



IP TRACEBACK

- Jaysheel Shah

Overview of the presentation

- Introduction
- Approaches
- Arriving at the current approach of Edge sampling
- Working of Edge sampling
- Encoding in Edge sampling
- Benefits & Drawbacks
- Future of Edge sampling
- Conclusion.

Introduction

- What is traceback?
 - A technique for tracing anonymous packet flooding attacks in the Internet back towards their source.
- Issues
 - Simple to implement
 - Difficult to prevent
 - Very difficult to trace
 - Laundering origin - explicitly or implicitly.

Introduction (Cont..)

- **Statistics:**

- 50% increase from 1989-50 (Howard)
- 1999 32% respondents complained DoS (CERT)

- **Goals:**

- Avoid interactive operational support
- Post-mortem
- Incrementally deployable
- Efficient implementation
- No/minimal changes to conventional technologies

Approaches

- Ingress Filtering
- Link Testing
 - Input Debugging
 - Controlled Flooding
- Logging
- ICMP Traceback
- Marking

1) Ingress Filtering

- **Description:**
 - Eliminate the ability to forge IP
 - Block Packets
- **Benefits**
 - Discourages the attacker
 - Limits the forge-IP-space
- **Drawbacks:**
 - Lack of co-operation from the ISPs.
 - Can still be forged.

2) Link Testing

- **Description:**
 - Recursively test all possible upstream paths
 - Only active attacks
 - Two types of link testing:
 - a) Input Debugging
 - b) Controlled Flooding

2a) Input Debugging

- **Description:**
 - Victim develops attack signature
 - Communicate this signature
 - Routers can determine ingress ports of packets on their egress ports.
- **Benefits**
 - Helps traceback accurately
- **Drawbacks:**
 - Human Interface
 - Lack of economic incentives
 - Lack of enough technical skills and capabilities

2b) Controlled Flooding

- Burch and Cheswick

- **Description:**

- Flood upstream traffic
- Behavior/drop of attack

- **Drawbacks:**

- Attack innocent networks
- Method is noisy for multiple attacks

3) Logging

- **Description:**
 - Logging at key routers
 - Data mining techniques
- **Benefits:**
 - Good for postmortem tracing besides live tracing
- **Drawbacks:**
 - Enormous resources
 - Large-scale inter-provider integration

4) ICMP Traceback

- **Description:**

- Packet + Details in ICMP – low probability at routers

- **Benefits:**

- Live and post-mortem
- No human interface
- No change in header

- **Drawbacks:**

- ICMP - filtered off or rate limited
- False ICMP - misleading trace

5) Marking the packets Arriving at the approach

- Marking of packets, probabilistically or deterministically, seems like a more suitable approach than the rest.
 - Live and postmortem traceback
 - No support from the ISP
 - No significant additional traffic
 - No significant overhead on routers
 - It can trace multiple attacks

Arriving at the approach (Cont...)

- Exact traceback and Approximate traceback
- Two components of this algorithm
 - 1) marking procedure
 - 2) path reconstruction procedure
- Convergence time of algorithm is the number of packets that victim observes to reconstruct the path

Assumptions of this approach

- 1) An attacker may generate any packet
- 2) Multiple attackers may conspire
- 3) Attackers may be aware of trace back
- 4) Attackers send numerous packets
- 5) Packets may be lost or reordered
- 6) Routers are both CPU and memory limited
- 7) Route between attacker and victim is fairly stable
- 8) Routers are not widely compromised.

Various Marking Algorithms

5a) Node Append

5b) Node Sampling

5c) Edge Sampling

5a) Node Append

- **Description:**
 - Append each routers IP address
- **Benefits:**
 - Robust
 - Quick to converge
- **Drawbacks:**
 - High router overhead
 - Required space in the packet
 - Unnecessary fragmentation
 - Attacker can fill empty space

