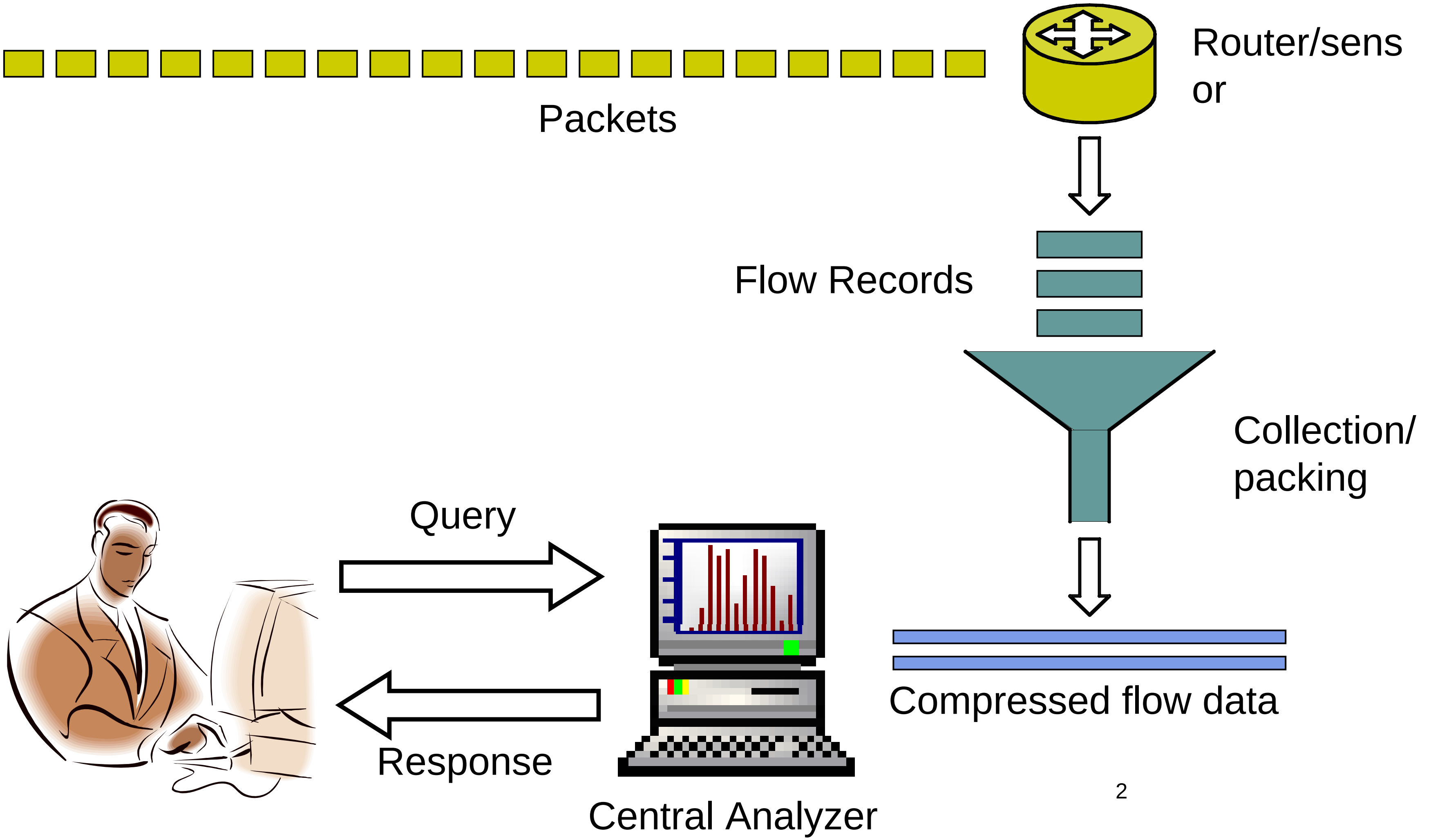
Scalable Flow Analysis

Abhishek Kumar
Sapan Bhatia
Georgia Tech

Flow Collection and Analysis Architecture



Packets

Router/sensor

Flow Records

