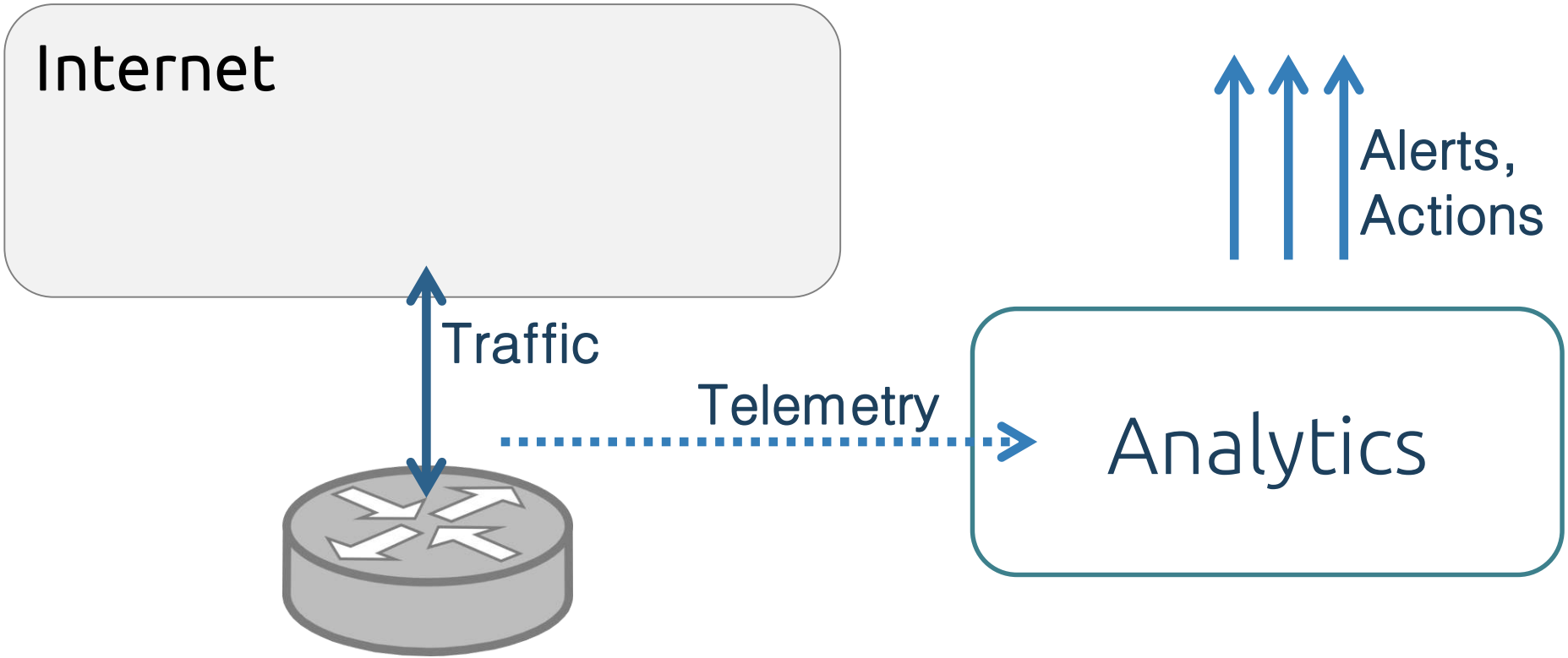
<epam>

Data Science for Network Security

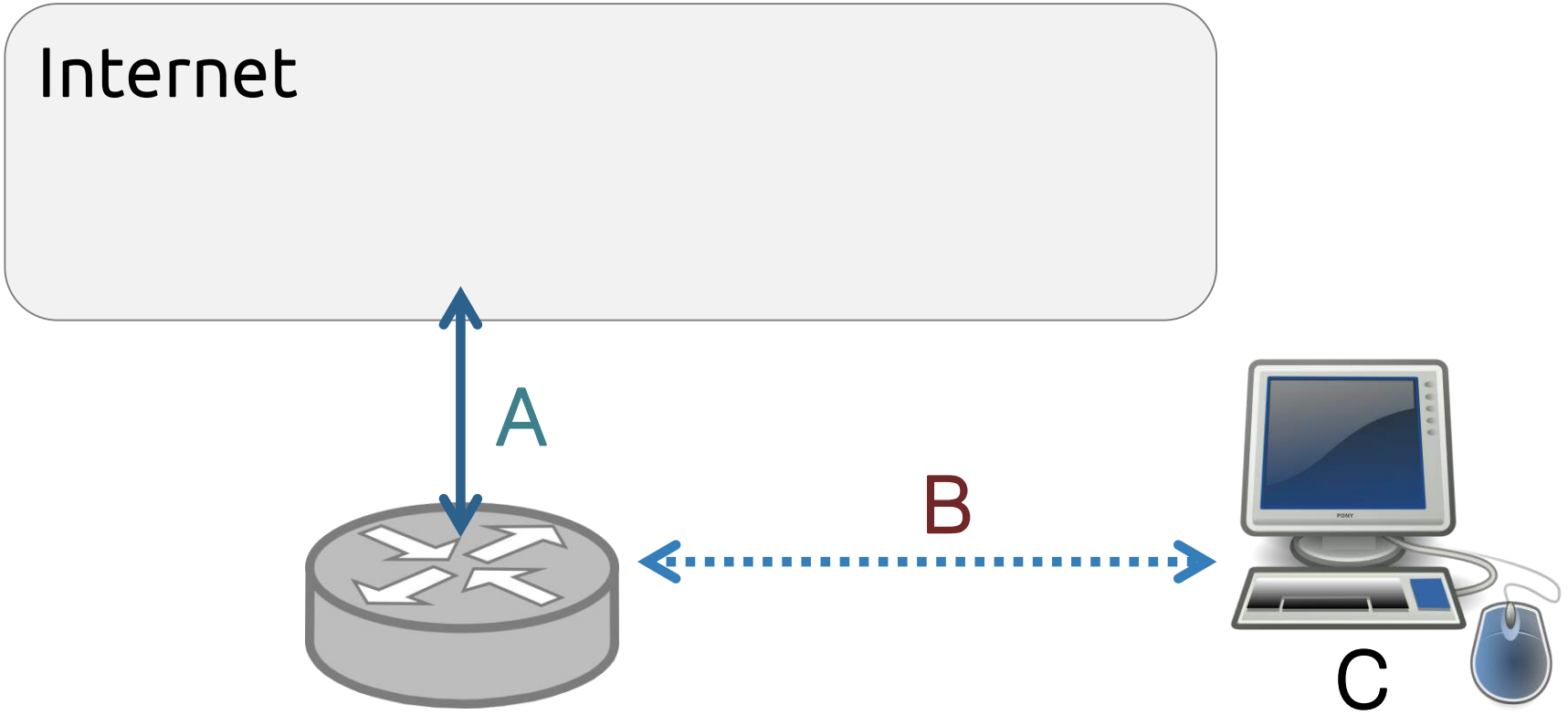
Dmitry Orekhov

Collecting Data

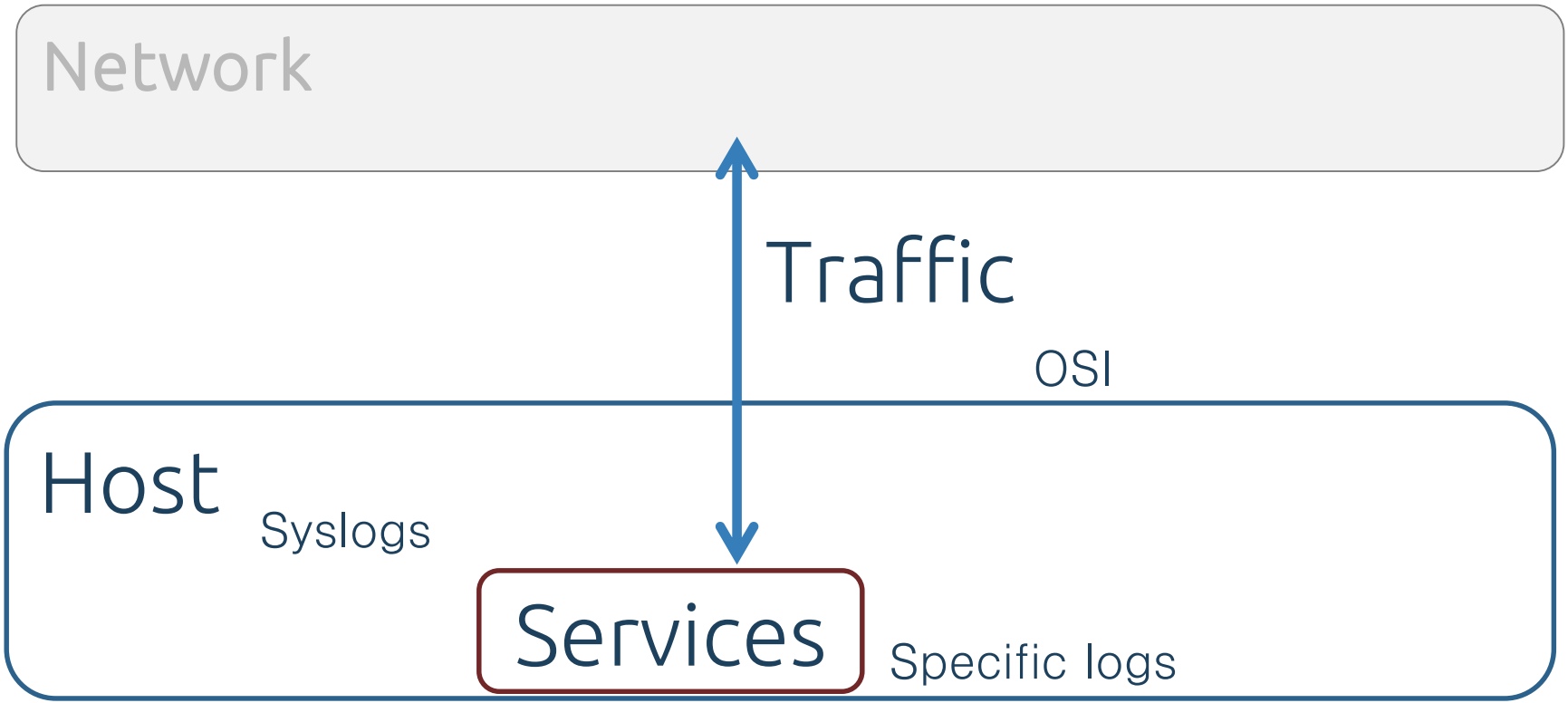