5b) Node Sampling

- **Description:**

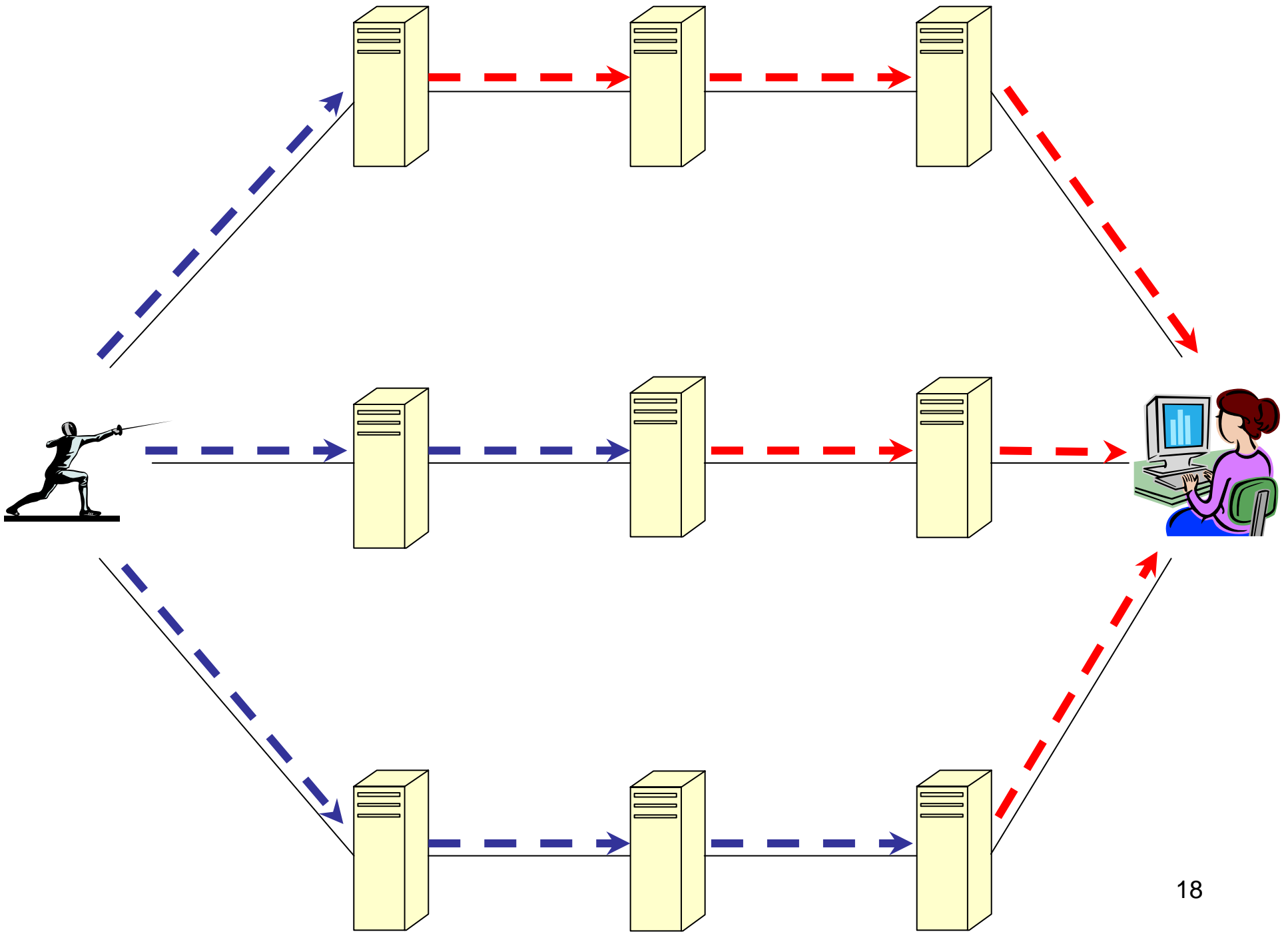
- Append single node per packet with probability p

- **Benefits:**

- $p > 0.5$ - robust trace back against a single attacker

- **Drawbacks:**

- Change the size of IP header
- Inferring the total path is a slow process
- Not robust against multiple attackers



5c) Edge Sampling

- **Description:**

- Add 3 fields in each packet.

- » Start address (w.start)

- » End address (w.end)

- » Distance (w.distance).