Collection/ packing

Compressed flow data

Query

Response

Central Analyzer

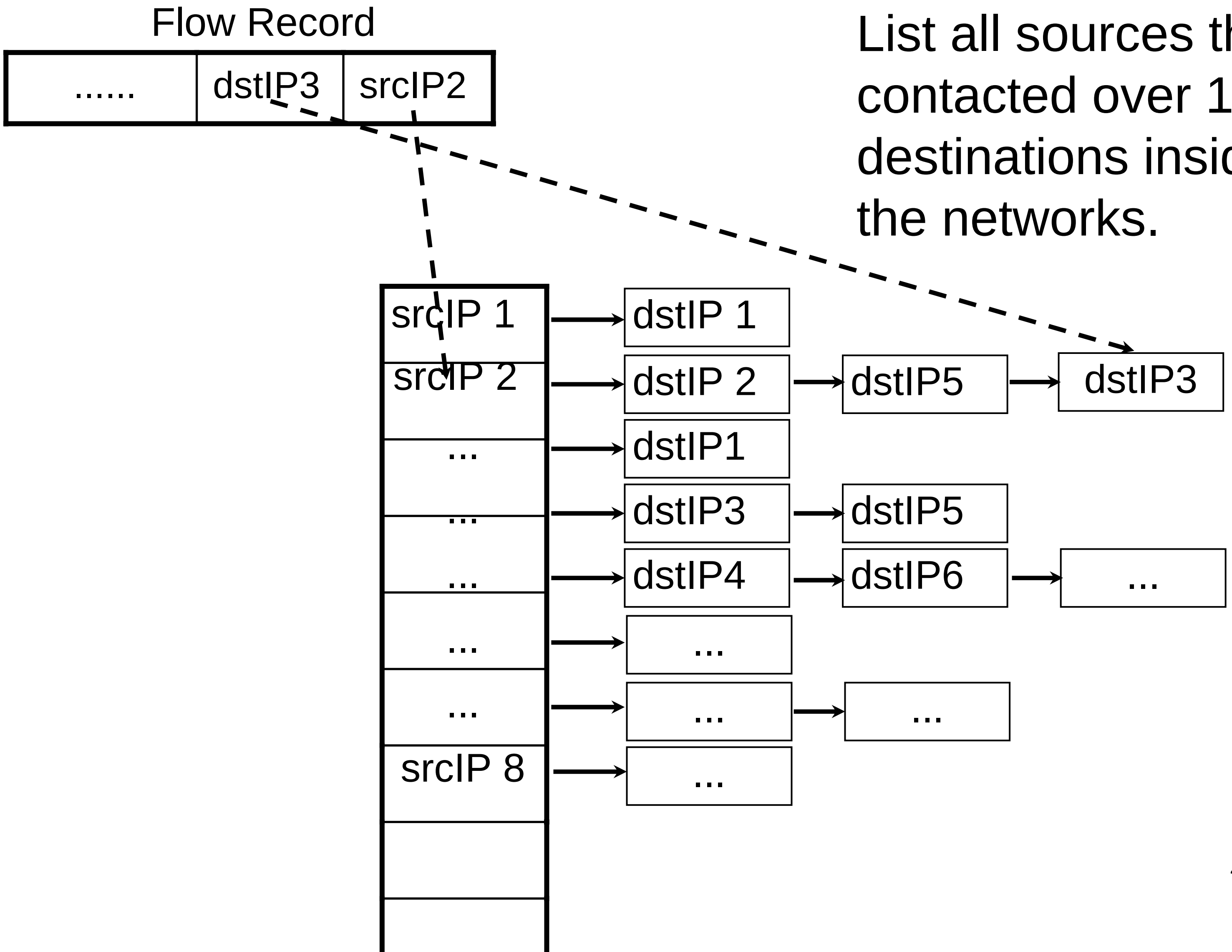
An informal Taxonomy

[illegible]

[illegible] 1,2,5,10 ... ?