Data Flow



Sensors: Vantage

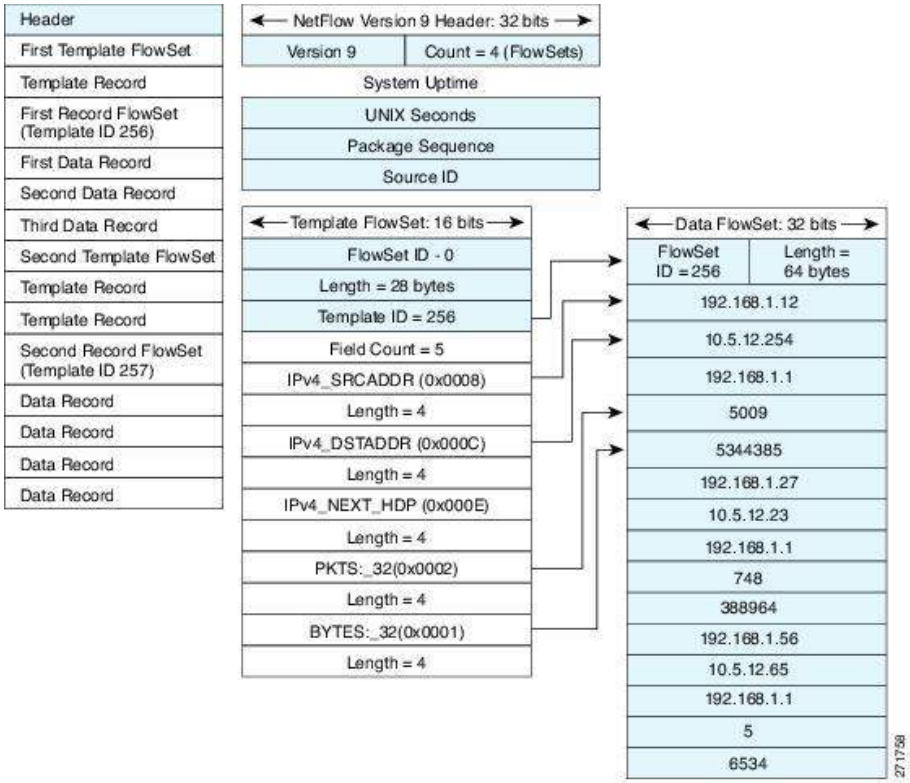


Sensors: Domain



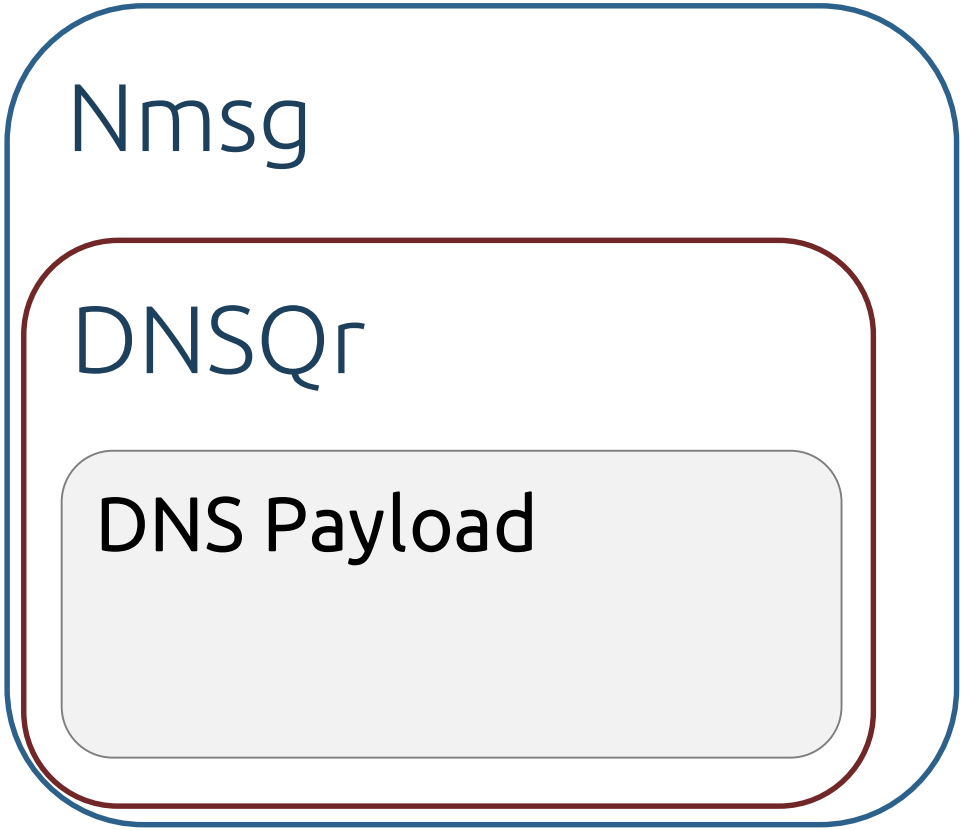
Transport: NetFlow

- NetFlow is a traffic summarization standard developed by Cisco Systems and originally used for network services billing.
- The heart of NetFlow is the concept of a flow, which is an approximation of a TCP session.
- Plenty of Open Source implementations