- Algorithm:

- Marking procedure at router R:*

- for each packet w

- let x be a random number from [0..1)

- if $x < p$ then

- write R into w.start and 0 into w.distance

- else

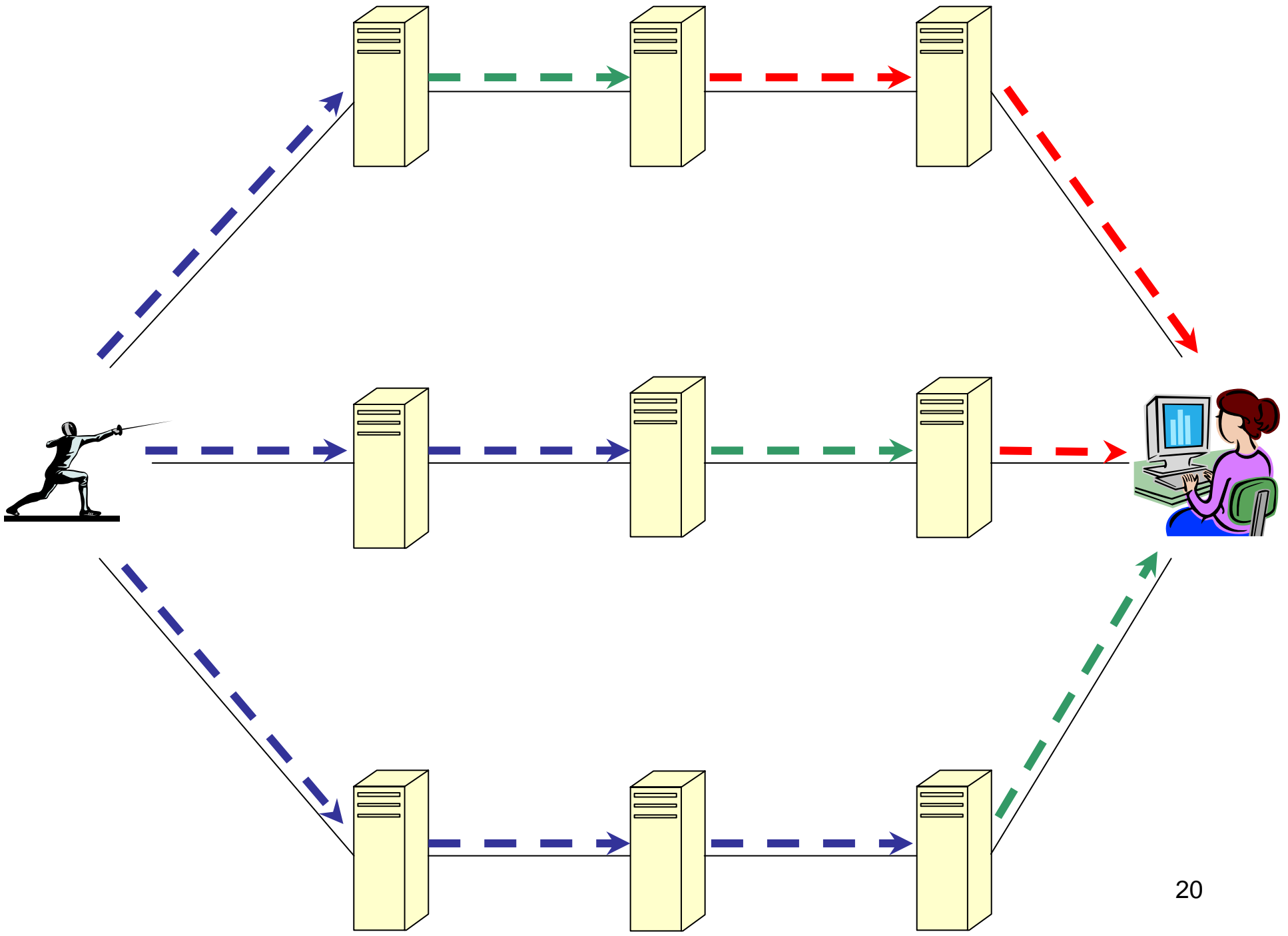
- if w.distance = 0 then

- write R into w.end

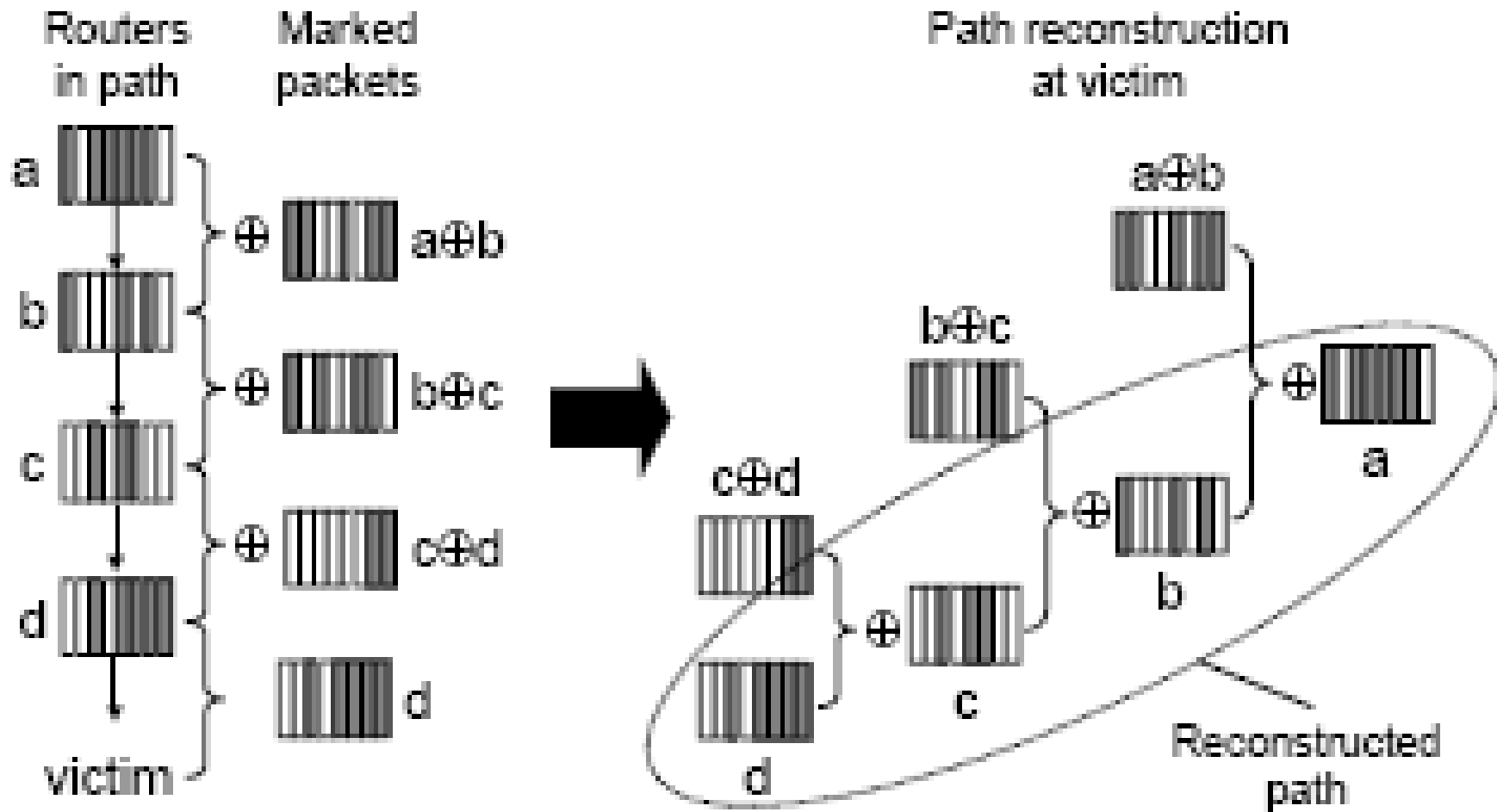
- increment w.distance

- Working:

