Identifying Anomalous Traffic Using Delta Traffic

Tsuyoshi KONDOH and Keisuke ISHIBASHI

Information Sharing Platform Labs.

NTT

Flocon2008, January 7–10, 2008, Savannah GA

Outline

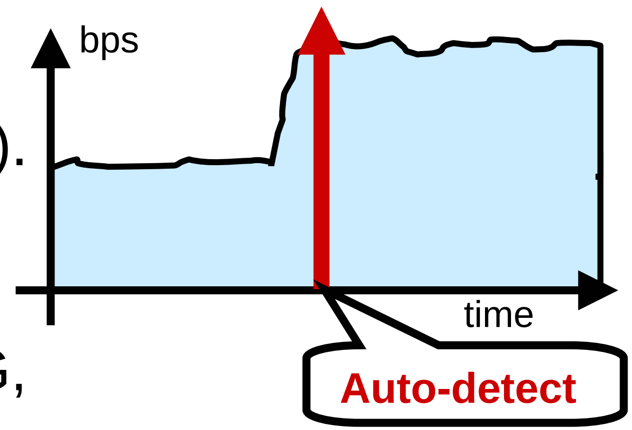
- Background and Motivation
  - Identifying anomalous traffic is the missing piece.
- Our Technique: DELTAA
  - Concepts
    1. Extract anomalous traffic as the delta of normal and anomalous time periods.
    2. Auto-aggregate extracted anomalous traffic.
  - Operation of our technique
    - How to implement the above concepts.
- Evaluation
  - Evaluation using synthesized DDoS traffic.
- Summary

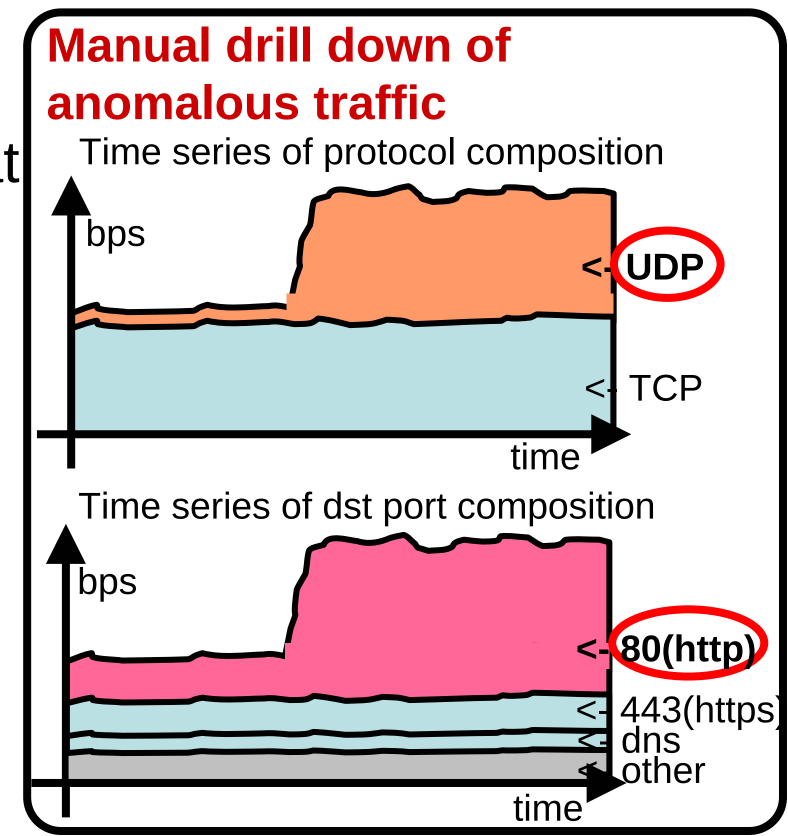
Background and Motivation

- Monitoring of traffic volumes is widely used for network operation (e.g. MRTG).
- Many techniques for detecting anomalous volume change have been proposed (NBAD, Holt-winters in MRTG, … etc.).
- Some tools to mitigate damage from anomalous traffic. (e.g. drop/rate limit at router, detour to Cisco Guard, etc.)

- However, accurate mitigation needs accurate ACL sets.
- Generating accurate ACL sets requires manual drill down by operator.
  - **Too costly.**

Time series of total traffic by bps





Our Technique: DELTAA

- **DELTAA** outputs ACL sets for filtering or rate limiting to mitigate the damage from anomalous traffic.
  - DELTAA: Delta Traffic Automatic Aggregator

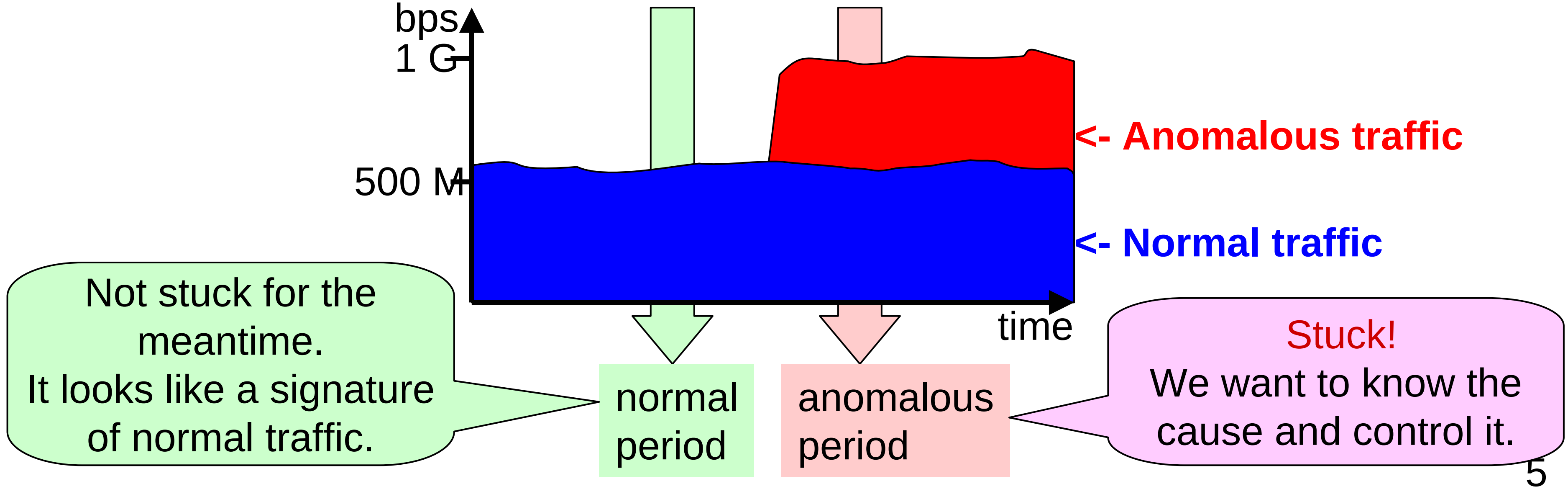
Today, I will focus on two concepts

- Three concepts of DELTAA:
  1. Reveal anomalous traffic using delta traffic.
  2. Aggregate delta traffic and generate optimized ACL sets on a single dimension (e.g. source IP address dimension).
  3. Generate multi-dimensional ACL sets by integrating each dimensional anomalous traffic range.

