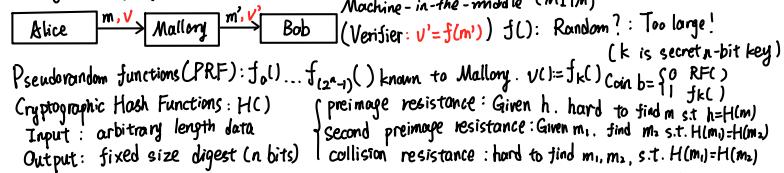


**Message Integrity:** ensure attackers cannot modify message without being detected



SHA-256 (Output 256 bit)

Input: arbitrary length data  
Output: 256-bit digest

Built from a compression function  $h$ :

- Inputs: (256 bits, 512 bits), Output: 256 bits  
"compresses" 768 bits to 256 bits using a complicated function (details out of scope for 388)



Pseudorandom Generator (PRG):  $g_k: 1 \rightarrow \{0,1\}^n$  for  $n = \text{poly}(1/k)$  ( $k$  is a truly random seed)

Build PRG from PRF: Given random key  $k$ , PRF  $f(\cdot)$ ,  $g_k := f_k(0) // f_k(1) // f_k(2) \dots$

Randomness is an input to your program! Provided by OS, via special APIs

Confidentiality: keep the content of message  $p$  secret from an eavesdropper ( $\$^{\text{off}}$ )  $k_{ABC}, k_{1-2}, \dots$

Alice  $[c := E_k(p)] \xrightarrow{\text{Eve}} \text{Bob} [p := D_k(c)]$  Caesar Cipher:  $(C_i + k_i) \bmod 26$  Vigenère:  $(P_i + K_{i \bmod n}) \bmod 26$

One-Time Pad (OTP): share a secret, long string of random bits  $k$   $C_i = P_i \oplus k$   $P_i = C_i \oplus k$  (In practical)

Stream Cipher (use PRG): share  $k$ ,  $S := g_k(\cdot)$   $C_i = P_i \oplus S_i$

Block Cipher: function that encrypts fixed-size( $n$ ) blocks with reusable  $k$ :  $E_k(p) := (p_0 \| p_1 \| \dots \| p_{n-1}) \rightarrow (p_0 \| p_1 \| \dots \| p_{n-1})$

AES (Advanced Encryption Standard) Block Cipher: (block size 128, key size: 128, 192, 256)

Padding: add bytes to end of message to make it a multiple of block size (PKCS7: M M M M M 0 0 0 0)

Cipher modes: 1. Encrypted codebook (ECB) mode  $C_i = E_k(p_i)$  (preserve pattern, replay)

### Cipher-block chaining (CBC) mode

"Chains" ciphertexts to obscure later ones

Choose a random initialization vector IV

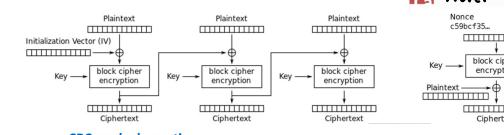
Encrypt:  $c_0 := IV; c_i := E_k(p_i \oplus c_{i-1})$

Decrypt:  $p_i := D_k(c_i) \oplus c_{i-1}$

[Why do we need the IV?]

Have to send IV with ciphertext

Can't encrypt blocks in parallel or out of order



**Malleable:** can transform ciphertext into another that decrypts to related plaintext without knowing plaintext.

### MAC-then-Encrypt



Note: Use separate keys for different purposes.  
(Can derive from a single key using PRG.) [Why?]

[Which approach is safest?] Our encryption methods (so far) only secure against passive eavesdroppers. Only EtM can ensure ciphertext isn't tampered with before decryption.

Common flaw when using MAC-then-Encrypt

Suppose an implementation uses CBC mode.

Decryption involves the following steps:

- $m // v // \text{pad} := D_k(c)$  Ciphertext  $c$  + Decryption  $\rightarrow k$
- Check that pad is valid PKCS7, else raise PadError
- Check that  $v = MAC_k(m)$ , else raise MacError

This is how TLS 1.0 worked. Seems reasonable?

Any method to distinguish these two error types (even tiny timing differences) leaks the plaintext!

