Exploiting Network Structure for Proactive Spam Mitigation

Shobha Venkataraman, Subhabrata Sen, Oliver Spatscheck,
Patrick Haffner, Dawn Song

Portcullis: Protecting Connection Setup from Denial-of-Capability Attacks

Bryan Parno, Dan Wendlandt, Elaine Shi, Adrian Perrig,
Bruce Maggs, Yin-Chun Hu

Presented by: Jason Croft

cs598pbg
Fall 2010

Outline

- Spam
  - Network-Level Properties
  - Historical Nature of IP Addresses
  - Characteristics Network-Aware Clusters
  - Exploiting Properties
- Denial-of-Service Attacks
  - DoS-Limiting Architectures/Techniques
  - Capabilities
  - Puzzles
  - Portcullis Architecture
  - Applications

Exploiting Network Structure for Proactive Spam Mitigation

Shobha Venkataraman, Subhabrata Sen, Oliver Spatscheck, Patrick Haffner, Dawn Song

USENIX Security '07

Properties of Spam

- Ramachandran and Feamster studied 17 months of spam
  - Compared to BGP route advertisements
- Results:
  - Only a few IP address spaces contribute a majority of spam
  - Most spam sent by Windows, each host sending a small amount
  - Spammers use short-lived route announcements to remain untraceable

Ramachandran and Feamster, “Understanding the Network-Level Behavior of Spammers”, *SIGCOMM '06*

Properties of Spam

- 80.* - 90.* majority spam
- 60.* - 70.* majority legitimate
- IP's are transient, 85% < 10 emails

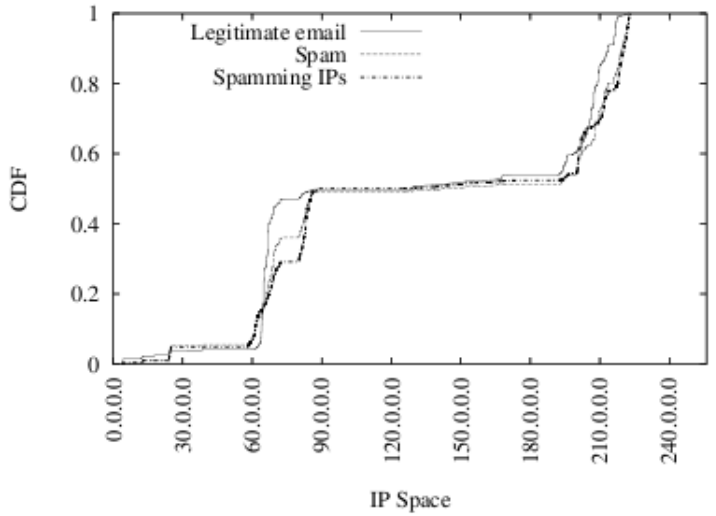


Figure 2: Fraction of spam email messages and comparison with legitimate email received (as a function of IP address space); also, fraction of client IP addresses that sent spam, binned by /24.

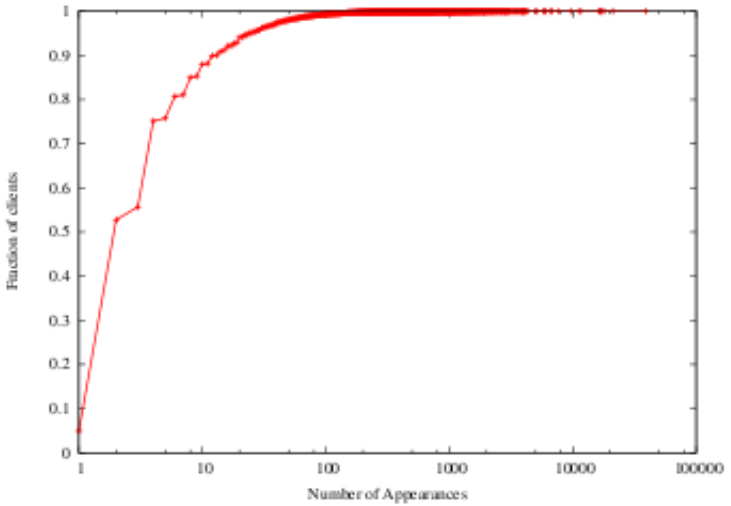


Figure 3: The number of distinct times that each client IP sent mail to our sinkhole (regardless of the number emails sent in each batch).

Ramachandran and Feamster, “Understanding the Network-Level Behavior of Spammers”, *SIGCOMM '06*

Properties of Spam

- > 10% originated from 2 ASes
- 36% originated from 20 ASes
- 40% of spam from top 20 ASes were from US

AS Number	*# Spam*	*AS Name*	*Primary Country*
766	580559	Korean Internet Exchange	Korea
4134	560765	China Telecom	China
1239	437660	Sprint	United States
4837	236434	China Network Communications	China
9318	225830	Hanaro Telecom	Japan
32311	198185	JKS Media, LLC	United States
5617	181270	Polish Telecom	Poland
6478	152671	AT&T WorldNet Services	United States
19262	142237	Verizon Global Networks	United States
8075	107056	Microsoft	United States
7132	99585	SBC Internet Services	United States
6517	94600	Yipes Communications, Inc.	United States
31797	89698	GalaxyVisions	United States
12322	87340	PROXAD AS for Proxad ISP	France
3356	87042	Level 3 Communications, LLC	United States
22909	86150	Comcast Cable Corporation	United States
8151	81721	UniNet S.A. de C.V.	Mexico
3320	79987	Deutsche Telekom AG	Germany
7018	74320	AT&T WorldNet Services	United States
4814	74266	China Telecom	China