Concept #1:
(1) Definition of “Normal” and “Anomalous” Traffic

Throughout this presentation, I use the following definitions.

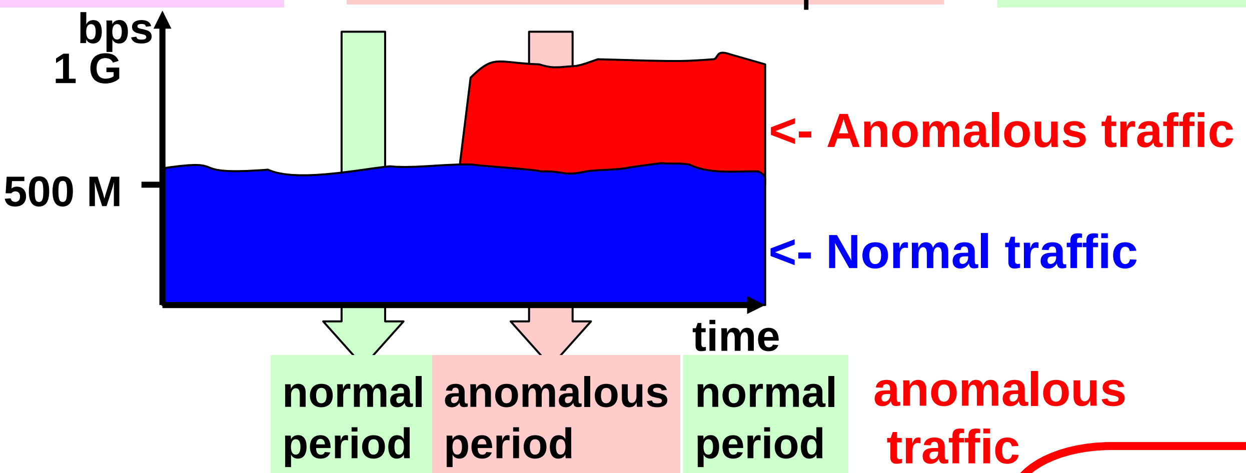
- Anomalous traffic: Traffic that causes a change in traffic volume (bps/pps/fps).
  - BitTorrent and server intrusion are out of scope because they always exist or do not cause a volume change.
- Normal period: Period when traffic volume is normal.
- Anomalous period: Period when traffic volume is anomalous.



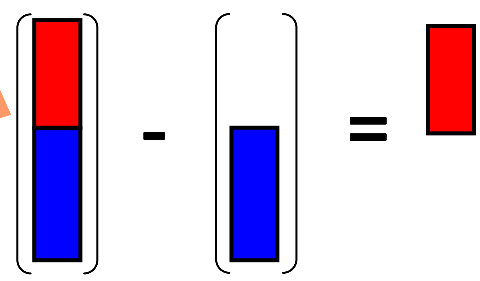
Concept #1 :
(2) Reveal Anomalous Traffic

- Make two assumptions
  1. traffic of normal period = normal traffic
  2. traffic of anomalous period = normal traffic + anomalous traffic
- We can then extract anomalous traffic as the delta of the above two periods.

anomalous traffic = traffic of anomalous period – traffic of normal period



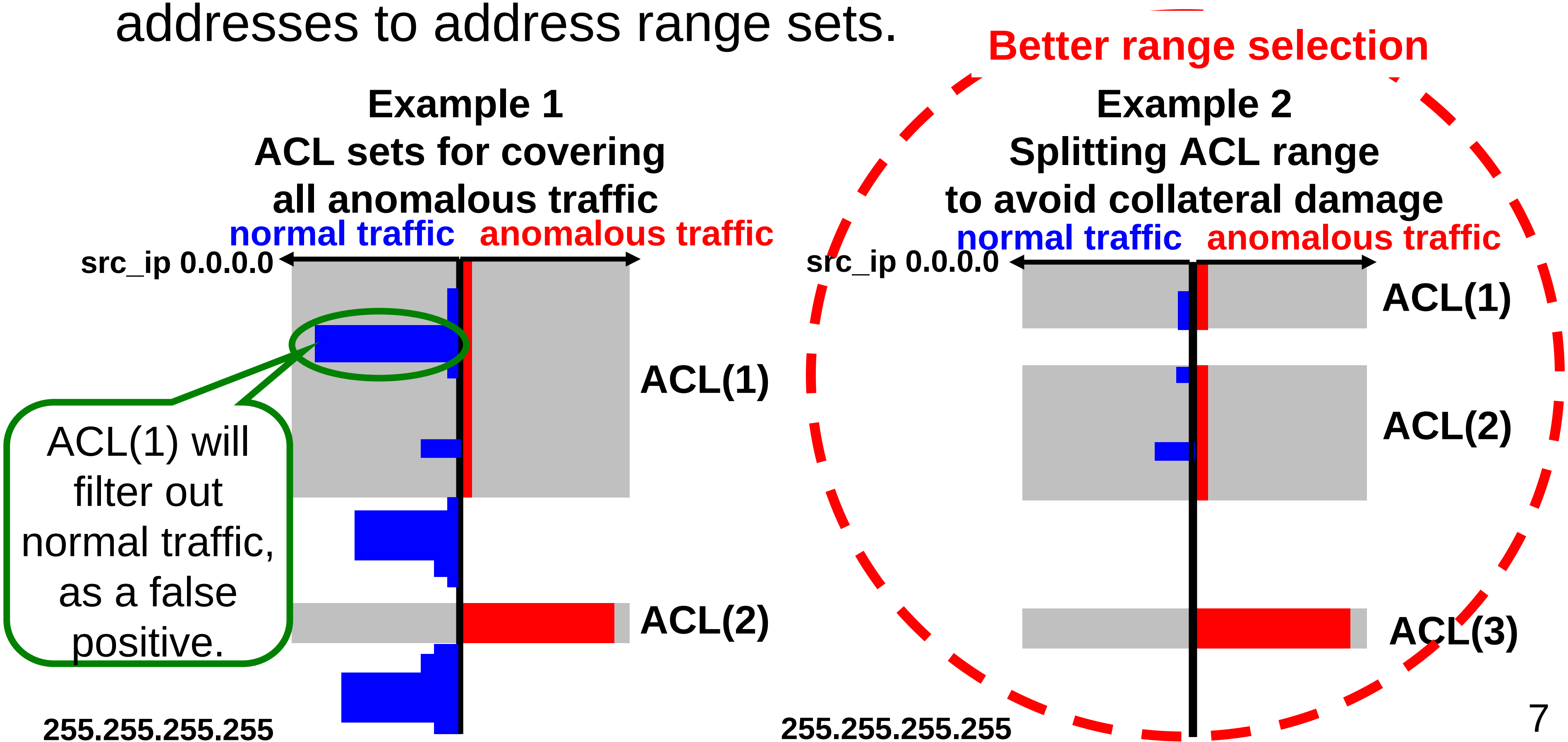
Extracting anomalous traffic from “traffic of anomalous period” is difficult because it is a mixture of normal and anomalous traffic.



Taking the delta between “traffic of normal period” and that of anomalous period, we can effectively extract anomalous traffic.