**Padding oracle:** attacker submits any ciphertext and learns if last bytes of plaintext are a valid pad

Preferred modern approach:

### Authenticated encryption with associated data (AEAD)

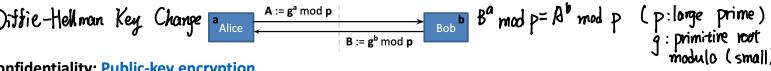
ChaCha20-Poly1305

Integrity and encryption in a single primitive:

$c, v := \text{Seal}(k, p, [\text{associated\_data}])$

encrypts plaintext  $p$  and returns ciphertext  $c$  and a verifier  $v$  (called a "tag")

$p, \text{err} := \text{Unseal}(k, c, v, [\text{associated\_data}])$  returns  $p$  or an error if  $v$  does not match the supplied  $c$  and  $\text{associated\_data}$



Confidentiality: Public-key encryption

Alice encrypts  $m$  using Bob's public key to get  $c$

Bob decrypts  $c$  using Bob's private key to get  $m$

Integrity/sender authenticity: Digital signatures

Alice signs  $m$  using Alice's private key to get  $s$

Bob verifies  $m, s$  using Alice's public key

Both properties: Use both, with two key pairs

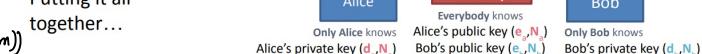
Alice encrypts  $m$  using Bob's public key to get  $c$ , then Alice signs  $c$  using Alice's private key to get  $s$

Bob verifies  $c, s$  using Alice's public key, then Bob decrypts  $c$  using Bob's private key to get  $m$

This is called "hybrid" encryption. Advantages:

- Identical messages yield different ciphertexts
- Don't have to worry about RSA padding
- Don't have to worry about message length

In practice, almost always should use



Safely encrypting with RSA

Encrypt a message  $m$  to RSA public key  $(e, N)$ :  $x := \text{PSS}(m)$

Use RSA to encrypt a random  $x \in N$  use a KDF to derive a key from  $x$ , then encrypt message using a symmetric cipher and key  $k$ .

Verifier checks padding:  $x := \text{SHA-256}(m)$  w/ Bleichenbacher attack

$x := s^d \bmod N$  You'll exploit in Project 1!

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

4.  $x := \text{AES-GCM}(m)$

This is called "hybrid" encryption. Advantages:

- Identical messages yield different ciphertexts
- Don't have to worry about RSA padding
- Don't have to worry about message length

In practice, almost always should use

The web platform is the collection of tech developed as open standards that powers web sites/applications

protect users from malicious sites and networks integrity: affect session confidentiality: steal data

scheme://host:port/path?query# HTTP: client send : method, path & query, headers

fragment: server return: response code, headers, content data

Cookie: maintain session state clients can read/change/erase cookie

Same-Origin Policy (SOP): origin is defined by scheme://domain:port

Protect: cookies, DOM storage & tree, JS namespace, permission

Cookie Security Attributes

Weakness: By default, cookies set over HTTPS will also be sent on requests over HTTP (and visible to network eavesdroppers).

Solution: Server can set "Secure" attribute to limit cookie to HTTPS requests.

Set-Cookie: id=a3fwa; Secure; **Secure**

Weakness: By default, cookies can be read by any JavaScript running in the origin. E.g., if bank.com includes Google Analytics script, Google can read authentication cookies.

Solution: Server can set "HttpOnly" attribute to prevent cookie from being accessed by DOM

Set-Cookie: id=a3fwa; Secure; **HttpOnly**

parameterized SQL

sql = 'SELECT \* FROM users WHERE username = ?'; cursor.execute(sql, [zhli])

hang up

TCP: Plaintext transport protocol, Provide semantics: dial, send/receive data

Doesn't Provide: confidentiality, integrity, authenticity

TLS (Transport Layer Security) is a cryptographic protocol that is layered above TCP to provide a secure channel. HTTP over TLS  $\Rightarrow$  HTTPS

Network Adversaries  $\begin{cases} \text{Passive: only eavesdrops} \\ \text{Active: can see, inject, modify, or block} \end{cases} \rightarrow$  provide confidentiality & integrity

How does client obtain server's public key?

Server presents a certificate: a message asserting the server's identity and its public key, signed by a certificate authority (CA).

A CA is an entity trusted by clients to verify server identities and issue certificates.

Each major platform and browser includes a set of public keys for the CAs that it trusts, called root CAs. Clients use these keys to verify certificates.

TLS uses the X.509 certificate format (specifies data structure, encoding, etc.)