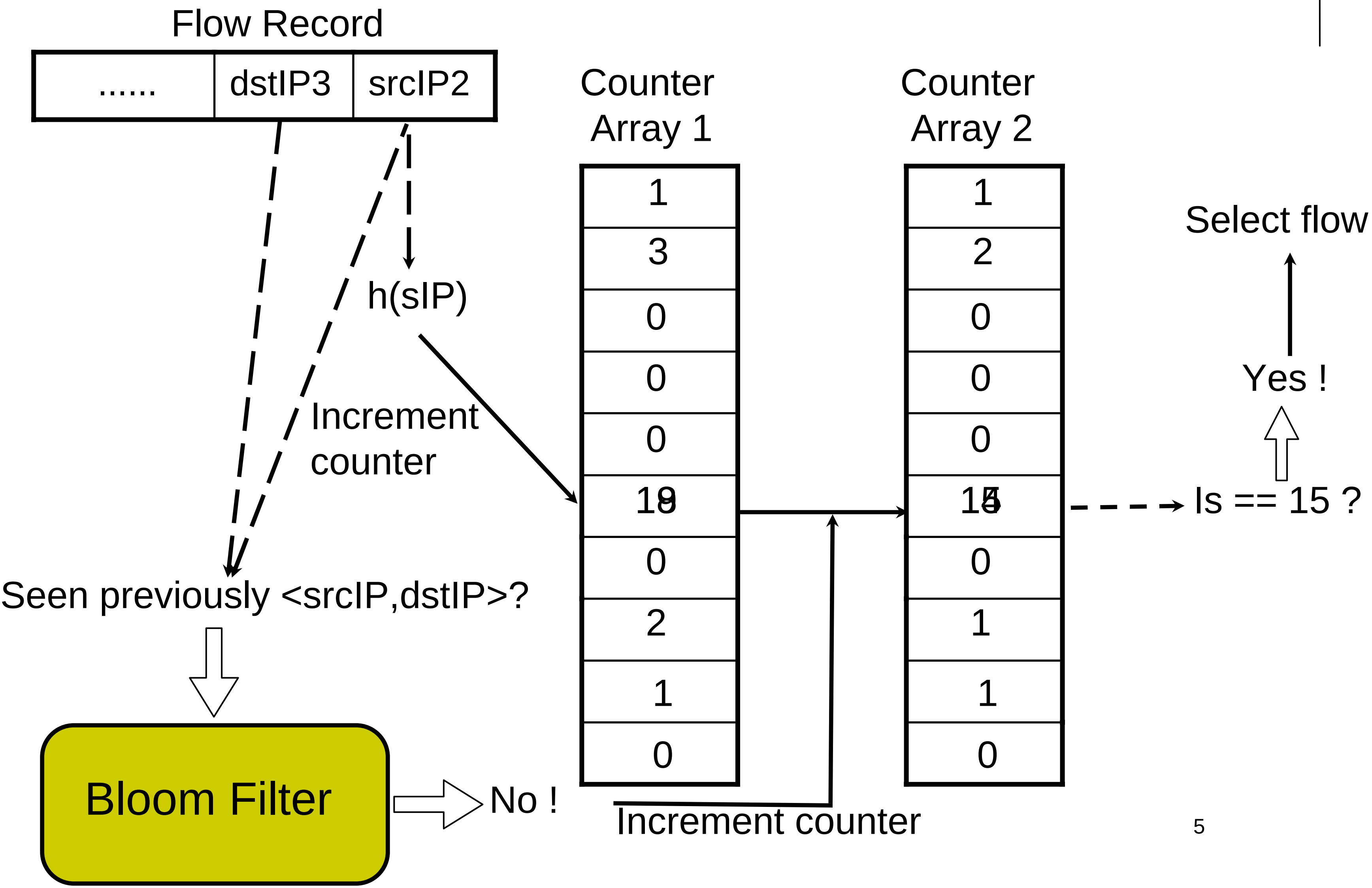
	Aggregate Query	Distributional Query	Identity Query
Routine Query			
Drill-down (forensic) Query			

An example query



List all sources that contacted over 15 destinations inside the networks.