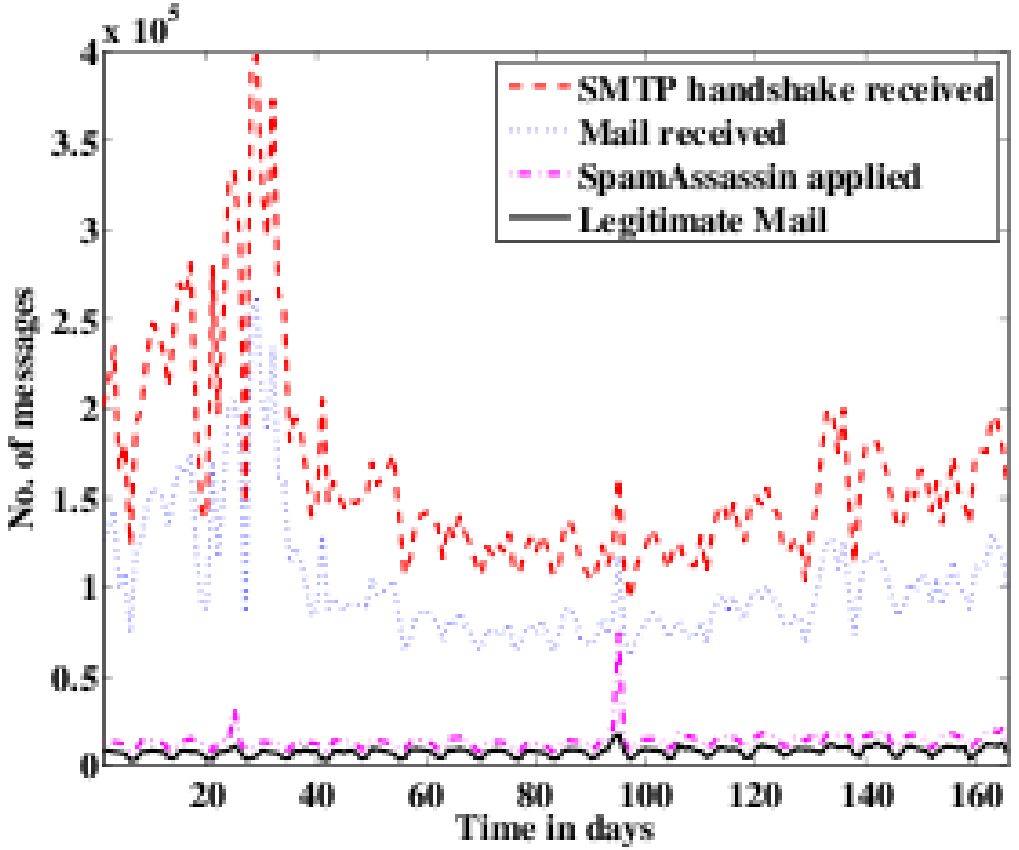
Ramachandran and Feamster, “Understanding the Network-Level Behavior of Spammers”, *SIGCOMM '06*

Properties of Spam (II)

- Venkataraman et al.: Can we predict the legitimacy of mail based on historical nature of the IP addresses?
- Collect traces from large company's mail server
  - 700 mailboxes
  - 166 days (1/2006 – 6/2006)
  - All attempted SMTP connections (IP address, time stamp)
- Assume mail servers under some load, running content filtering (SpamAssassin)