Concept #2:
Auto-aggregate Delta Traffic

- In aggregation, **optimize a trade-off** (false negative, false positive, number of ACLs) by **using the best range-selection algorithm**.
- Aggregation example: Aggregate from distinct source IP addresses to address range sets.



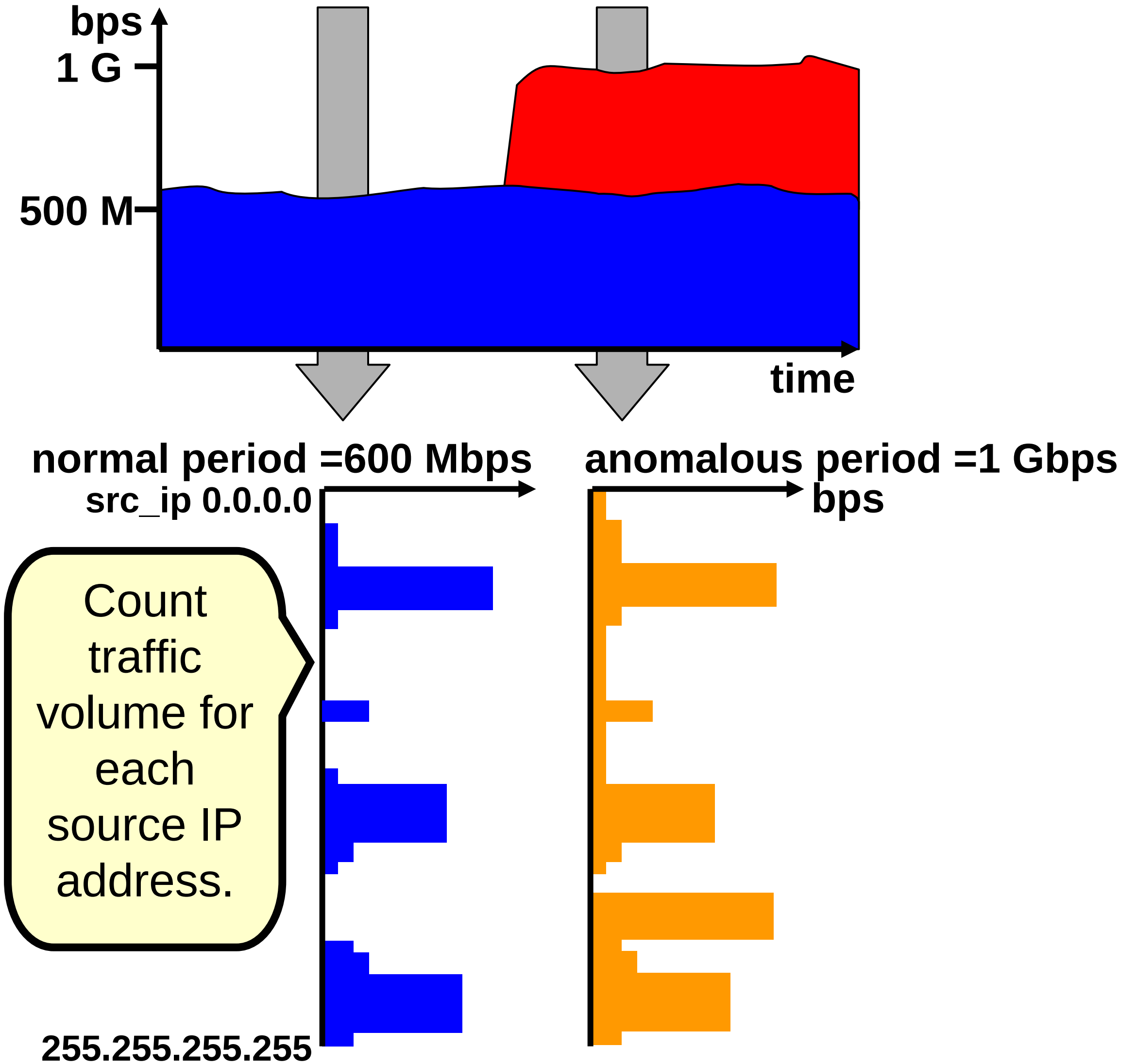
Explanation of Our Technique

- Our technique can generate multi-dimensional ACL sets.
  - e.g. source/destination IP address, source/destination port, protocol, flow exporter, and router interface
  - Multiple dimensions do not mean independent of above information sets.
  - Our technique merges above information to make multi-dimensional ACL sets.

- In this presentation, I focus on source IP dimension identification as an example and explain step by step.

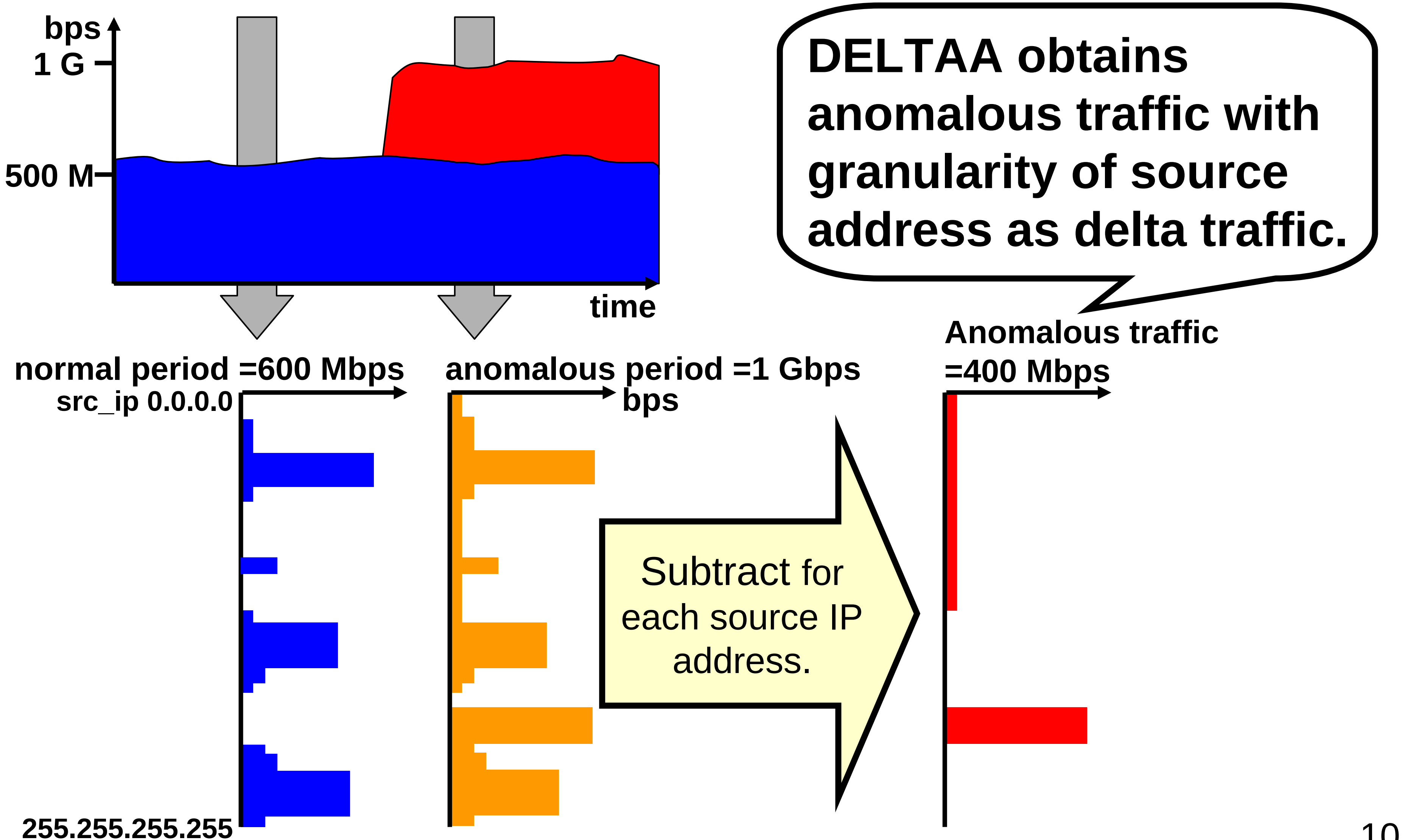
Step 1: (1) Counting Up

Count normal and anomalous periods of traffic for each source IP address.



