EXA48200 & EXA48800 Deduplication

Application Note

March 2026

Author: Herbert Etlinger

TABLE OF CONTENTS

Deduplication Background

Tapping multiple points across a network can cause the same packet to be copied multiple times, particularly for East/West traffic and SPAN port traffic. This can negatively affect the performance of various tools, as they not only need to process the same traffic multiple times but, in many cases, also expend additional processing resources performing deduplication themselves. In some cases, duplicate packets can cause false positives or skew reporting.

Performance resilience is the ability to ensure appliance performance in any data centre environment. Eliminating duplicate traffic is critical for improving network performance. It is critical for performance monitoring, cybersecurity, and forensics appliances. Duplicates can cause significant issues. The obvious issue is that doubling the data requires doubling the processing power, memory, etc.

However, the main issue is false positives: errors that are not really errors or threats that are not really threats. One common way that duplicates affect analysis is by an increase in TCP out-of-order or retransmission warnings. Debugging these issues takes a lot of time, usually time that an overworked, understaffed network operations or security team does not have. In addition, any analysis performed on the basis of this information is probably not reliable, so this only exacerbates the issue.
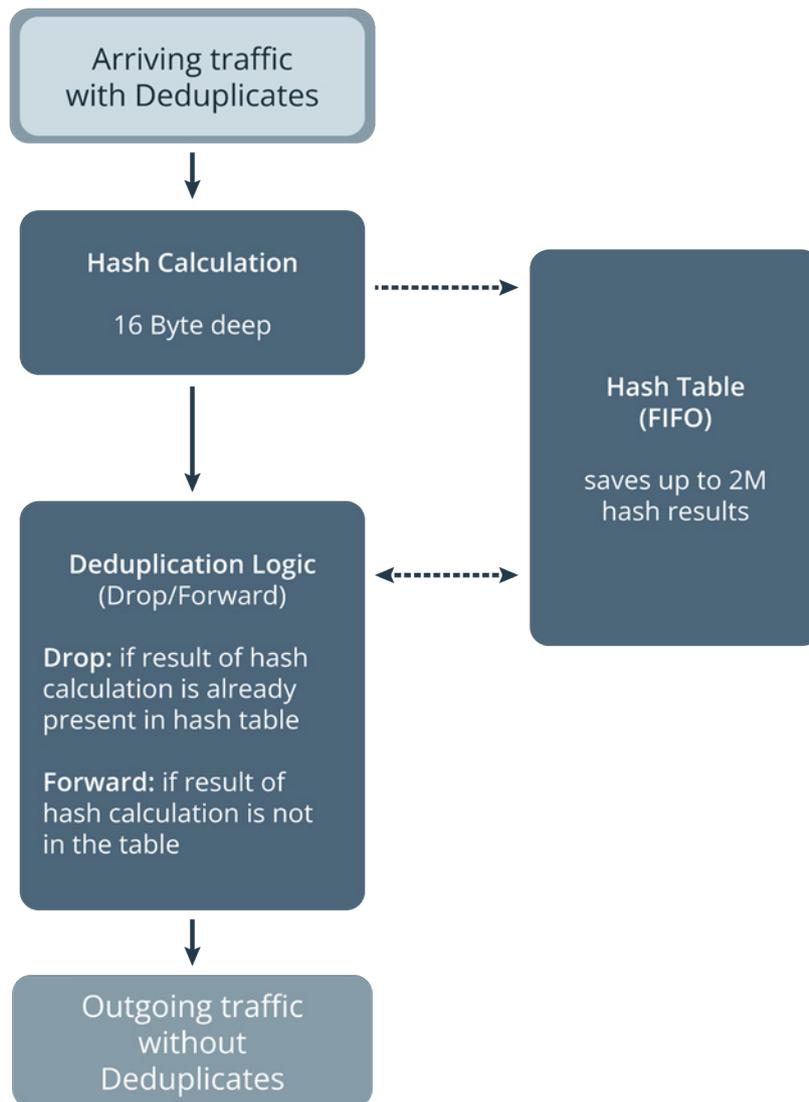
The Cubro EXA48200 and EXA48800 offer a deduplication function to eliminate duplicate packets, thereby protecting monitoring equipment from overload.
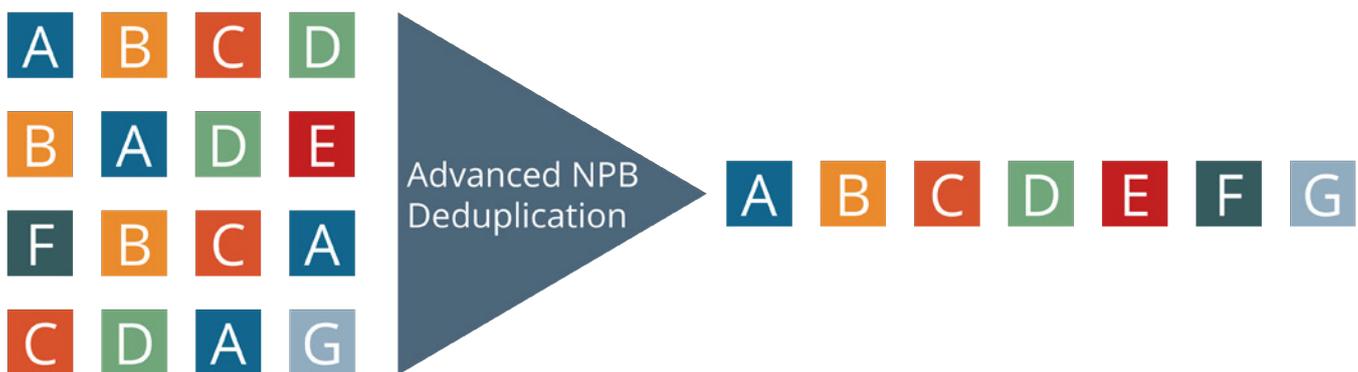
Deduplication on Cubro EXA48200 & EXA48800