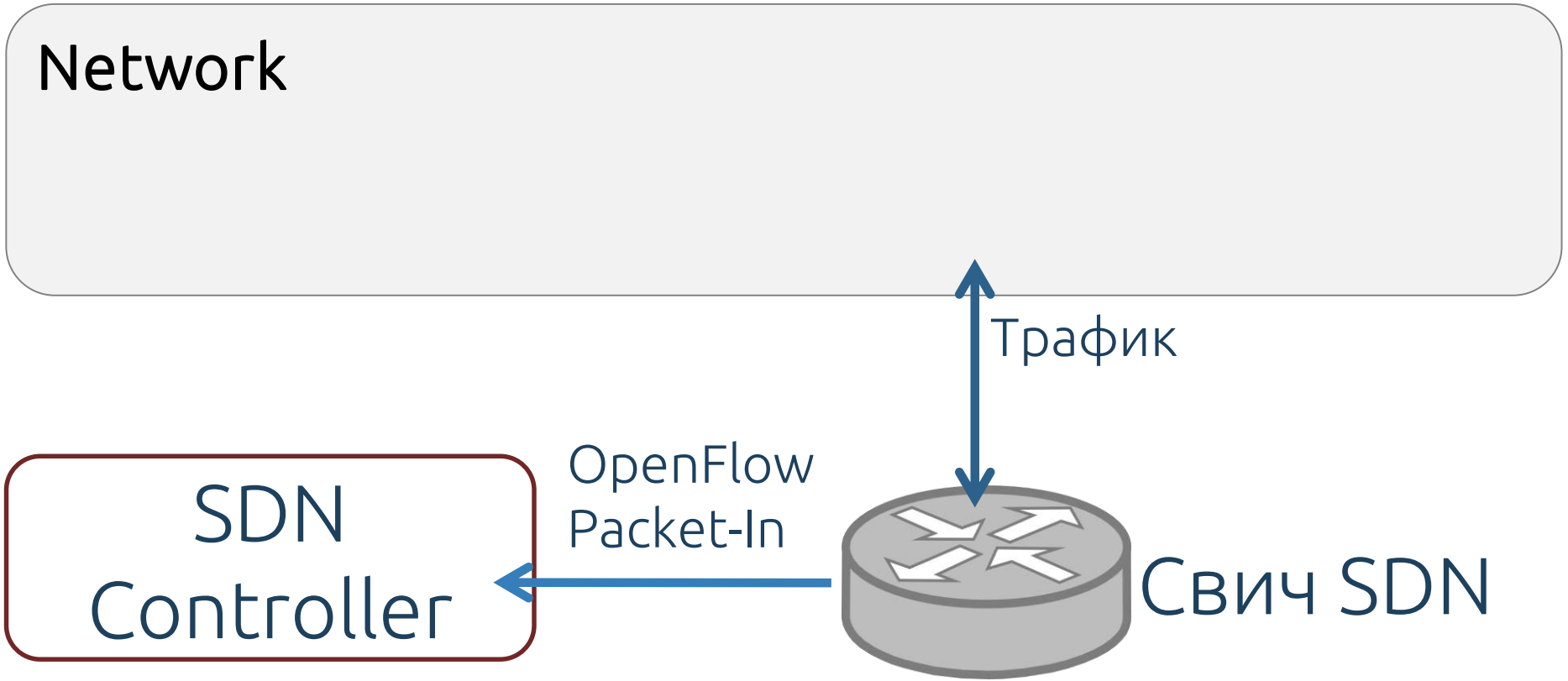


Transport: Nmsg

- Developed by Farsight Security for transporting network packets, particulary – DNS
- Low-latency and compactness
- Support binary and presentation forms; protobuf for protocols
- Open impelentation (GNU)