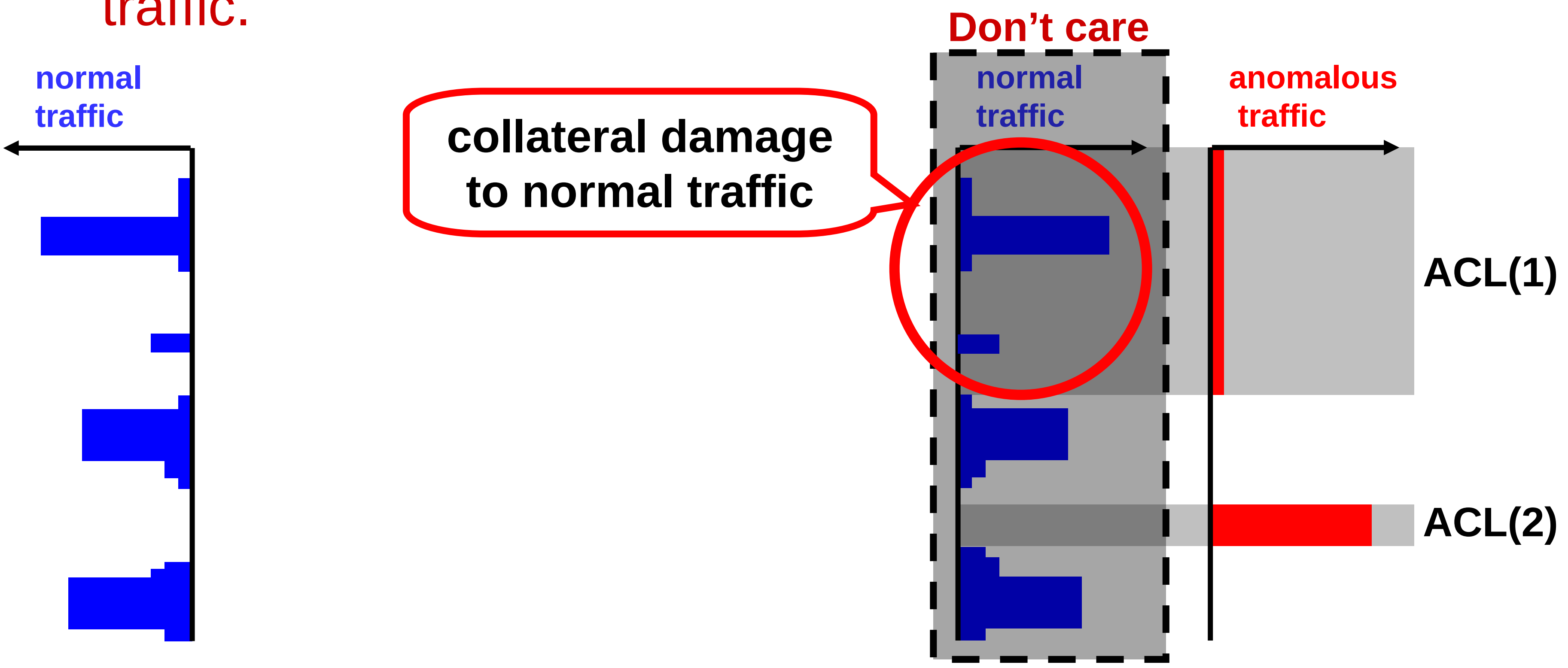
Step 1: (2) Making Delta Traffic

Make delta traffic by subtracting traffic of normal period from that of anomalous period.



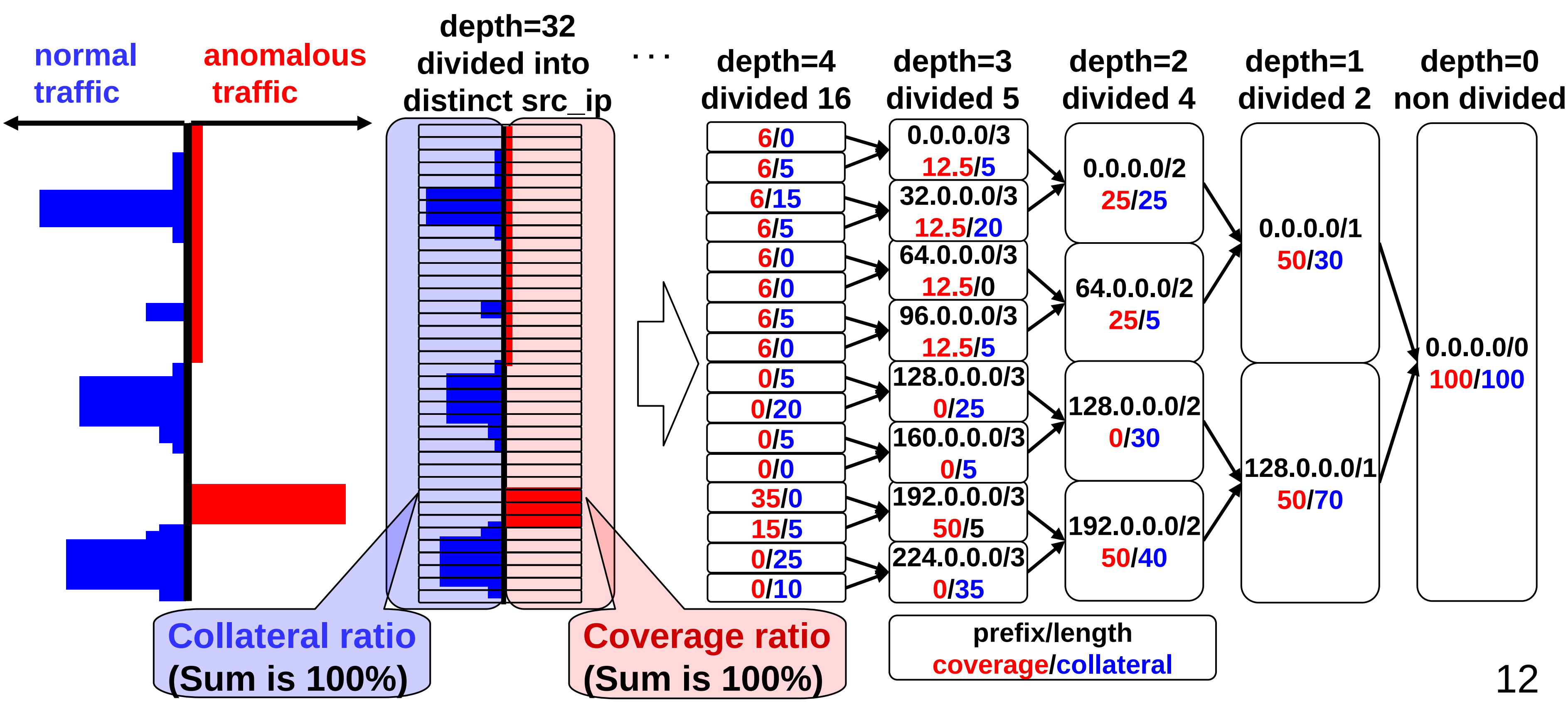
Step 2: (1) Building Tree of Normal and Anomalous Traffic