Proposed Solution (Preprocessing)



Flow Record

......	dstIP3	srcIP2

h(sIP)

Increment counter

Counter Array 1	Counter Array 2
1	1
3	2
0	0
0	0
0	0
19	15
0	0
2	1
1	1
0	0

Increment counter

Is == 15 ?

Yes !

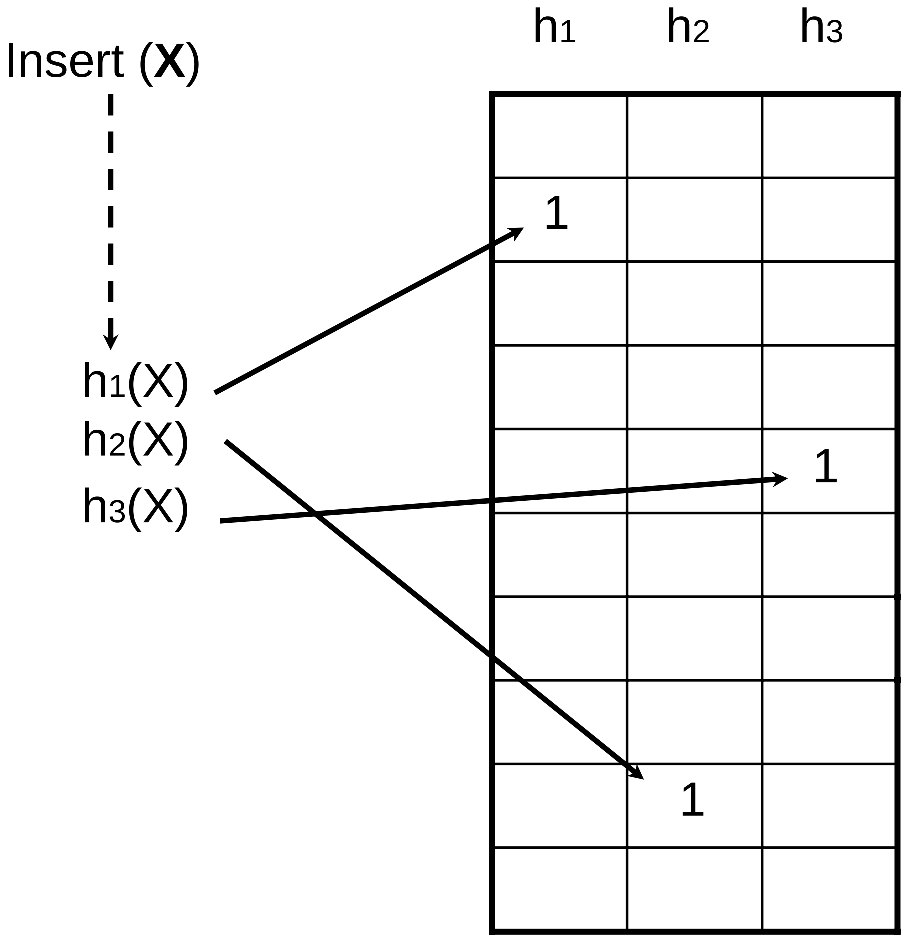
Select flow

Seen previously <srcIP,dstIP>?