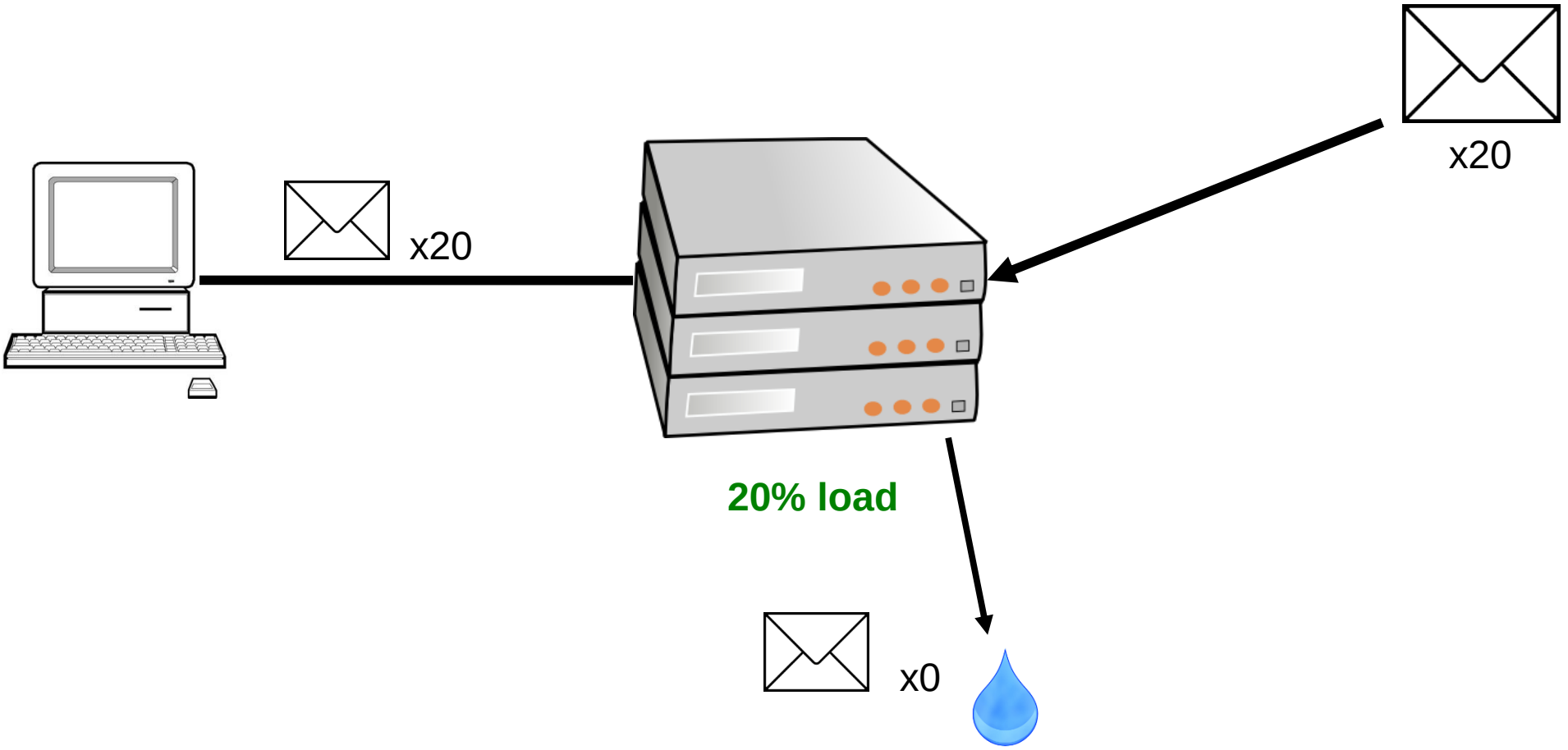
Properties of Spam (II)