- Example: When we use only anomalous traffic information, collateral damage cannot be avoided.
  - Causes mis-filtering of normal traffic.
- So, build a traffic tree using both normal and anomalous traffic.



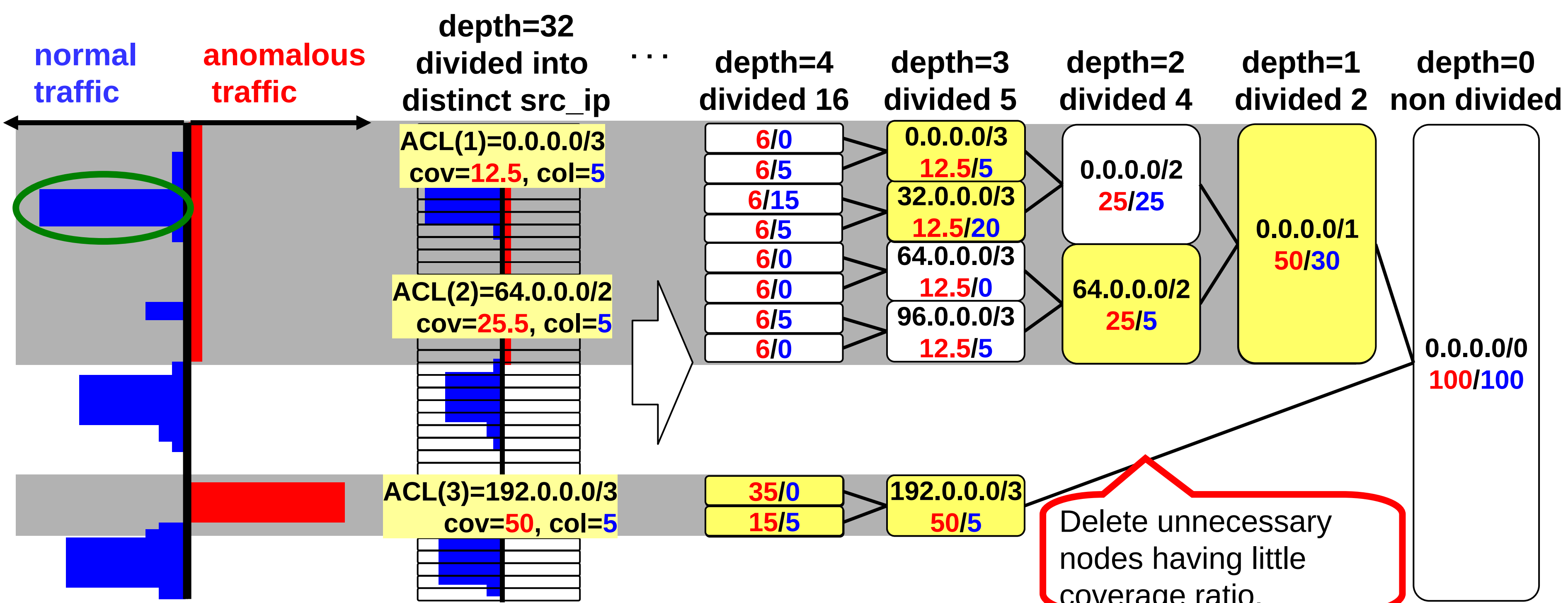
Step 2: (2) Building Tree of Normal and Anomalous Traffic

- **Traffic tree making**
  - Build up from individual source IP addresses (depth=32).
  - Each node has information about coverage and collateral ratio.
    - Collateral ratio: normal traffic of the node ÷ total normal traffic
    - Coverage ratio: anomalous traffic of the node ÷ total anomalous traffic
  - Make parent nodes by merging child node information.



Step 3: Selecting Best Node Sets (ACL sets)

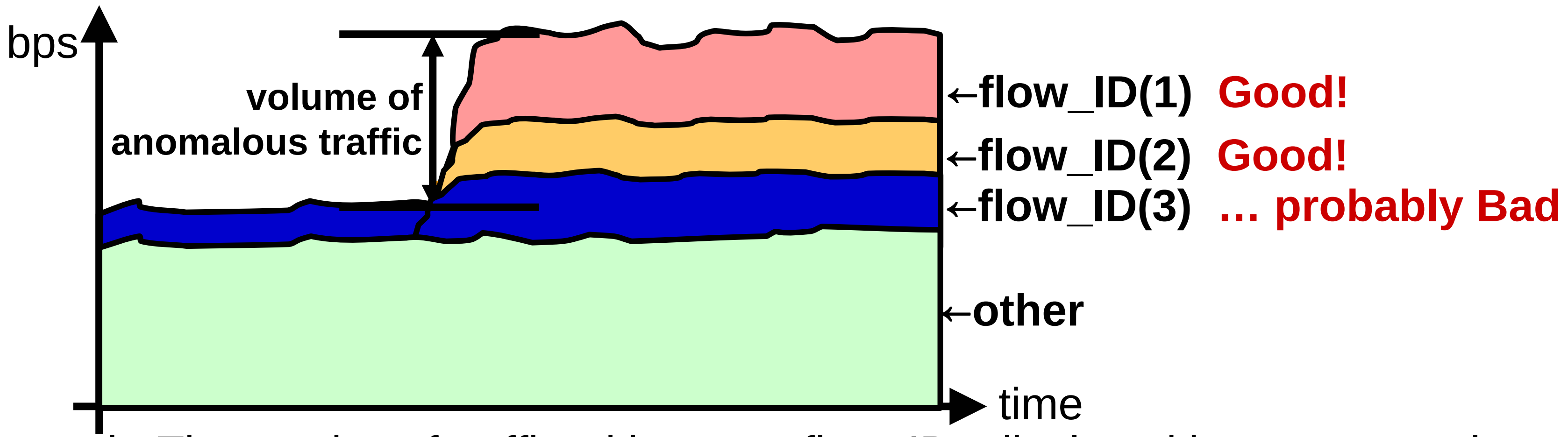
1. To reduce search space, delete unnecessary nodes.
   - Unnecessary node: node having little coverage ratio (little anomalous traffic) or little difference from its descendant nodes.
2. Search for best node sets by evaluating goodness of every node combination.
   - Best node combination = **Best ACL sets for source IP dimension**
- But, how to decide goodness of the node sets?