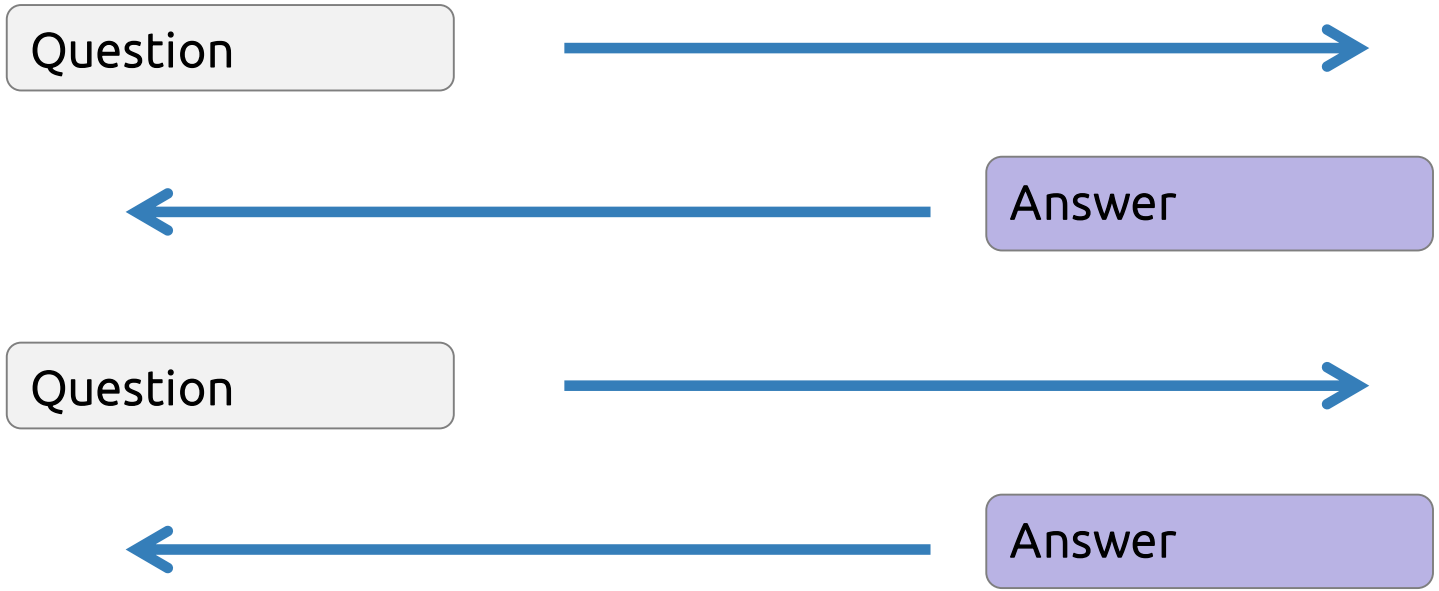
Transport: OpenFlow



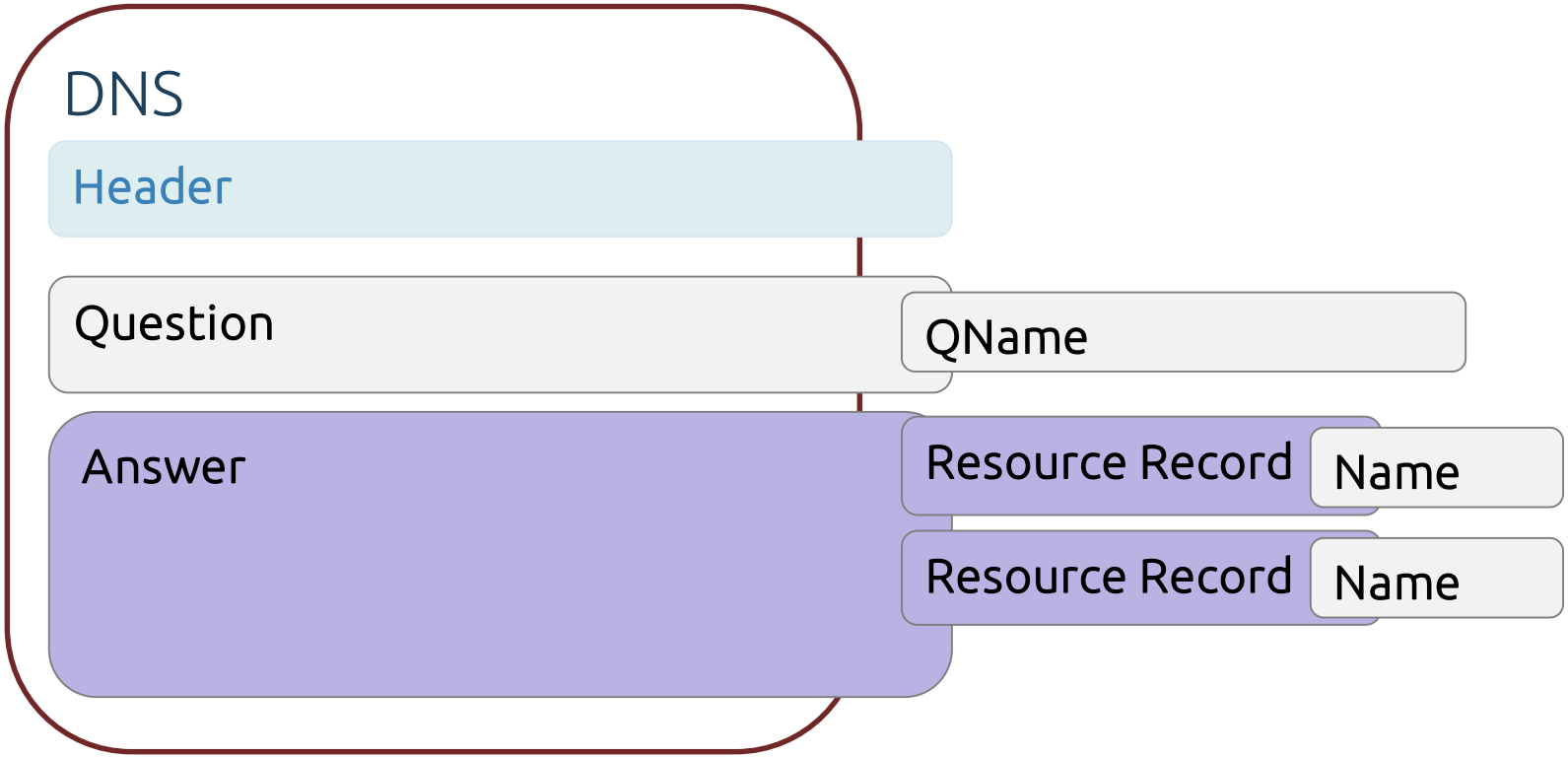
Use Case

DNS Tunneling

DNS



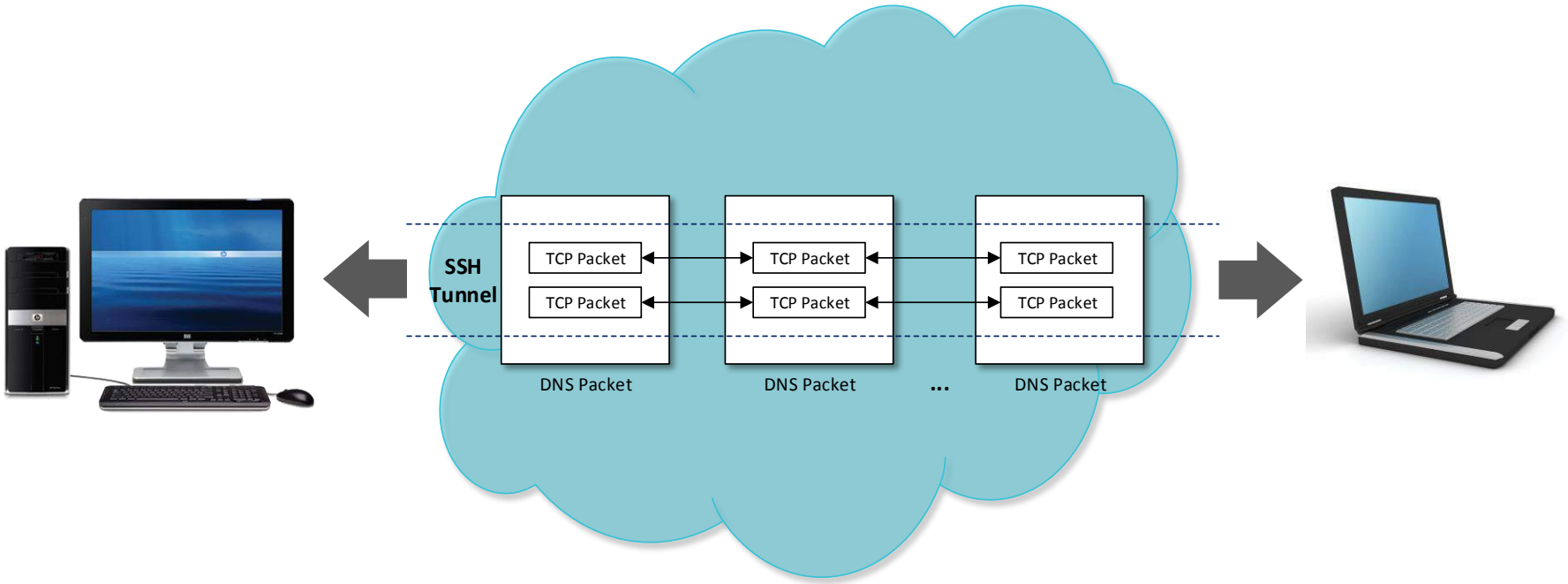
DNS Message: Big Picture



Base32

0	A	8	I	16	Q	24	Y
1	B	9	J	17	R	25	Z
2	C	10	K	18	S	26	2
3	D	11	L	19	T	27	3
4	E	12	M	20	U	28	4
5	F	13	N	21	V	29	5
6	G	14	O	22	W	30	6
7	H	15	P	23	X	31	7

