SIBRA: Scalable Internet Bandwidth Reservation Architecture

Cristina Basescu, Raphael M. Reischuk, Pawel Szalachowski, Adrian Perrig, Yao Zhang, Hsu-Chun Hsiao, Ayumu Kubota, Jumpei Urakawa

NDSS 2016, San Diego, CA
Headless-browser DDoS
Botnet IPs: Day 1

150 hours

180,000 IP addresses
+690,000,000 hits per day
861 user agents

source: http://www.securityweek.com/ddos-attacks-cost-40000-hour-incapsula
picture: https://www.incapsula.com/blog/headless-browser-ddos.html
Record-breaking DDoS attack in Europe hits 400Gbps
A distributed denial-of-service attack peaked 38 percent higher than last year's Spamhaus attack, the previous DDoS record-holder.

DDoS Attacks Hit Record 500 Gbps in 2015

An Internet of Treacherous Things
A zombie network of home routers highlights the importance of prioritizing smart appliance security.

IoT Security Systems in Alarming Security Fail

By Glenn Fleishman on January 13, 2015
Why are current DDoS defenses inadequate?
Defense Strategies

- **Traffic Scrubbing**: clean incoming traffic from malicious flows
  
  *Useless if a link upstream is flooded*

  ![Diagram showing traffic scrubbing](image)

- **Network Capabilities**: isolate attack traffic from benign traffic
  
  *Useless if links are congested (DoC attacks [32])*

*The Coremelt attack [38] (ESORICS 2009)*

- Exploits a characteristic of today’s Internet: (legitimate) end hosts cannot control the path to bypass congested links
Current defenses lack a crucial property:

**Availability does not diminish** — regardless of the botnet size

"Botnet-size independence"
What ingredients do we need for DDoS defense?
SIBRA: Key Ingredients

Group ASes into Isolation Domains (ISDs)

Autonomous System (AS)

Internet

Distribute control for path construction & resource allocation
between
- source AS,
- destination AS,
- core ASes
Which notion of fairness is required for botnet-size independence?
SIBRA Paths

Fairness between ISDs: core paths

- between ISD Core ASes
- negotiated between direct neighbors
- initiated from destination
- according to previous traffic volumes
- long-term (months)
- optional guarantees e.g., 99.99% availability

ISD United States

AS\textsubscript{C1} \rightarrow \text{ISD Germany} \rightarrow \text{ASD1} \rightarrow \text{ISD Japan} \rightarrow \text{ASB1} \rightarrow \text{ASB2}

2\text{Tbps} \rightarrow \text{ASD1} \rightarrow 1\text{Tbps}

ISD Austria

S \rightarrow \text{ISD United States} \rightarrow \text{ASD1} \rightarrow \text{ASB1} \rightarrow \text{D}
SIBRA Paths

Fairness inside ISDs: steady paths

- requested by inner ASes
- low-bandwidth traffic (control traffic, DNS, ICMP)
- intermediate-term (order of minutes)
- periodically extendable
- basis for launching high-bandwidth reservations
- cryptograph. protected (using local keys)

Fairness between ISDs: core paths

- 30 Mbps
- 50 Mbps

STEADY
Fairness between ISDs: core paths

Fairness inside ISDs: steady paths

E2E reservations: ephemeral paths

** fairness: ** per-source and dest. AS

- bandwidth proportional to steady paths and core paths

---

- requested by end hosts
- high-bandwidth traffic (proportional to steady bw.)
- short-term (tens of seconds)
- periodically extendable
- similar to leased lines (more flexible and cheaper)
- similar to virtual paths (with security protection)
How much bandwidth do ephemeral paths obtain?
2-Dimensional Bandwidth Decomposition

1. **vertical**
   (hierarchical, per-location)

2. **horizontal**
   (per-link)
2-Dimensional Bandwidth Decomposition

1. **vertical**
   (hierarchical, per-location)

   **case 1)** *source ISD*

   **case 2)** *between ISDs*

   **case 3)** *destination ISD*

2. **horizontal**
   (per-link)
2-Dimensional Bandwidth Decomposition

1. **vertical**  
   (hierarchical, per-location)