Example 4: Best node sets: Can filter almost all anomalous traffic with little collateral

Criteria of “Goodness”

- Three criteria of identification
  1. Coverage ratio:
     Maximize filtered anomalous traffic = (1 – FNR)
  2. Collateral (damage) ratio:
     Minimize filtered (normal) legitimate traffic = (FPR)
  3. **Number of ACLs**:
     ACL entry budget is limited, so having few ACLs is better.
- But, these three criteria have a trade-off relationship with each other.



Dummy graph: Time series of traffic with output flow_IDs displayed in separate colors

Evaluation Formula for Goodness

- To evaluate goodness of best ACL sets, we use the formula:

  coverage : cov, collateral ratio : $coll$, no. of ACLs : n

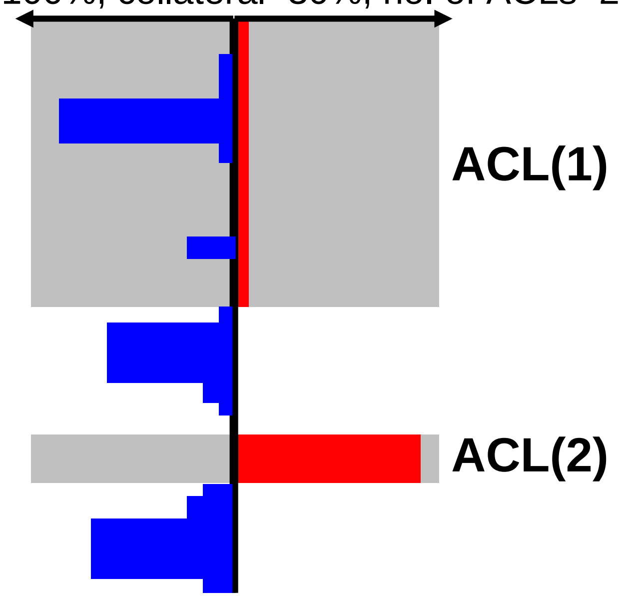
  $$rate = \frac{(\beta - \alpha) + \alpha \cdot cov - \beta \cdot coll}{n^{\gamma}} \quad (\alpha, \beta, \gamma \text{ : weighting coefficients})$$

  – Weighting coefficients can be tuned to reflect network policy or customer requirements.

ACL sets for covering all anomalous traffic

rate= 2.61

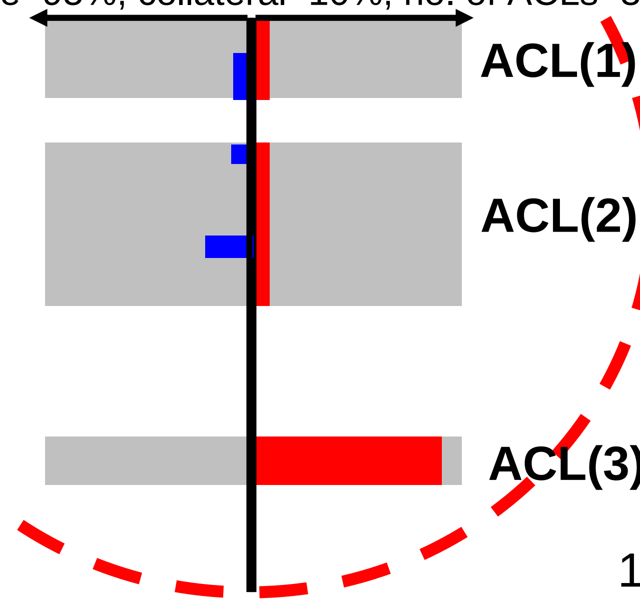
coverage=100%, collateral=30%, no. of ACLs=2