- Result: 20x more spam than legitimate mail
  - 1.4 million vs. 27 million



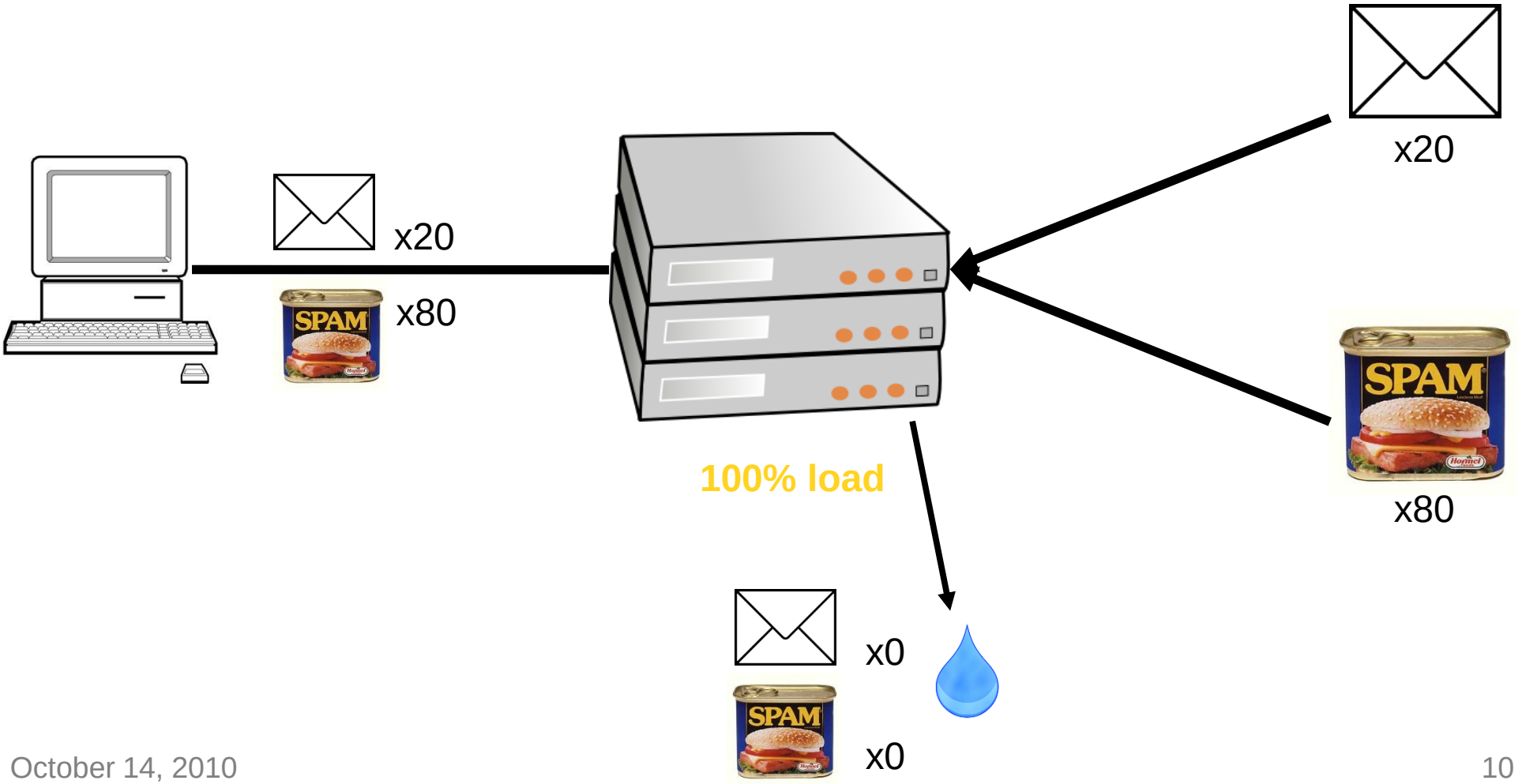
Server under Load

- Server can process 100 emails per second, crash at 200



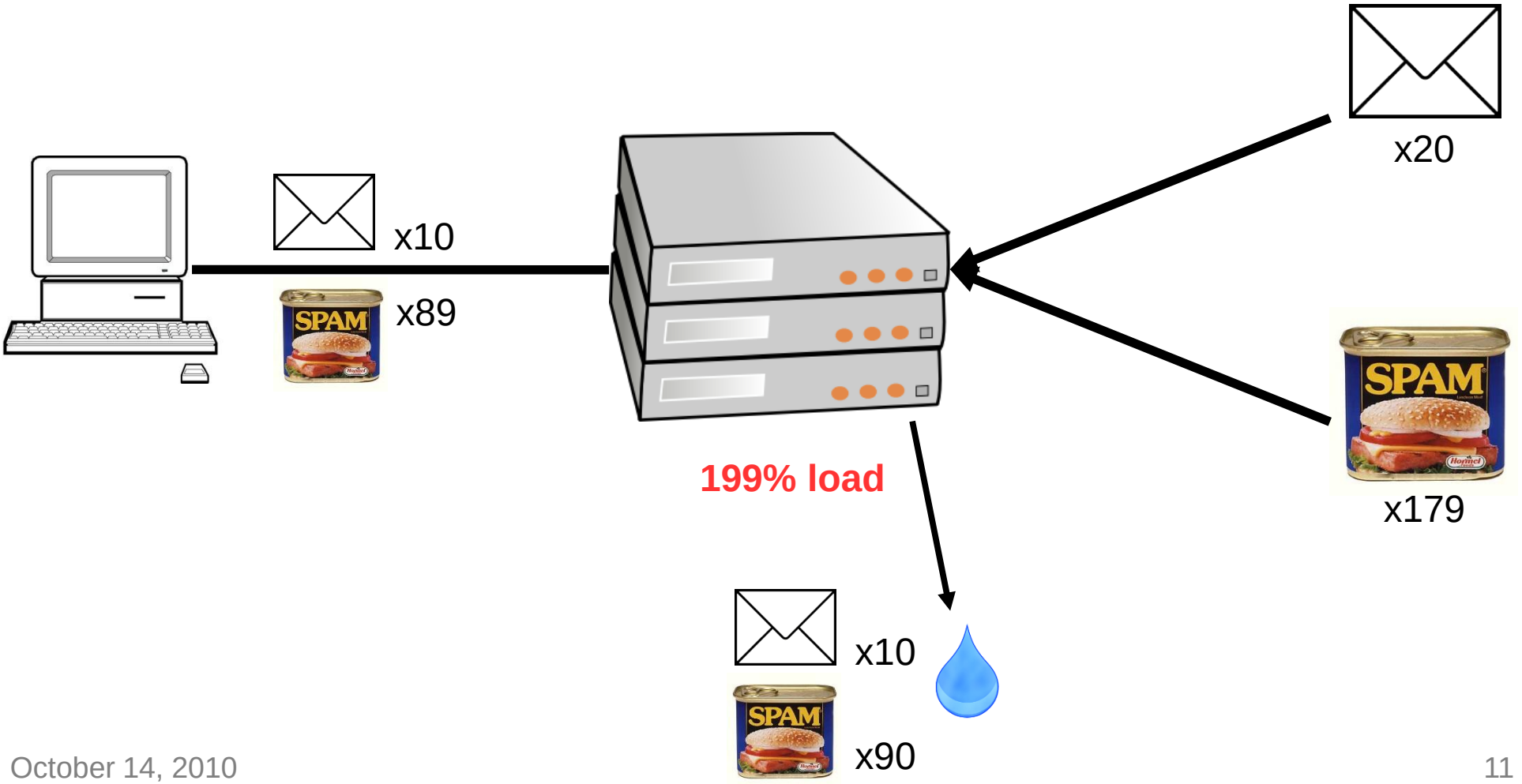
Server under Load

- Server can process 100 emails per second, crash at 200



Server under Load

- Server can process 100 emails per second, crash at 200



Definitions

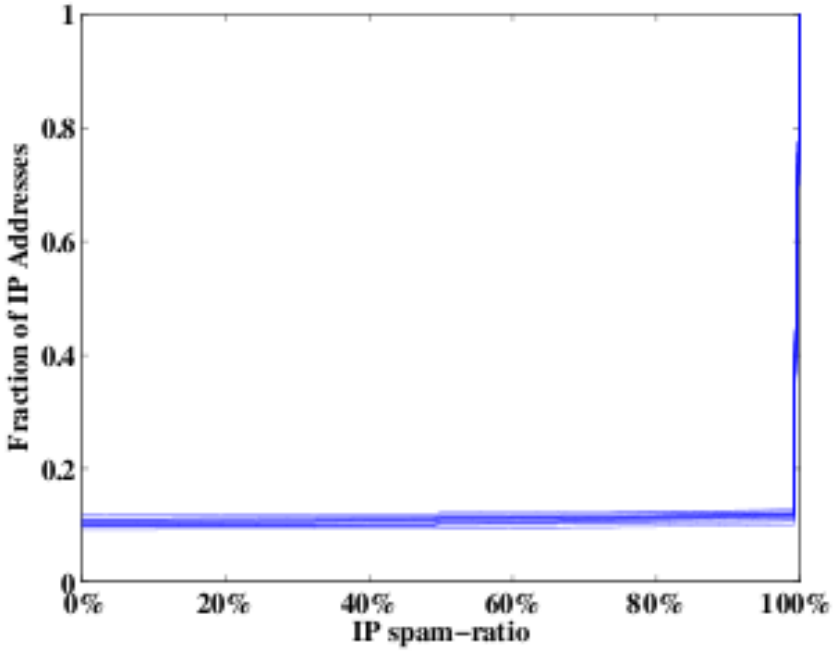
- *Spam-ratio*: fraction of mail sent by IP addresses that is spam
  - Lower => more legitimate mail
- *k-good*: the lifetime spam-ratio of an IP address is at most k
- *k-good set*: set of IP addresses whose lifetime spam-ratios are at most k

Analysis

- Distribution by IP spam-ratio
  - What fraction of legitimate mail or spam is contributed by IP addresses with different spam-ratios?
- Persistence
  - How long does an IP address contribute a major proportion of total legitimate mail?
- Temporal spam-ratio instability
  - How much fluctuation is there in an IP's spam-ratio?

Distribution by IP Spam-Ratio

- Less than 1-2% of IP's have spam ratios between 1%-99%
- 90% of IP's on a given day have spam ratios between 99%-100%
- 99% of spam on a given day comes from an IP with a high spam ratio (> 95%)