Here you are: Tunnel



Discovery methods

Payload analysis

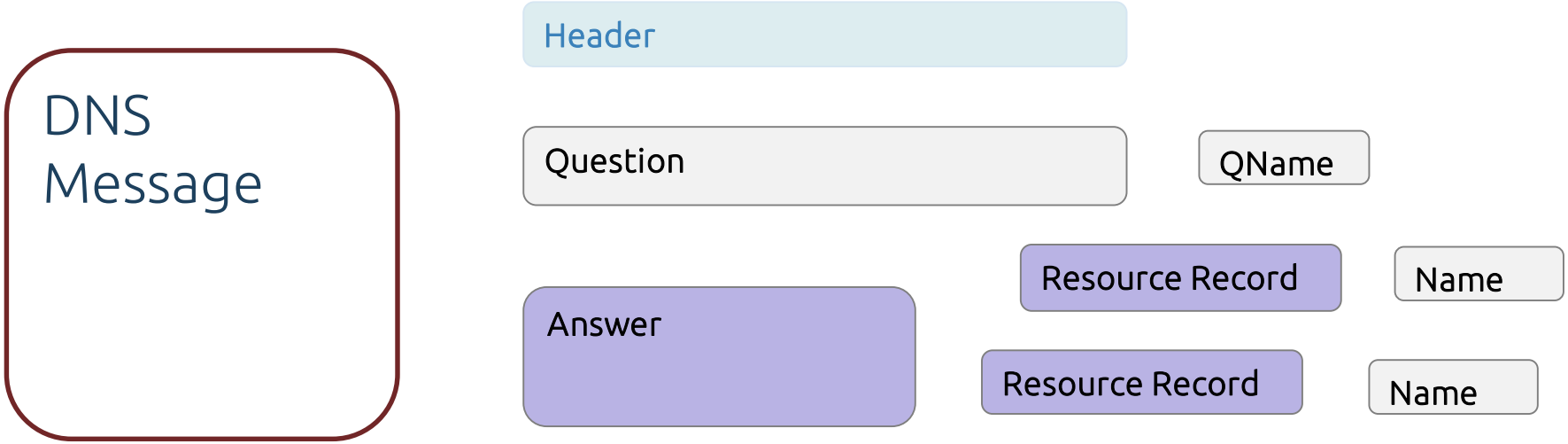
- Size of request and response
- Entropy of hostnames
- Statistical Analysis
- Uncommon Record Types

Traffic Analysis

- Volume of DNS traffic per IP address
- Volume of DNS traffic per domain
- Number of hostnames per domain
- Geographic location of DNS server
- Domain history
- Orphan DNS requests