The Web's public key infrastructure is operated by a community that includes:

• Certificate authorities Verify identities, issue certificates, manage revocation

• Browser/platform developers Implement cert. validation and UI; trust or distrust CAs

• CA/Browser Forum Consortium that sets industry-wide issuance policies

Confidentiality: Public-key encryption

Alice encrypts  $m$  using Bob's public key to get  $c$

Bob decrypts  $c$  using Bob's private key to get  $m$

Integrity/sender authenticity: Digital signatures

Alice signs  $m$  using Alice's private key to get  $s$

Bob verifies  $m, s$  using Alice's public key

Both properties: Use both, with two key pairs

Alice encrypts  $m$  using Bob's public key to get  $c$ , then Alice signs  $c$  using Alice's private key to get  $s$

Bob decrypts  $c$  using Bob's private key to get  $m$

This is called "hybrid" encryption. Advantages:

- Identical messages yield different ciphertexts
- Don't have to worry about RSA padding
- Don't have to worry about message length

In practice, almost always should use

Safely encrypting with RSA

Encrypt a message  $m$  to RSA public key  $(e, N)$ :  $x := \text{PSS}(m)$

Use RSA to encrypt a random  $x \in N$  use a KDF to derive a key from  $x$ , then encrypt message using a symmetric cipher and key  $k$ .

Verifier checks padding:  $x := \text{SHA-256}(m)$

$x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

4.  $x := \text{AES-GCM}(m)$

This is called "hybrid" encryption. Advantages:

- Identical messages yield different ciphertexts
- Don't have to worry about RSA padding
- Don't have to worry about message length

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

3.  $x := s^d \bmod N$

Verifier checks padding:  $x := \text{SHA-256}(m)$

3. Verify that  $x$  is correct PSS-padded  $\rightarrow$

In practice, almost always should use

Safely signing with RSA

Sign a message  $m$  with RSA private key  $(d, N)$ :  $x := \text{PSS}(m)$

Use RSA to sign a carefully padded version of the digest of  $m$ : Many gotchas! A good padding scheme is "Probabilistic Signature Scheme" (PSS).

1.  $x := \text{SHA-256}(m)$

2.  $x := \text{PSIV}$  Caution if you don't correctly verify the padding, attacker can forge signatures

&lt;p

## HTTPS Attacks and Defenses

1. **Fooling Users: Homographs:** visually similar domains Sol: use punycode to block

Fooling users: **Stripping:** (HTTP) SSL Strip attack Sol: HTTP Strict Transport Security

Fooling users: **Phishing:**钓鱼. Sol: Google Safe Browsing/Phishing-resistant authentication

### Mixed Content

(resources loaded over HTTP from inside an HTTPS page) can be modified or leak information.

Browsers block HTTP scripts but may allow images to be loaded over HTTP.

**Mitigation:** Use HTTPS for all resources.

### HTTPS reveals certain information about a site.

TLS: sends certificate chain in plaintext, and domain (but not URL) in plaintext: **Server Name Indication (SNI)** header.

Privacy and censorship concerns.

### CA Weakness: fooling validation

CA identity validation isn't foolproof.

If attacker can **falsely convince a CA** that they control a domain, they can obtain a certificate and fool browsers.

Example: Getting a cert for umich.edu

Attack 1: (Email validation)

CA emails confirmation code to administrator@umich.edu.

Attack 2: (HTTP validation)

Attacker becomes MITM between CA and U-M (e.g., by route hijacking) and redirects CA's GET request to own site.

Web PKI uses a distributed architecture. Thousands of intermediate CAs can issue certificates for any domain.

### Occasionally CAs get hacked.

In 2011, a cert for \*.google.com was issued to an attacker who broke into DigiNotar, a small Dutch CA.

Cert was used for MITM attacks in Iran.

Nobody noticed the attack until somebody found the certificate in the wild...

Google, Microsoft, Apple and Mozilla distrusted all DigiNotar certs, and the CA ceased operation.

### Bugs

### in TLS

### Mitigation:

**Formal verification techniques** can be used to check software correctness using a proof.

miTLS is a formally verified TLS library.