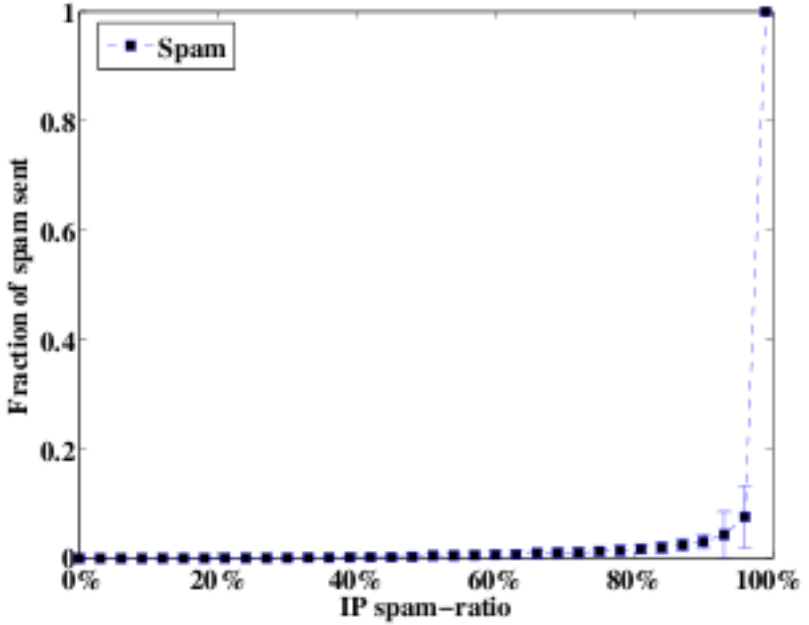


Persistence

- IP's with low lifetime spam ratios contribute a major proportion of total legitimate mail
- The longer an IP address lasts, the more stable its contribution to legitimate mail
- IP's with high spam ratios are present for only a short time

Temporal Spam-Ratio Stability

- Frequency-fraction excess: how often an IP (in a k-good set) exceeds k on a given day
- Majority of IP addresses in each k-good set have frequency-fraction excess of 0
- 95% of IP's have frequency-fraction excess of at most 0.1



Summary

- Good mail servers mostly send legitimate mail and persist for long periods of time
- IP's tend to exhibit stable behavior
- Bulk of mail comes from IP addresses that mostly send spam

Exploiting Findings

- How to use these findings to determine how to prioritize incoming connections?
- Individual IP's don't help too much
- Better: can we determine if the reputation of an unseen IP can be derived from an aggregation of IP's to which it belongs?

Network-Aware Clusters

- Set of unique network IP prefixes collected from a set of BGP routing table snapshots
- Analyze:
  - Granularity: is mail cluster mostly spam or legitimate mail?
  - Persistence: do individual clusters appear over long periods of time?

