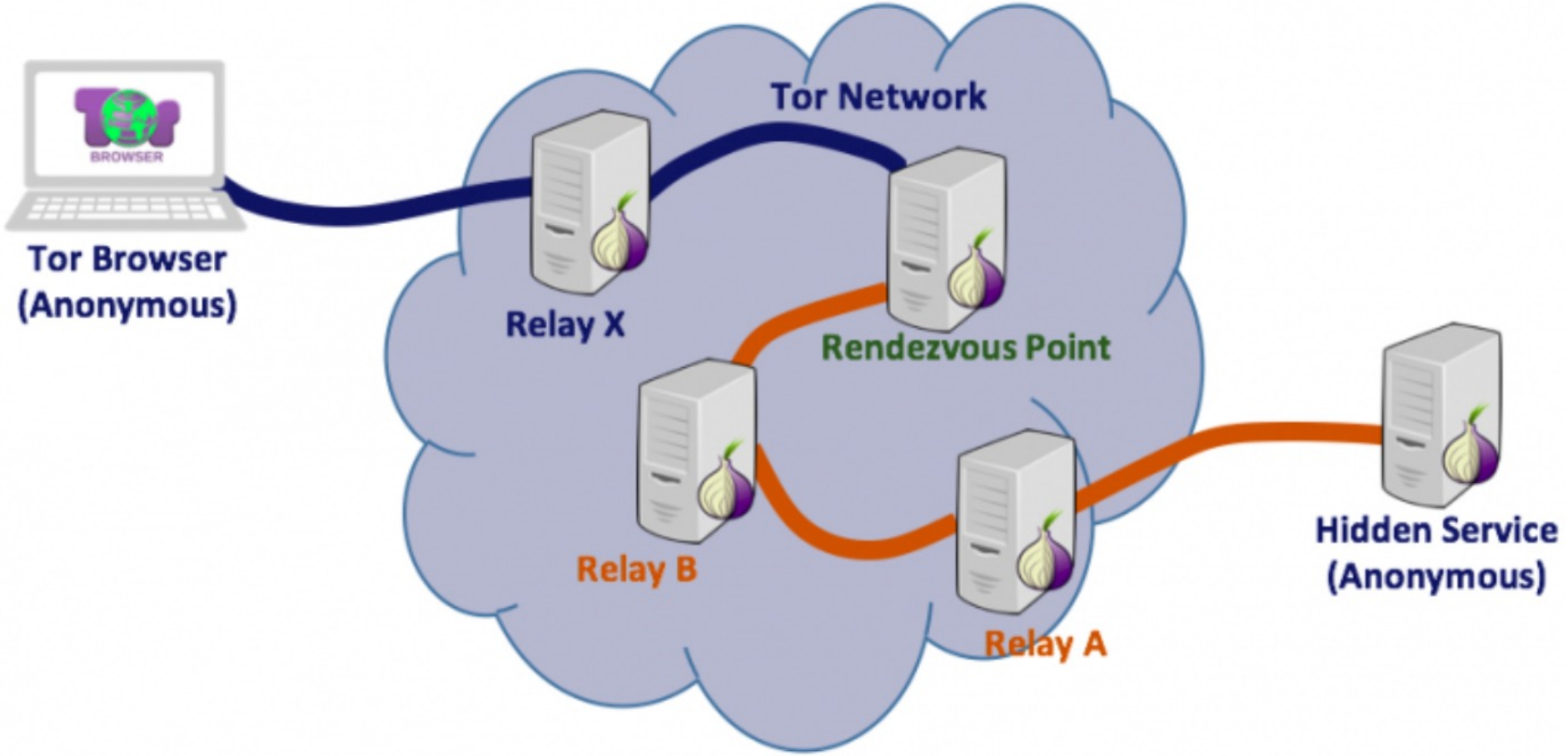
Anonymity attacks: congestion



- Main idea: R_8 can send Tor traffic to R_1 and measure load on R_1
- Exploit: malicious server wants to identify user
  - Server sends burst of packets to user every 10 seconds
  - R_8 identifies when bursts are received at R_1

    Follow packets from R_1 to discover user's ID

Tor Hidden Service



https://www.torproject.org/docs/hidden-services.html.en