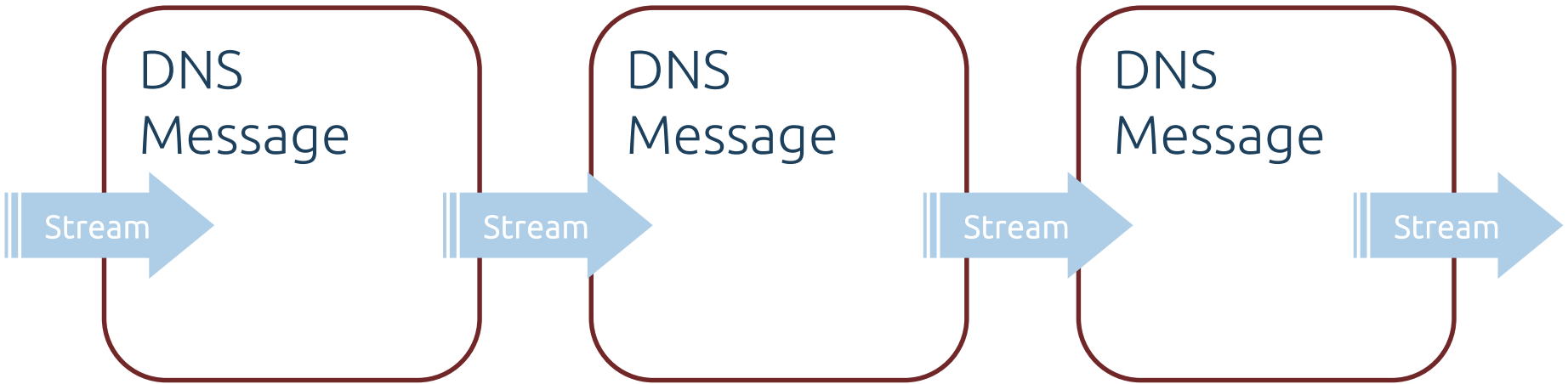
Effective Intrusion Detection

… needs deep packet inspection



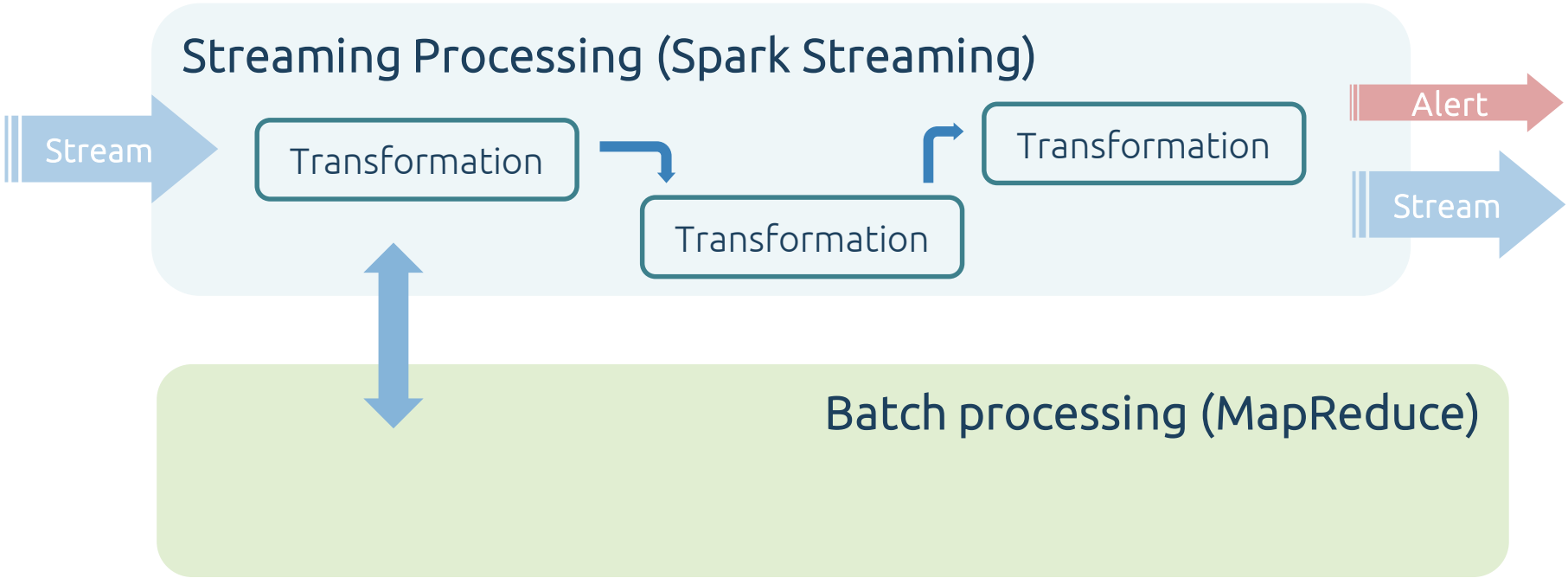
… and

… low-latency



Solution:
Lambda Architecture

Online model



Algorithms

Incremental algorithms

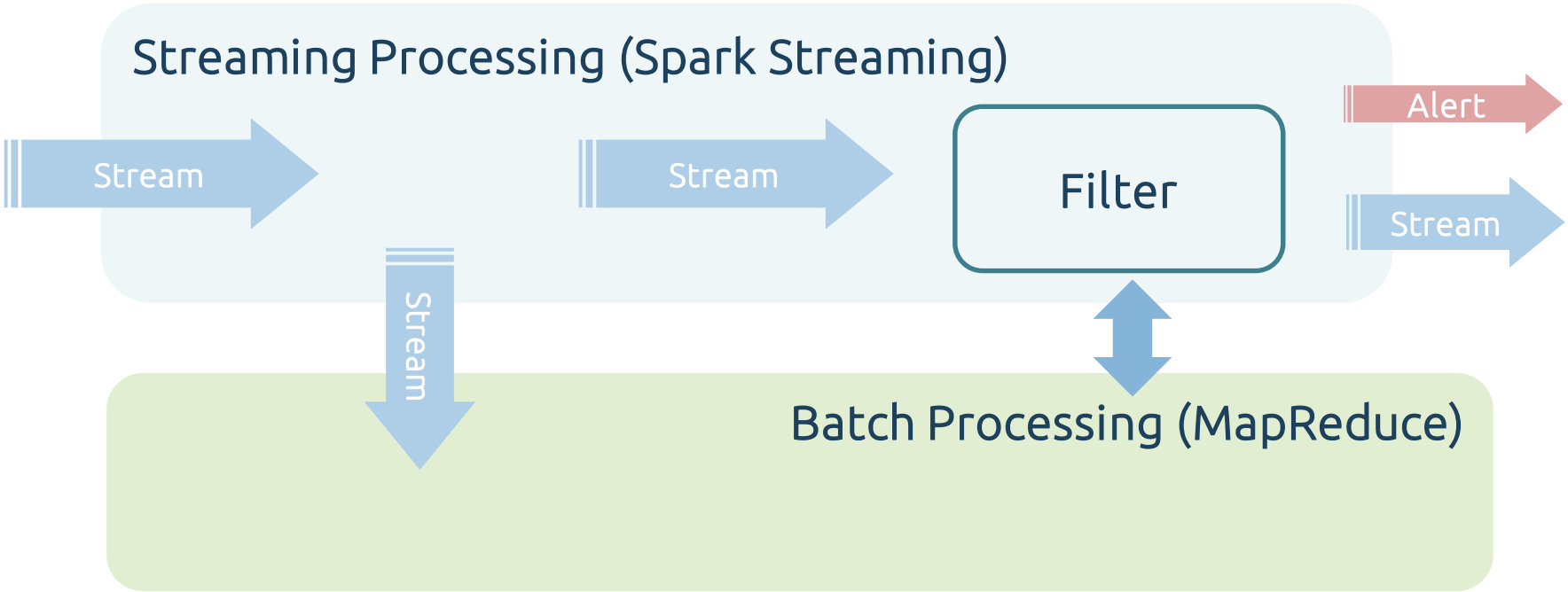
Outlier Detection

- Median Absolute Deviation, MAD
- Standard Deviation from average
- Standard Deviation from Moving Average

Потоковая классификация

- Incremental decision tree
- Hoeffding Tree (VFDT)
- Half-Space Trees

Offline model



Offline Model – Алгоритмы

Analyze entire data set at once

Hypothesis tests

- Simple outlier detection far a period
- Statistical criteria
- Kolmogorov-Smirnov test

Decision Trees

Auto Regressive (AR) Moving Average (MA)

Architecture