Results

- Similar to individual IP addresses
- Clusters are at least as temporally stable as individual IP addresses
- Distribution of clusters by daily cluster spam-ratio is similar to distribution of IP addresses by IP spam ratio
- Clusters present for long periods with high cluster spam-ratio contribute large fraction of spam

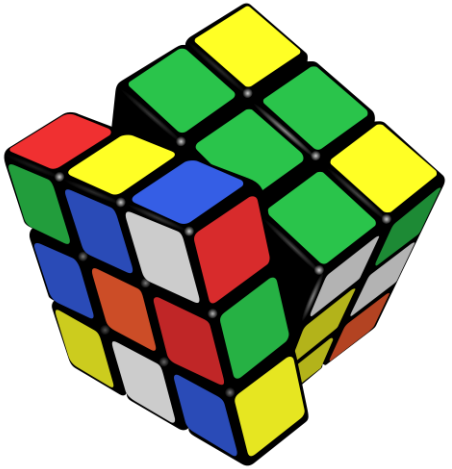
Exploiting Findings (II)

- Mail server under load
- Only for prioritizing based on IP, not a replacement/comparable to content-based filtering
- To selectively accept connections to maximize acceptance of legitimate mail:
  - History-based reputation function R(i)
  - Maximize sum of R(i) over all connections

Portcullis: Protecting Connection Setup from Denial-of-Capability Attacks

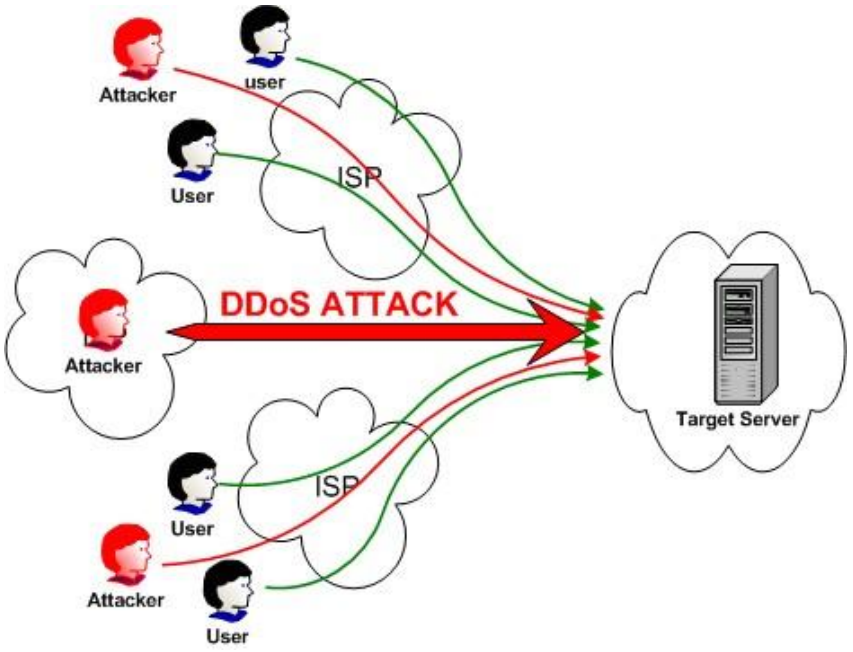
Bryan Parno, Dan Wendlandt, Elaine Shi, Adrian Perrig, Bruce Maggs, Yin-Chun Hu

SIGCOMM '07



Denial-of-Service Attack

- Problem:
  - Victim of DDoS can identify legitimate flows but cannot give flows priority
  - Routers can prioritize traffic but cannot easily identify legitimate traffic (without input from receiver)



Network Capability

- Owner of limited resource should have control over resource usage
- Idea: request to send
  - Source sends capability request packet to destination
  - Routers on path add cryptographic markings to packet header
  - When request arrives, accumulated markings represent capability
  - Capability added to packets to receive priority service
- Prioritize flows based on capability
- What about DoS on capability channel?

Anderson, Roscoe, Wetherall, “Preventing Internet Denial-of-Service with Capabilities”, Hotnets II (2003)

DoS-Limiting Architectures: TVA

- Traffic Validation Architecture (TVA) – capabilities with tags/identifiers
- Trust boundaries – AS edge
  - Tag with small, unique value
  - Tag is identifier for path
  - Fair-queue requests by most recent tag

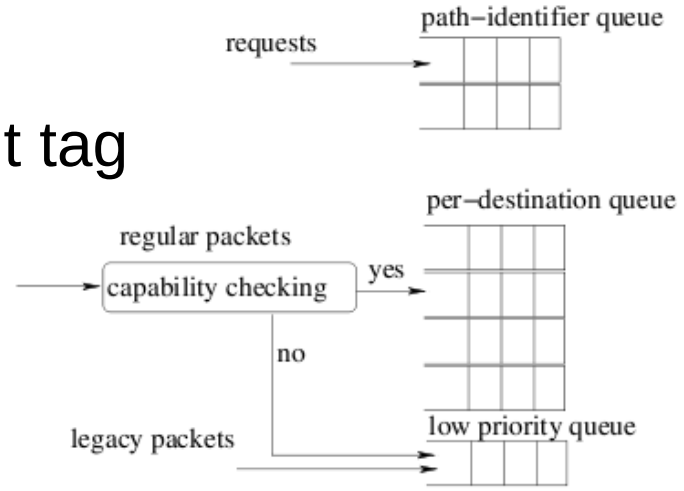
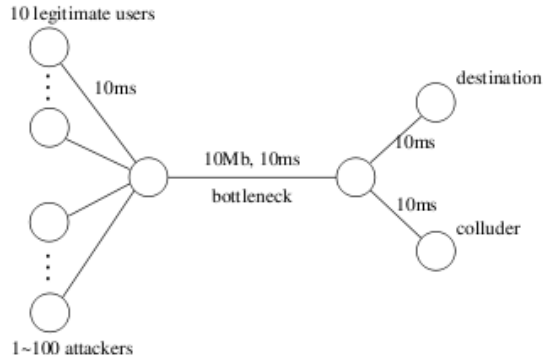


Figure 2: Queue management at a capability router. There are three types of traffic: requests that are rate-limited; regular packets with associated capabilities that receive preferential forwarding; and legacy traffic that competes for any remaining bandwidth.