   - **case 1)**  
     source ISD

   - **case 2)**  
     between ISDs

   - **D** ISD United States

   - **S1**
   - **S2**

2. **horizontal**  
   (per-link)

   - **ASD1**
   - **ASB1**
   - **ASB2**
   - **ASH**
   - **ASK**

   - **core path**
   - **ephemeral path**
   - **steady path**

   - **80% ephemeral**
   - **5% steady**
   - **15% best-effort**
2-Dimensional Bandwidth Decomposition

1. **vertical**
   (hierarchical, per-location)

   - **case 1)**
     source ISD

   - **case 2)**
     between ISDs

   - **case 3)**
     destination ISD

2. **horizontal**
   (per-link)

   - ISD United States
   - ISD Germany

---

100 Gbps
80 Gbps _ephemeral_
5 Gbps _steady_
15 Gbps _best-effort_

---

**Source ISD**: AS\(_A2\)

**Destination ISD**: AS\(_B1\)

---

**Path Details**:
- **Core Path**: 80% _ephemeral_, 5% _steady_, 15% _best-effort_
- **Ephemeral Path**: 80% _ephemeral_

---

**Network Locations**:
- **ASE**: ISD Austria
- **ASG**: ISD United States

2-Dimensional Bandwidth Decomposition

1. vertical
   (hierarchical, per-location)

   case 1) source ISD

   30 Mbps
   90 Mbps

   100 Gbps
   80 Gbps ephemeral
   5 Gbps steady
   15 Gbps best-effort

2. horizontal
   (per-link)
2-Dimensional Bandwidth Decomposition

- **Source ISD**: $S_1$, $S_2$
- **Core Path**: Core Path
- **Destination ISD**: $D$
- **Core Path Bandwidth**: 2 Gbps
- **Ephemeral Path Bandwidth**: 480 Mbps
- **Case 1**: Source ISD

**Mathematical Expression**: $80 \div 5 = 480$ Mbps
2-Dimensional Bandwidth Decomposition

Source ISD

\[ S_1 \]

\[ S_2 \]

core

\[ 480 \text{ Mbps} \]

\[ 960 \text{ Mbps} \]

Destintion ISD

\[ D \]

\[ 50 \]

\[ 30 \]

\[ 90 \]

\[ 880 \]

Mbps

\[ 2 \text{ Gbps} \]

steady path

core path

ephemeral path

\[ 30 \text{ Mbps} \]

\[ 30 / (30+90+880) \]

\[ * \]

\[ 2 \text{ Gbps} \]

\[ * 80 / 5 = 480 \text{ Mbps} \]

\[ * 80 / 5 = 960 \text{ Mbps} \]
2-Dimensional Bandwidth Decomposition

30 Mbps

\[ \frac{30}{(30+90+880)} \times 2 \text{ Gbps} = 480 \text{ Mbps} \]

\[ \frac{30}{(30+90+880)} \times \frac{2}{(2+8)} \times 50 \text{ Mbps} = 4.8 \text{ Mbps} \]
2-Dimensional Bandwidth Decomposition

1. **vertical**
   (hierarchical, per-location)

2. **horizontal**
   (per-link)

---

**1. Vertical (hierarchical, per-location)**

- **ASc1** connected to **ASC2**
- **ASC2** connected to **ASp1**
- **ASp1** connected to **ISD Germany**
- **ISD Germany** connected to **ASB1**
- **ASB1** connected to **ASD1**

- **480 Mbps** connection between **ASF** and **ASG**
- **960 Mbps** connection between **ASC1** and **ASC2**
- **4.8 Mbps** connection between **ASB1** and **ASD1**

**2. Horizontal (per-link)**

- **4.8 Mbps** path from **ASB1** to **ASB2**
- **960 Mbps** path from **ASF** to **ASG**
- **80% ephemeral**
- **15% best-effort**
- **5% steady**

---

**Network Design:**
- **S1** connected to **ASG**
- **S2** connected to **ASE**
- **ISD Austria**
- **ISD Japan**
- **ISD United States**
1. vertical
(hierarchical, per-location)