Cubro provides several options for deduplication, configurable by user-defined parameters, before forwarding traffic to monitoring tools. Offloading this resource-intensive task frees a tool's processing resources, allowing it to perform the tasks more efficiently they were designed to do.

When the deduplication function is switched on, EXA48200/EXA48800 will calculate a checksum (=hash- key) over the full packet. A full packet uses every bit, including the payload, for the hash-key calculation. The hash-key is an MD5 calculation, and the result is 16 byte long. Hash keys are stored, and every incoming packet (= one hash key per packet) is checked in the memory. Once a hash key is found in memory, the packet with the same hash key is dropped.

```
┌──────────────────────┐
│   Arriving traffic    │
│   with Deduplicates   │
└──────────────────────┘
           │
           ▼
┌──────────────────────┐            ┌──────────────────────┐
│   Hash Calculation    │ - - - - ▶ │                      │
│                      │            │     Hash Table        │
│     16 Byte deep      │            │      (FIFO)          │
└──────────────────────┘            │                      │
           │                         │   saves up to 2M      │
           ▼                         │   hash results        │
┌──────────────────────┐            │                      │
│  Deduplication Logic  │ ◀ - - - ▶ │                      │
│    (Drop/Forward)     │            └──────────────────────┘
│                      │
│  Drop: if result of   │
│  hash calculation is   │
│  already present in    │
│  hash table           │
│                      │
│  Forward: if result of │
│  hash calculation is not│
│  in the table          │
└──────────────────────┘
           │
           ▼
┌──────────────────────┐
│   Outgoing traffic    │
│       without         │
│     Deduplicates      │
└──────────────────────┘
```

Hash keys are aged out after a certain time. This time is called the detection window. The detection window for the EXA48200/EXA48800 is 1000ms. Thus, when a duplicate packet arrives within 1000ms, the Cubro Advanced NPB will find it and drop it. If the duplicate packet arrives after the 1000ms window, the duplicate is not found.

In addition to the detection window, the hash table size is important. The EXA48200 & EXA48800 support a hash-key storage of 4M values. When the hash table is getting full, it is aging out (FIFO principle). The size of the hash table also decides the performance limits.

Packet Size in Byte	Traffic Utilization in Gbit/s	Packets per second
64	10	14 880 952
300	10	3 906 250
500	10	2 403 846
1000	10	1 225 490
1518	10	763 125

Small packets, such as 64 bytes, fill the hash table much faster than 1518-byte packets. Typical performance limit for deduplication is about 70Gbit/s for the EXA48200 and 140Gbit/s for the EXA48800, having in mind usual traffic patterns and packet size distributions.