Yang, Wetherall, Anderson, "A DoS-limiting Network Architecture, SIGCOMM '05

DoS-Limiting Architectures: TVA

- Using identifiers to prioritize traffic is inadequate for large/diverse Internet
  - Can't trust all routers
  - Spoofable
  - Large variation in number of users represented by single identifier/IP (e.g., NAT)
- Legitimate traffic mixes with attack traffic at each AS hop
  - Traffic becomes indistinguishable for TVA's priority mechanism
  - TVA's original analysis used simple topology with single hop, no mixing

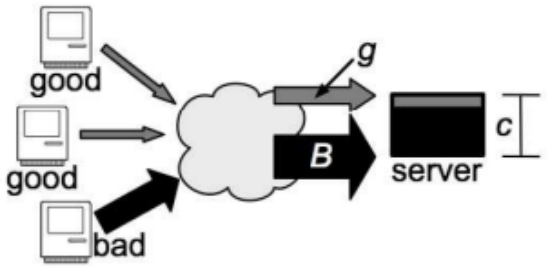


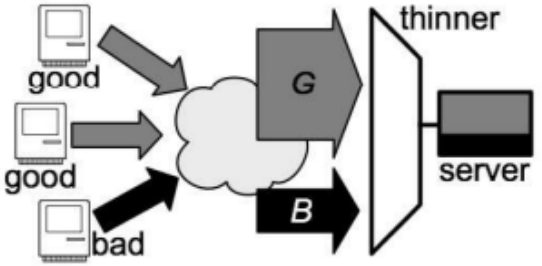
Yang, Wetherall, Anderson, “A DoS-limiting Network Architecture, SIGCOMM '05

DoS-Limitating Architectures: Speak-Up

- Bandwidth as “currency”
- Bandwidth available to users can greatly vary (up to 1500x)
- Assumes network is uncongested
- Focuses on application layer DDoS attacks
  - Protects only end-host resources
  - What about protection for network links?
  - What about effect on other hosts?

- Performance (time to establish capability) declines as number of attacks increases
  - Attackers have more bandwidth relative to legitimate users





Walfish, Vutukuru, Balakrishnan, Karger, Shenker, “DDoS Defense by Offense”, SIGCOMM '06

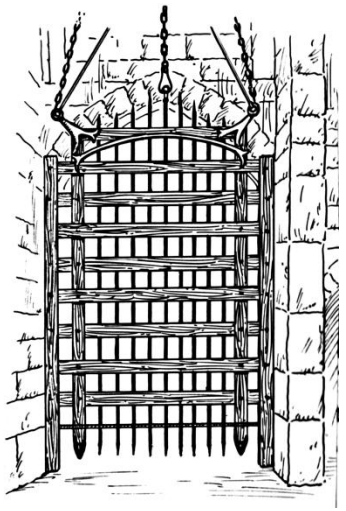
DoS-Limiting Techniques

- Source address filtering
  - Ingress filtering needs high degree of deployment
  - Spoofing among address sharing same prefix
- Pushback – dynamic traffic filters
  - Node tries to characterize types of packets causing a flood, sends requests closer to source to rate limit
  - Difficult at line rate
  - Vulnerable to spoofing, E2E encryption
- Overlay Filtering – reroute traffic to intermediate node and add a secret into header, downstream routers ignore packets without secret
  - Vulnerable to attack if secret is discovered

Anderson, Roscoe, Wetherall, “Preventing Internet Denial-of-Service with Capabilities”, Hotnets II (2003)

Portcullis

- Use capabilities to prevent DoS
- Add puzzles (computational proof of work) to enforce fair sharing of request channel to protect against DoC
- Bounds delay an adversary can impose on legitimate sender's capability establishment



Why Puzzles?

- Better than tagging
- Router provides fairness proportional to work performed by sender
  - Easily verifiable and difficult to spoof
- Computation disparities are smaller than network
  - Workstation vs cellphone: 38x
  - Dialup vs LAN: 1500x

Platform	SHA-1 hashes/minute	Normalized to Nokia 6620
Nokia 6620	25 K	1.00x
Nokia N70	36 K	1.44x
Sharp Zaurus PDA	56 K	2.24x
Xeon 3.20GHz	956 K	38.24x

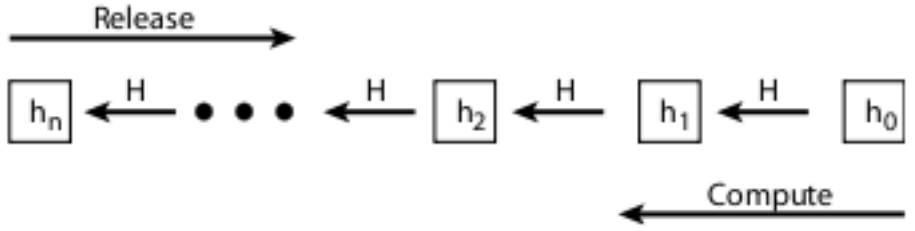
- Puzzle *level* reflects amount of computation required to solve
  - Higher levels have higher priority

Architecture

- Seeds
- Seed Distribution Service
- Puzzle
- Puzzle Verification
- Router Scheduling
- Sender Strategy