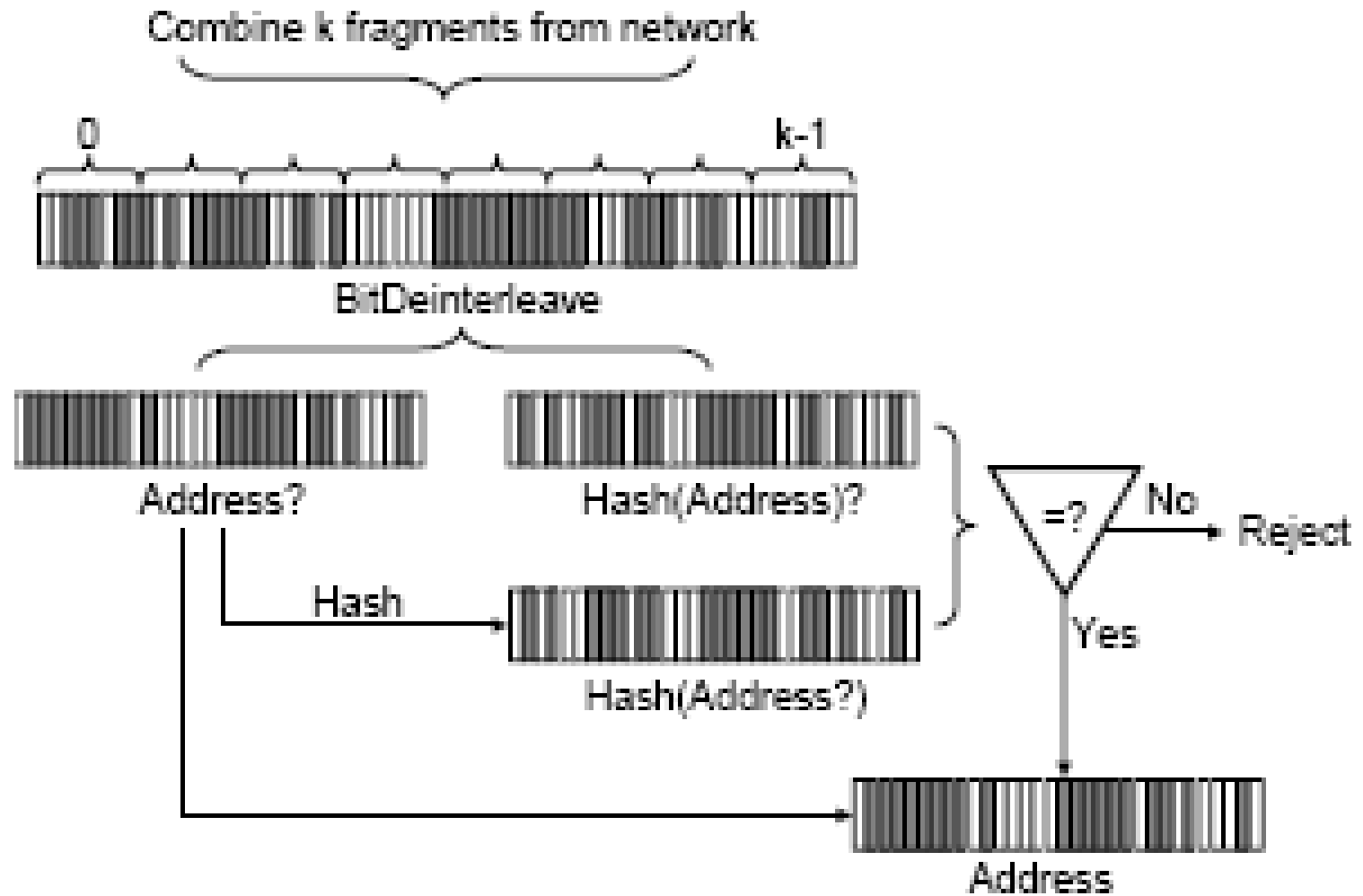
- With $p=1/25$ it takes only 108 packets to converge