The internet: global network that deliver packets between connected hosts

Packet { header: metadata used by network payload: data to be transported (Any packet might be dropped)}

Each host has a unique identifier (IP address) Routing: A series of routers

Network protocols define how hosts communicate (open standards specified in RFCs)

Protocols are used together via encapsulation HTTPS: [HTTP(REQ/RES)] TLS(8443) TCP(80)

### Protocol layers

Layer 5 Application	Defines how individual applications communicate (e.g., HTTP, SSH, DNS)
Layer 4 Transport	Adds features (ports, connections, encryption) on top of bare packets (e.g., UDP, TCP, TLS, QUIC)
Layer 3 Network	Gets packets to the final destination, over arbitrarily many hops (e.g., IP)
Layer 2 Link	Provides a point-to-point link to get packets to next hop (e.g., Ethernet)
Layer 1 Physical	How bits get translated into electrical, optical, or radio signals

### Common protocol vulnerabilities:

#### Plaintext transmission

Passive attackers can eavesdrop on unencrypted communication.

#### No source authentication

The source address on packets you receive can be spoofed, can't trust it.

#### No cryptographic integrity

Protocols don't prevent data modification. (Checksums stop data corruption, not attacks.)

#### No built-in bandwidth control

Attacks can deny service by flooding hosts or the network itself with traffic.

80 00 20 7A 3F 3E  
Destination MAC Address

80 00 20 20 3A AE  
Source MAC Address

08 00  
MAC Header  
(14 bytes)

IP, ARP, etc.  
Payload  
(46 - 1518 bytes)

00 20 20 3A  
CRC Checksum  
(4 bytes)

Ethernet Type II Frame  
(64 to 1518 bytes)

### Internet Protocol (IP) delivers packets to internet destinations

Packets are in plaintext. May be reordered, lost, corrupted, duplicated or attacked.

**Issue:** How does a host know what MAC address to use to reach a given IP address?

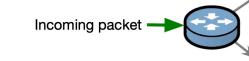
### Address Resolution Protocol (ARP)

is a layer-3 protocol for mapping IP addresses to MAC addresses:

1. Host that needs MAC address M corresponding to IP address A **broadcasts** an ARP packet to entire LAN asking, "who has IP address A?"
2. Host that has IP address A will reply, "IP address A is at MAC address M."
3. Host H caches <IP A: MAC M>

A **router** is a device with multiple link layers, each part of a different network.

Examines incoming packet's destination IP and **forwards** it out a link that will get it closer to the destination.



**Issue:** How does each router know where to send the packet next?

**Response:** ISPs use **Border Gateway Protocol (BGP)** to announce their network prefixes and connectivity to neighbors. Routers compute **route tables** from these announcements.

Host can record all packets arriving ('packet sniffing') data format:pcap software

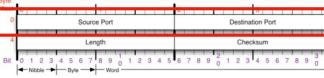
Send data to particular app: each app is identified by 16-bit port number (1~65535) (privileged)

By convention, HTTP:80, HTTPS:443, SMTP:25, DHCP:67, SSH:22, Telnet:23, Root: <0x24

### User Datagram Protocol (UDP)

is a transport-layer protocol that is essentially just a wrapper around IP that adds ports to demultiplex traffic by application.

#### UDP header:



Applications that use UDP:

- DNS
- QUIC
- Many multiplayer games and real-time communications apps

### Transmission Control Protocol (TCP)

is a transport-layer protocol providing:

- Connection-oriented semantics
- A reliable, bi-directional byte stream
- Congestion control

### TCP does not provide: strong security.

Data carried in **segments**, each one IP packet.

#### DNS compromise allows MITM attacks.

An attacker who subverts DNS can return malicious responses that direct clients to visit attacker's server in place of real server.

Examples of DNS attacks:

- **DNS hijacking:** Attacker hacks into home router, changes DHCP settings to point clients to attacker-controller DNS server...
- **Hosts file hijacking:** Malware edits OS hosts file to redirect target sites to malicious server.
- **DNS monitoring** and **DNS blocking:** ISP monitors DNS requests to track users. Injects fake responses to block censored sites.

### Encrypted DNS: DNSSEC

(no confidentiality) (doesn't guarantee results authenticated)

slow adoption by domain & resolver

May fall back to UDP (allow downgrade)

#### DNS stems from DoS attacks

overwhelm a host or network with traffic, such that it can't process legitimate requests

**DoS stems from asymmetry:** especially vulnerable if attacker's cost to make requests is low, victim's cost to process them is high

**DoS Bug:** Design flaw that allows a malicious client to easily disrupt service (e.g., if a malformed request can crash the server)

**DoS Flood:** Attacker sends a large number of messages to exhaust finite resources

**Botnets** are collections of compromised machines (**bots**) under the unified control of an attacker (**botmaster**). Used for DDoS

### Passive Monitoring

Intrusion Detection Systems (IDSes) are software/devices to monitor network traffic for attacks or policy violations

A **virtual private network (VPN)** creates an encrypted channel that "tunnels" IP packets to a distant network location.