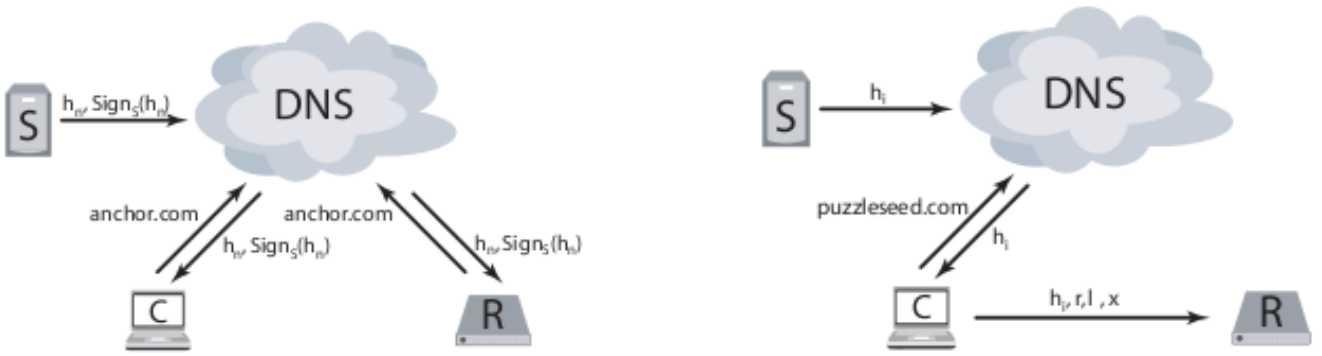
Seeds

- Unpredictable and efficiently verifiable
- Randomly choose h_0 and create hash chain of length n
  - $h_{k+1} = H(h_k \,||\, k)$
- Every t minutes create new seed by reversing chain
- Anchor: last value in chain
- Verification: hash and compare result to seed release from previous time slot
  - Example: in first time slot, sender includes seed h_{n-1}
    - Router verifies $H(h_{n-1} \,||\, n\text{-}1)$ = hash-chain anchor h_n



Seed Distribution Service

- Provide puzzle seeds and hash-chain anchor for roots/senders to verify subsequent seeds
- Can implement using:
  - Private content distribution service
  - Existing DNS infrastructure
    - Already resilient to DoS attacks: highly provisioned and widely replicated



(a) **Yearly Setup** (b) **Connection Establishment**

Puzzles

- Brute-force-like computation
- Solve: $p = H(x \,||\, r \,||\, h_i \,||\, \text{destination IP} \,||\, l)$
  - r = random number, to prevent duplicate puzzles (which routers drop)
  - l = level
  - x = 64-bit value such that the last l bits of p are 0
- Note: no use of source IP to prevent NAT/proxy issues

Puzzle Verification

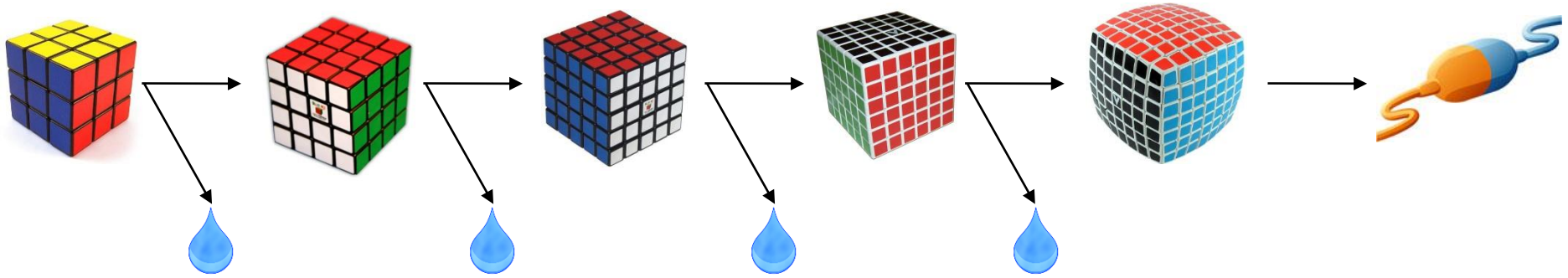
- Little computation for router: only one computation
- Verify seed for h_i: compute $H(h_i \,||\, l)$ and compare to h_{i+1} seed
- Hash provided values: x, r, l (destination IP is provided in packet)
- Verify last l bits are 0

Router Scheduling

- Router's request channel should:
  - Limit reuse of puzzle solutions
  - Give preference to senders solving high-level puzzles
- Bloom filter with solutions, tuple (r, h_i, l, dest IP)
  - Compact lookups
  - False positives, but no false negatives
- Drop packets not passing filter check

Sender Strategy

- Network is under attack
- Sender sends request packet
- On failure, solve puzzle that requires twice the computation, continue until request succeeds



Pricing Applied to Spam

- Pricing function: easy, moderate, hard
  - Proportional to time to compose message
  - To send message, must compute function verified by recipient's mail program
- Shortcut: easier to evaluate pricing function
  - Bypass the access control mechanism
  - Desirable bulk mail (e.g., conference CFP)
- Frequency correspondence list: messages accepted without verification
  - E.g., friends/relatives, mailing lists

Dwork and Naor, “Pricing via Processing or Combatting Junk Mail”, CRYPTO '93

Thanks!

Questions?