2. horizontal
(per-link)

bottom line:
**ephemeral BW is proportional to steady BW**
(source-ISD paths, core paths, dest-ISD paths)

unused st./eph. BW is loaned to best-effort BW
(through statistical multiplexing)
SIBRA Guarantees

• Source AS S initiates a reservation. Each AS on path accepts or declines and provides a cryptographic token:

\[ \text{MAC}_{K_i} ( \text{ingress}_{AS_i} \| \text{egress}_{AS_i} \| \text{Request} \| RT_{AS_{i-1}} ) \]

• Efficiency & Scalability: ASes verify these tokens, embedded in the forwarded packets, i.e., no per-flow state.

CBC-MAC (AES) Intel’s AESni [16] 4.15 cycles/byte

\[ RT_{AS_i} = \text{ingress}_{AS_i} \| \text{egress}_{AS_i} \]

16
SIBRA under Attack

Per-neighbor monitoring at transit ASes (fastpath)

Botnet A

Per-flow monitoring at the edge (slowpath, [37])

Botnet C

Botnet B

Probabilistic monitoring at transit ASes (fastpath, [43])

Botnet D
Is there enough bandwidth in today’s Internet?
Case study: core links to Australia

- The entire world connects to Australia (32,428 leaf ASes)

### Capacity (Gbps)
- (2) Australia - Papua New Guinea-2
- (3) PIPE - Pacific Cable-1
- (4) Australia - Japan Cable
- (5) Gondwana-1
- (6) Southern Cross Cable Network
- (7) Telstra Endeavor
- (8) Tasman-2

---

**463.9 Mbps**
(371.1 Mbps ephemeral bandwidth)

for each AS

**5.64 Gbps**
in 2018
How effective is SIBRA?
Evaluation: Defense against Coremelt

![Graph showing file transfer time vs number of attacker pairs](image)

- **SIBRA**: Blue line, consistently low file transfer times across all attacker pairs.
- **TVA**: Green line, shows a steady increase in file transfer time with increasing attacker pairs.
- **Portcullis**: Orange line, similar to TVA but slightly slower.
- **STRIDE**: Yellow line, the most affected, with a sharp increase in file transfer time from 5,000 attacker pairs onwards.

Fig. 12: Simulation results on SIBRA’s availability. (a) shows the existence of the reservable bound for bandwidth request, (b) presents the tolerability of bandwidth reservation against packet loss.
How efficient is SIBRA?
Per-flow Stateless Operations

10 Gbps core link (load ~40%): $2.2 \times 10^5$ flows per second
1 Tbps core link (load ~40%): $2.2 \times 10^7$ flows per second

Storing per-flow state is prohibitively expensive — especially under attack

### Router Action

<table>
<thead>
<tr>
<th>Action</th>
<th>Time (avg)</th>
<th>Per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing 1 reservation request</td>
<td>9.10 $\mu$s</td>
<td>110 K</td>
</tr>
<tr>
<td>Processing 1 packet (1500 bytes) using Intel's DPDK and AESni</td>
<td>0.04 $\mu$s</td>
<td>25 Mio</td>
</tr>
</tbody>
</table>

280 Gbps
Conclusions

- **Botnet-size independence** is the key property against DDoS attacks
- SIBRA is the first **bandwidth reservation architecture** to achieve botnet-size independence at Internet scale
- Two-dimensional **bandwidth decomposition**
- Very fast operations, per-flow stateless forwarding

Related Work


Backup
Parameter Choice: Traffic Types

- **ephemeral (80%)**
  - Netflix’s video constitutes >50% of the entire Internet traffic
  - together with YT and FB, 70-90% are realistic for ephemeral traffic

- **steady (5%)**
  - based on a 10-day measurement of a tier-1 ISP: connection establishment (TCP-SYN) uses 0.5% of the bandwidth
  - SIBRA allocates 10x that amount

- **best-effort (15%)**
  - email, news, SSH, DNS (3.9%)
  - very short-lived flows, less than 256ms (5.6%)