Example ACL splitting

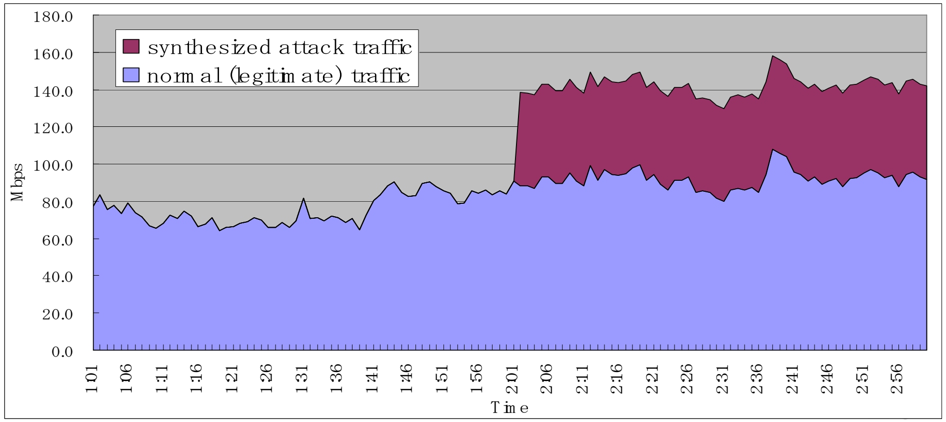
rate= 3.18

coverage=95%, collateral=10%, no. of ACLs=3



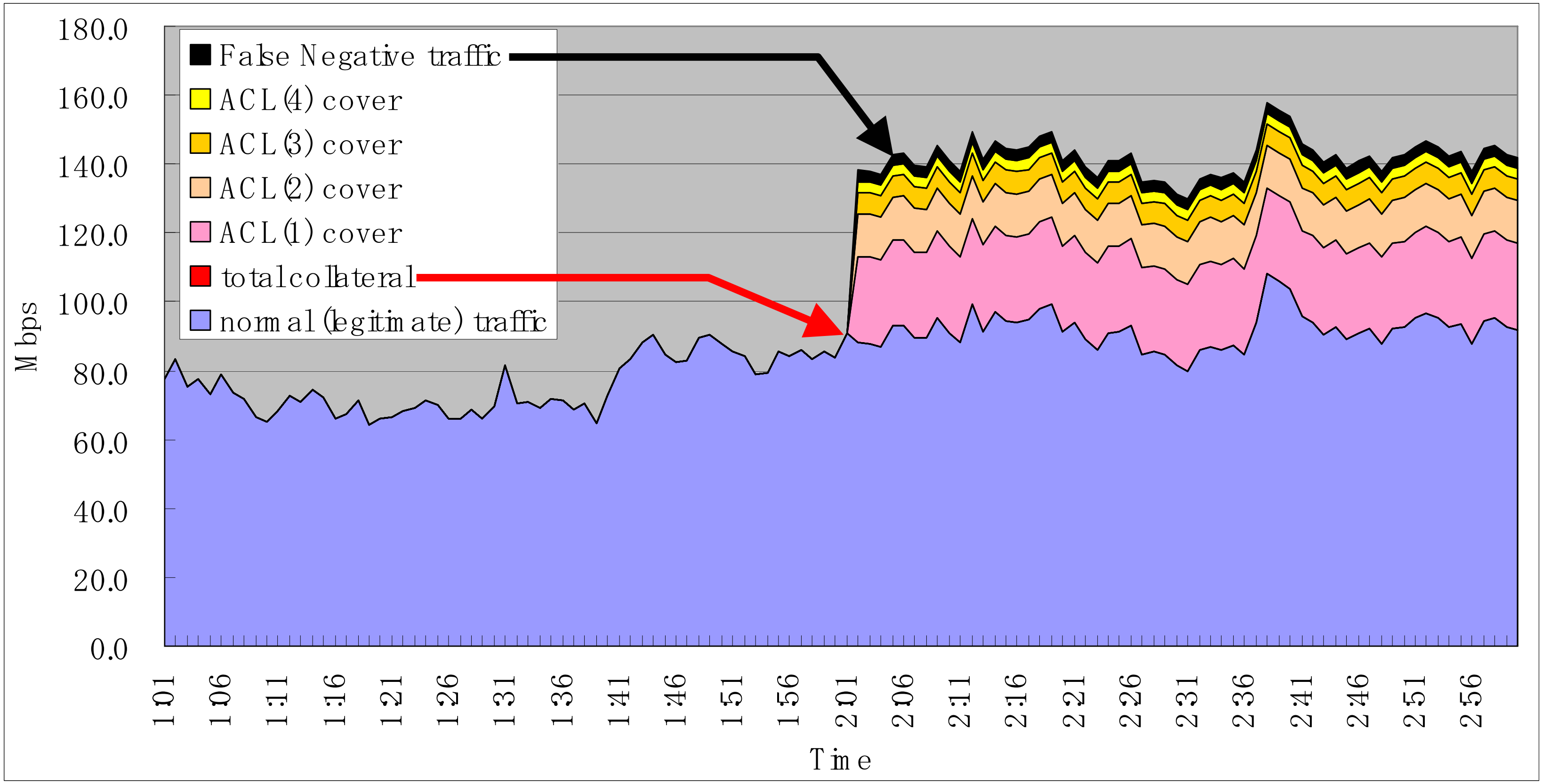
Evaluation and Results: Test Data Set

- Normal traffic: publicly available traffic data captured on transpacific line (100 Mbps)
- Anomalous traffic: injected synthesized DDoS attack traffic
  - Mimic large DDoS attack
    - We choose source/destination addresses that have large normal traffic because simple identification would cause collateral.
  - Destination: Popular server appeared in normal traffic
  - Source: Choose IP address blocks (/16) from which volume of normal traffic to the destination is largest.
  - Port numbers and protocol of attack traffic are the same as those of normal traffic.



Evaluation and Results: Results (1)

- Results: We get four ACL sets with below conditions
  - coverage: 93.75%
  - collateral: 0.00%
  - no. of ACL sets: 4



Evaluation and Results: Results (2) OUTPUT

basic information

basetime_len=	60.0 (sec) : (1168362060.0 - 1168362120.0)	
anomtime_len=	60.0 (sec) : (1168362180.0 - 1168362240.0)	
base_total_bps=	89,121,539.5	
anom_total_bps=	137,729,812.7	
diff_total_bps=	48,608,273.2	+54.5 %

single dimension identification results

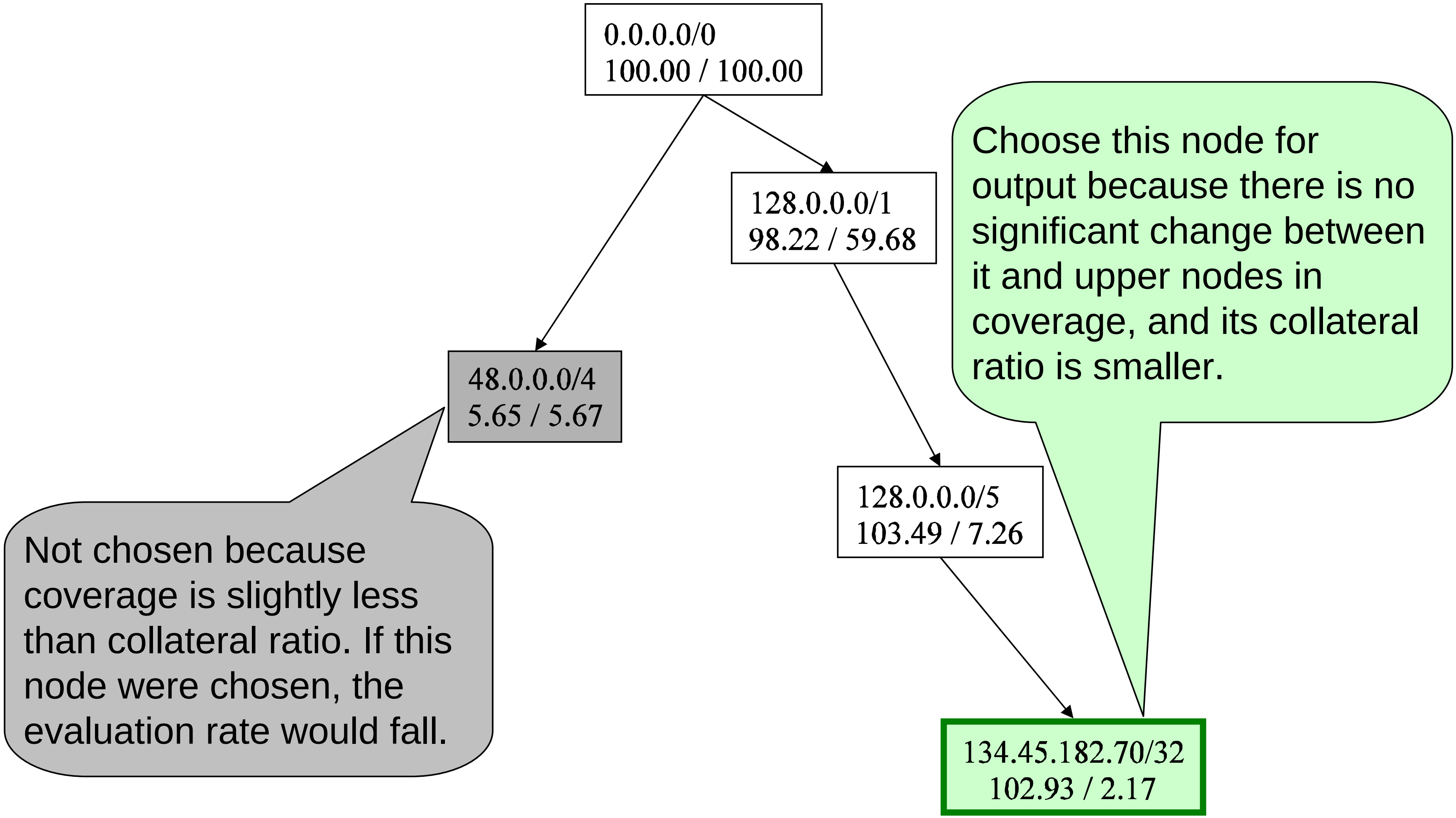
1-D_OUTPUT: PROTOCOL=	6	coverage= 100.42	collateral= 95.52
1-D_OUTPUT: SRC_PORT=	high	coverage= 108.27	collateral= 33.42
1-D_OUTPUT: DST_PORT=	high	coverage= 100.09	collateral= 96.40
1-D_OUTPUT: **SRC_IP**		coverage= 96.43	collateral= 0.00
	119.170.0.0/17	coverage= 51.43	collateral= 0.00
	119.170.128.0/18	coverage= 25.72	collateral= 0.00
	119.170.192.0/19	coverage= 12.86	collateral= 0.00
	119.170.240.0/20	coverage= 6.43	collateral= 0.00
1-D_OUTPUT: **DST_IP**		coverage= 102.93	collateral= 2.17
	134.45.182.70/32	coverage= 102.93	collateral= 2.17