Deduplication Operation

Setup of Deduplication is straight-forward and easy. In case the traffic arrives at 10G port 1 (X1) and should be sent out at 10G port 2 (X2) it requires following steps:

a)   Create a Forwarding Policy via drag and drop from X1 to X2
b)   Enable Deduplication on X1
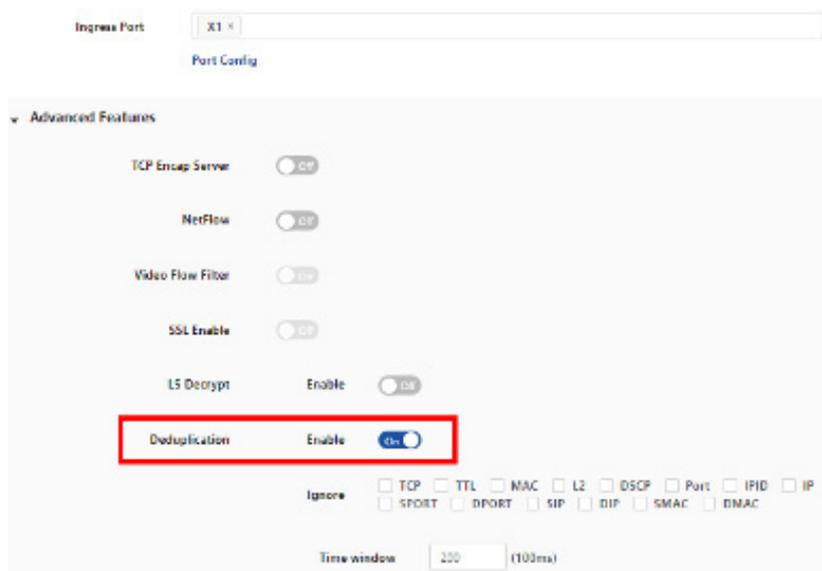
Another important characteristic of deduplication on the EXA48200 & the EXA48800 is that it allows for blanking out certain fields in the packet.

By default, the EXA48200 & the EXA48800 compare full packets and drop a packet if it is 100% identical.

Another important characteristic of the EXA48200 & the EXA48800 is that it allows to blank-out certain fields in the packet. By default, the EXA48200 & the EXA48800 compare full packets and drop a packet if it is 100% identical. In certain circumstances, it might happen that a packet is still a duplicate in terms of payload, but the header might look different. This situation is very common when dealing with mirror ports, when, e.g., ingress mirroring has a single VLAN while egress mirroring has two VLAN tags.

To cope with this situation and still be able to find the duplicates, the EXA48200 & the EXA48800 allow blanking out the following fields from the hash calculation:

- TCP... ignores Seq/Ack in the TCP header
- TTL... ignores Time to Live
- MAC... ignores MAC Source and MAC Destination
- L2... ignores all Layer 2, including MAC and VLANs
- DSCP... ignores Layer Quality of Service
- Port... ignores the input port
- IP... ignores the full IP header
- SPORT... ignores L4 Source Port
- DPORT... ignores L4 Destination Port
- SIP... ignores Source IP
- DIP... ignores Destination IP
- SMAC... ignores Source MAC
- DMAC... ignores Destination MAC

Ignore	☐ TCP ☐ TTL ☐ MAC ☐ L2 ☐ DSCP ☐ Port ☐ IPID ☐ IP ☐ SPORT ☐ DPORT ☐ SIP ☐ DIP ☐ SMAC ☐ DMAC
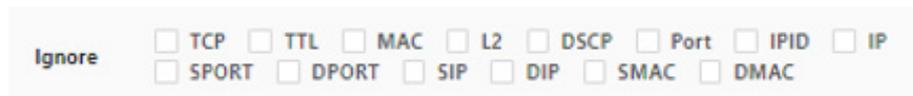
Deduplication Statistics

Statistics give you a good indication of how many duplicates have been found and eliminated.

Interface Statistics

sFlow Statistics · Deduplication Statistics

Interface Statistics: X1

Same_deduplication_packet: 1   Handle_deduplication_packet: 1048

Summary

The EXA48200 & the EXA48800 are ideal tools for deduplication, helping protect monitoring equipment from overload caused by duplicate packets. These products are the latest generation of Advanced Network Packet Broker based on a newly designed ASIC and ARM CPUs. The EXA48200 & the EXA48800 products provide maximum flexibility by combining a classic ASIC-based NPB with high-performance CPU(s), enabling them to run powerful applications.

Note: This document applies to EXA48200 & EXA48800 Advanced NPB. For any further queries, please contact **support@cubro.com**.