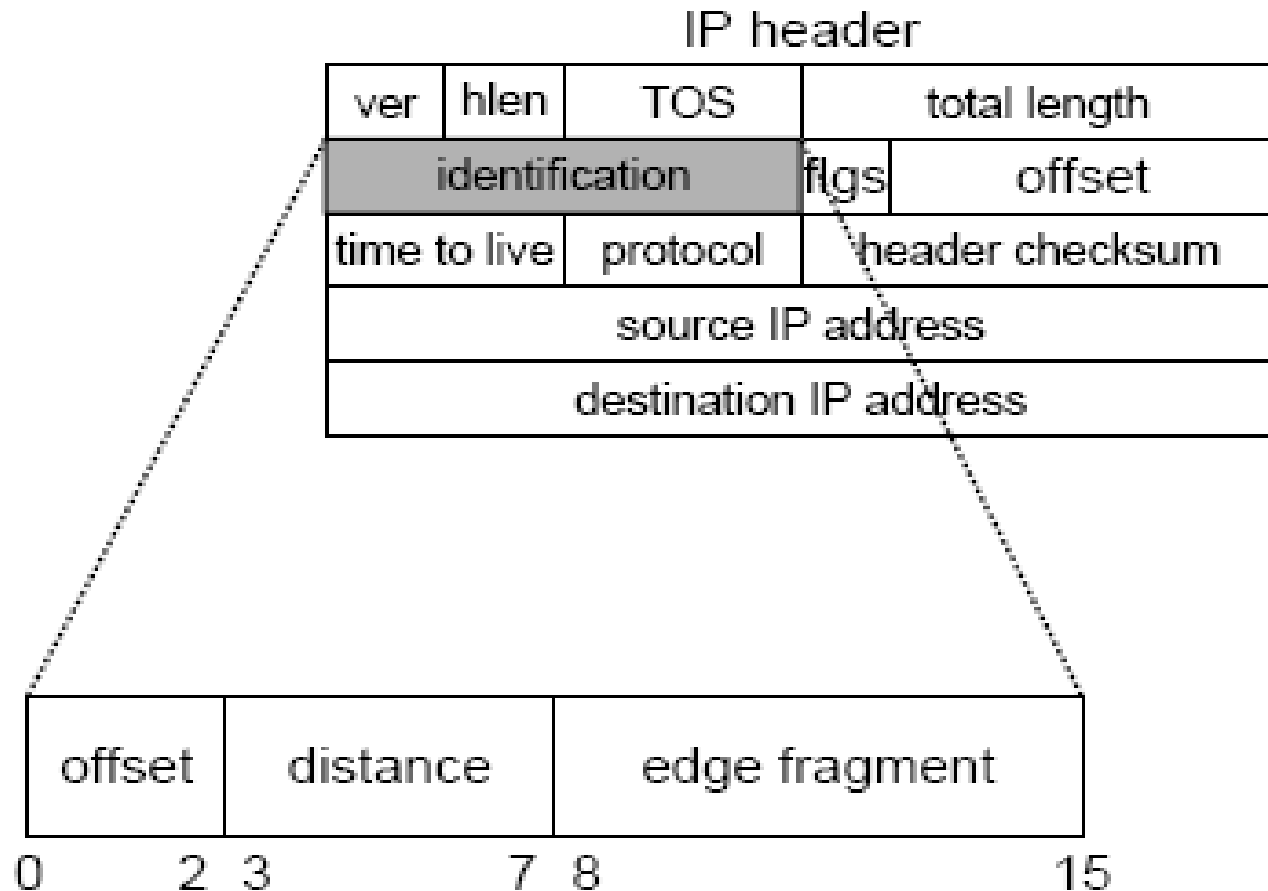
Compressed Edge Sampling



Encoding



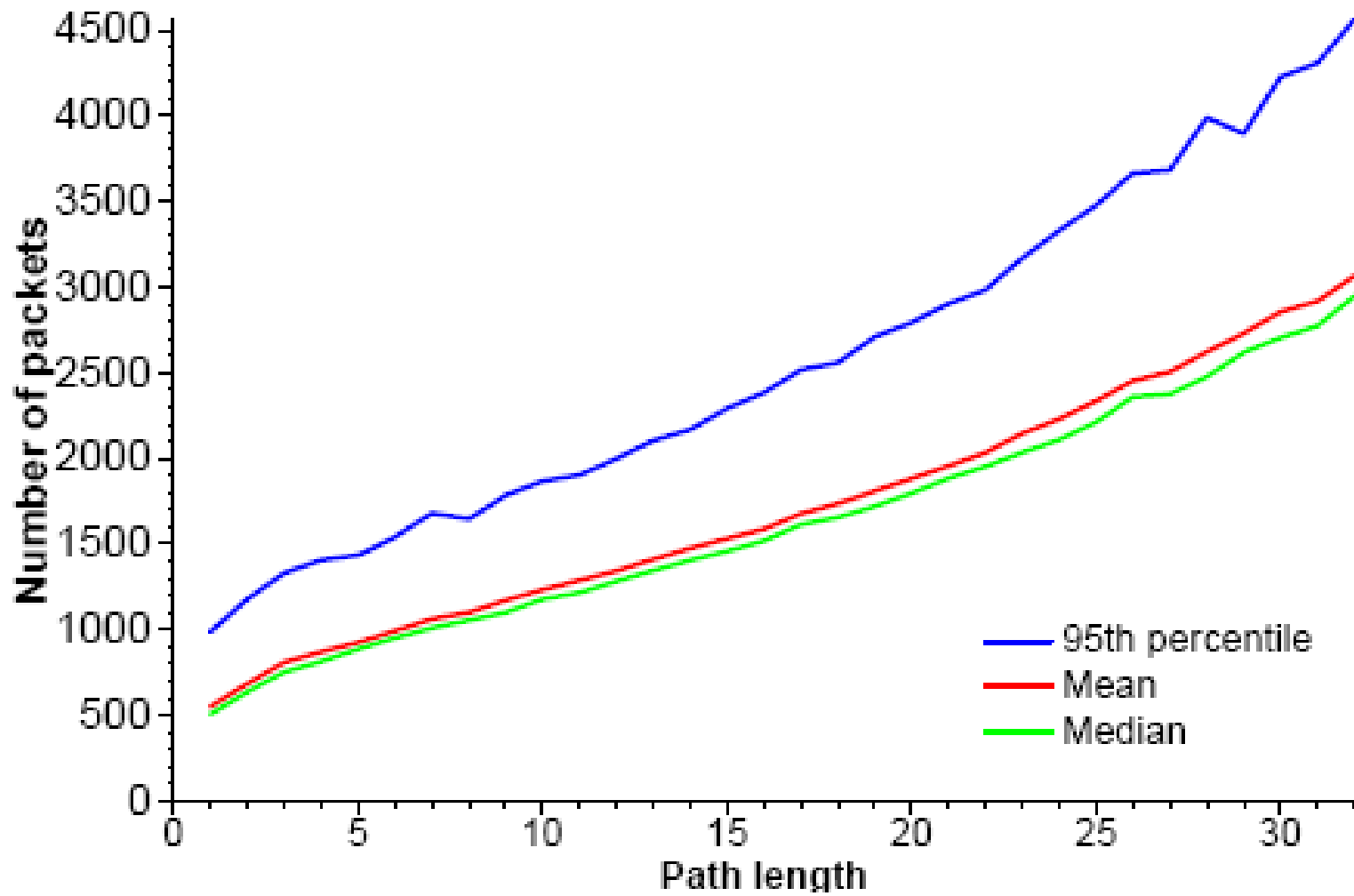
Overloading in IPv4



Real Life example

- Single Attacker
 - k , fragments/edge-id = 8
 - d , distance in hops = 10
 - p , probability = $1/25$
 - *Packets required to converge* < 1300
- Multiple Attacker
 - h , hash length
 - $p(\text{accept arbit edge-id}) = 1/(2^h)$
 - m , number of attackers at d
 - $p(\text{accept incorrect edge-id}) = 1 - (1 - \frac{1}{2^h})^{m^k}$
 - $h=32$, $k=4$, $p(\text{resolving path}) = 97\%$ for 100 routers

Statistics



Benefits & Drawbacks

- **Benefits:**
 - Spoof resistant
 - Path of the closest attacker
 - Robust
- **Drawbacks:**
 - Backwards Capability
 - Distributed Attacks
 - Path validation
 - Attack origin determination approach

1) Backwards Capability

- Negative impact on user that require fragmented IP Datagrams.
- Currently incompatible with parts of IPSec.
 - Solution: request for traceback encoded as a BGP attribute from a particular network.

2) Distributed Attacks

- Difficulty in correctly grouping fragments together.
- Increase in
 - probability of misattributing an edge,
 - amount of state needed to evaluate the decision.
- Significant work required to scale robustness.

3) Path validation

- Unmarking by routers
- Insertion of fake edges
 - Manipulating the identification field
 - Spoof extra edges
- Identification of a valid suffix
 - Differentiate between transit n/w and stub n/w

4) Attack origin detection

- Difficulty due to:
 - Whom are we tracing?
 - ‘laundering’ attacks
 - Spoof their IP and MAC address
 - This requires legal remedies, advanced forensic means from various parties, which is neither in the hands of the victim nor the algorithm.

Future Work

- Overloading 24-bit *flow-label* in IPv6.
- Encoding methodologies.
- Ways to identify valid path suffix.
- Points of indirection such as reflectors not addressed.
- Support from beyond the protocol.

Conclusion

- This is not the final solution but definitely the stepping stone
- Several areas have to be addressed
 - Widely distributed attacks
 - Points of indirection

Reference

- Practical Support for IP Traceback
 - Stefan Savage, David Wetherall, Anna Karlin and Tom Anderson.

Questions?