MULTI-DIMENSION_FLOW_OUTPUT coverage= 96.43 collateral= 0.00

	↑ **coverage**	↑ **collateral**:	↑ src_ip	↑ dst_ip	protocl	scr_port	dst_port
flowID_0:	cov= 51.43	col= 0.00:	**119.170.0.0/17**	**134.45.182.70/32**	6	high	high
flowID_1:	cov= 25.72	col= 0.00:	**119.170.128.0/18**	**134.45.182.70/32**	6	high	high
flowID_2:	cov= 12.86	col= 0.00:	**119.170.192.0/19**	**134.45.182.70/32**	6	high	high
flowID_3:	cov= 6.43	col= 0.00:	**119.170.240.0/20**	**134.45.182.70/32**	6	high	high

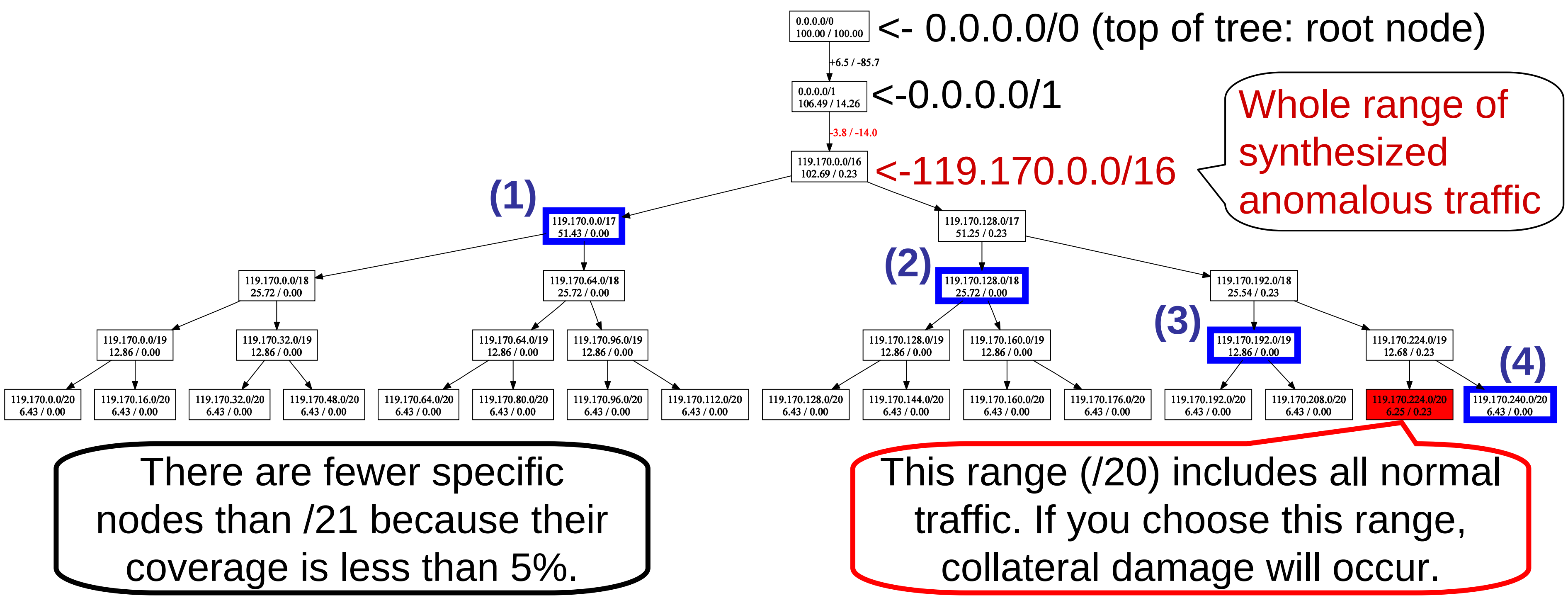
Evaluation and Results (3): Destination IP Tree

1-D_OUTPUT: **DST_IP** coverage= 102.93 collateral= 2.17
134.45.182.70/32 coverage= 102.93 collateral= 2.17



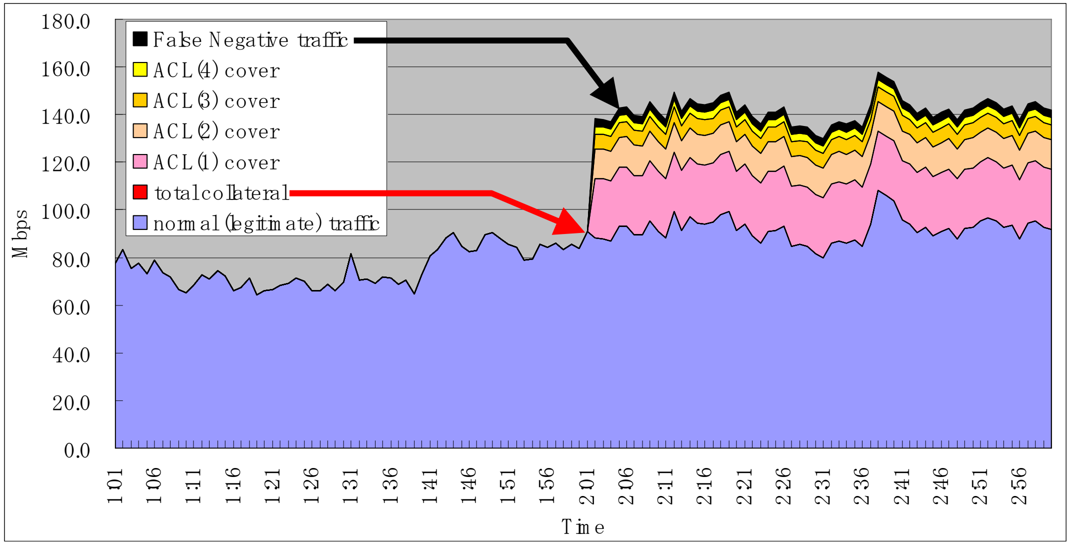
Evaluation and Results (4): Source IP Tree

1-D_OUTPUT: **SRC_IP**		coverage= 96.43	collateral= 0.00
(1)	**119.170.0.0/17**	coverage= 51.43	collateral= 0.00
(2)	**119.170.128.0/18**	coverage= 25.72	collateral= 0.00
(3)	**119.170.192.0/19**	coverage= 12.86	collateral= 0.00
(4)	**119.170.240.0/20**	coverage= 6.43	collateral= 0.00



Summary

- Revealed three criteria of optimal ACL sets.
  - for mitigating DDoS attacks on router
- Proposed DELTAA technique: Optimizes trade-off among the these criteria, using normal and anomalous traffic.
- Showed effectiveness of DELTAA.
  - Evaluation results using prototype and synthesized data sets:
    - coverage: 93.75%
    - collateral: 0.00%
    - no. of ACL sets: 4



Thank you.

Any questions are welcome.

This study was supported by
the Ministry of Internal Affairs and Communications of Japan.