### Content Delivery Networks (CDNs):

Large CDNs have enough bandwidth to absorb very large DDoS attacks. Security teams work with ISPs to filter DoS traffic upstream

**Network 1: Integrate into apps** (e.g., SSH)

**Network 2: At transport layer** (e.g., TLS)

**Network 3: Below network** (e.g., VPNs)

**Network 4: In the link layer** (e.g. 5G, WPA3)

Since any host on the LAN can send ARP requests and replies, any host can claim to be another host on the local network!

**ARP Spoofing:** Host X can force IP traffic between hosts A and B to flow through X:

- Claim IP\_A is at attacker's MAC address M\_X
- Claim IP\_B is at attacker's MAC address M\_X
- Re-send IP\_A's traffic to M\_X, and vice versa.

By spoofing the gateway, attacker can MITM all LAN-to-Internet traffic.



**BGP has no authentication: Compromised ISP can announce someone else's network!**

**BGP hijacking** is common, usually due to operator error, but sometimes malicious. Packets to hijacked network are routed to attacker, who can eavesdrop or MITM.

**BORDER GATEWAY PROTOCOL ATTACK — Suspicious event hijacks Amazon traffic for 2 hours, steals cryptocurrency**

Almost 1,300 addresses for Amazon Route 53 rerouted for two hours. ASN: 65000 - ASN: 65000 - 323 PM

### Defense: RPKI

Cryptographic method of signing records that associate a BGP route announcement with the correct originating ISP. Slow adoption...

Host can record all packets arriving ('packet sniffing') data format:pcap software

Send data to particular app: each app is identified by 16-bit port number (1~65535) (privileged)

By convention, HTTP:80, HTTPS:443, SMTP:25, DHCP:67, SSH:22, Telnet:23, Root: <0x24

### UDP properties:

- Connectionless.
- Unreliable data transfer.
- No sequencing or flow control.
- No congestion control.
- **No security.**

UDP is highly vulnerable to impersonation:

If off-path attacker knows what ports a client and server are using, it can inject packets (with spoofed source IP addresses) that appear to be part of the communication.

Using UDP allows/forces applications to provide missing properties at a higher level as needed, which sometimes allows better performance.

TCP provides these despite IP packets being dropped, re-ordered, and duplicated.

Each segment has 32-bit sequence number (where data sent fits in the stream) and acknowledgement number (what data is expected next).

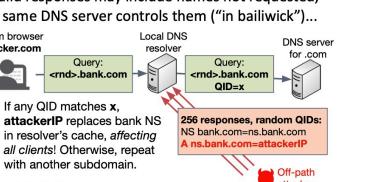
Receiver reassembles segments as output stream. Sender retransmits unacknowledged segments.

### TCP header:

#### Off-path DNS cache poisoning

Attack discovered by Dan Kaminsky in 2008.

DNS authenticates responses with 16-bit QueryID. Valid responses may include names not requested, if same DNS server controls them ("in bailiwick")...



### Partial defense:

Randomize UDP source port (16 bits of entropy).

Randomize capitalization in query (1 bit per letter).

### SYN flood attack:

attacker sends SYN packets from many spoofed source IPs

**SYN cookies** is a defense against SYN flooding that encodes the initial connection state in the SYN-ACK packet itself. Invented by D. J. Bernstein

### DDoS Defenses:

**Egress filtering:** To prevent IP source address spoofing, ISPs should drop packets when source IP is outside the local CIDR block

- **Disadvantage:** Requires global coordination.

All ISPs need to filter for this to be effective, but little incentive for an ISP to implement.

**Content Delivery Networks (CDNs):** Large CDNs have enough bandwidth to absorb very large DDoS attacks. Security teams work with ISPs to filter DoS traffic upstream

**Network 1: Integrate into apps** (e.g., SSH)

**Network 2: At transport layer** (e.g., TLS)

**Network 3: Below network** (e.g., VPNs)

**Network 4: In the link layer** (e.g. 5G, WPA3)