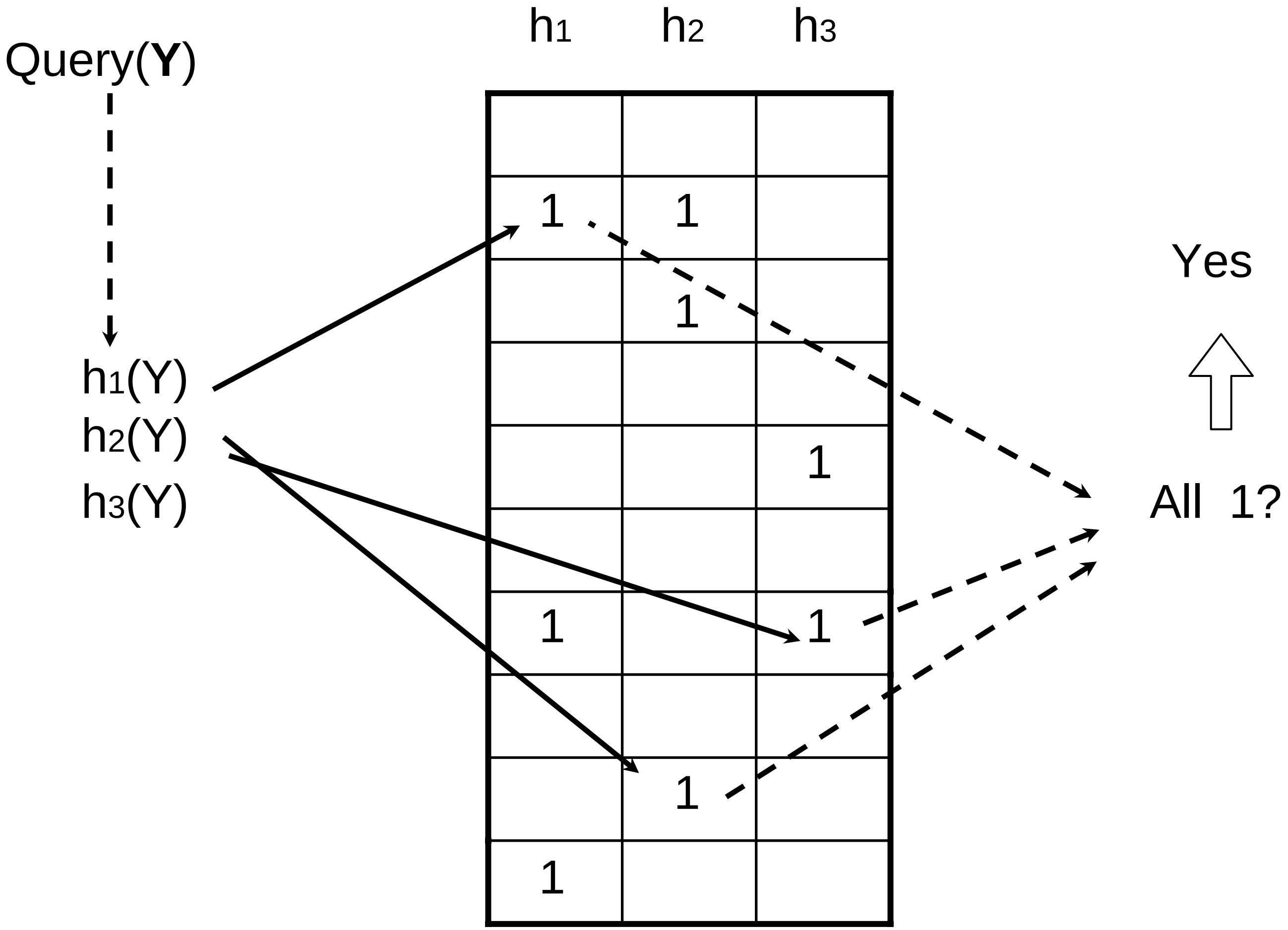
Bloom Filter

No !

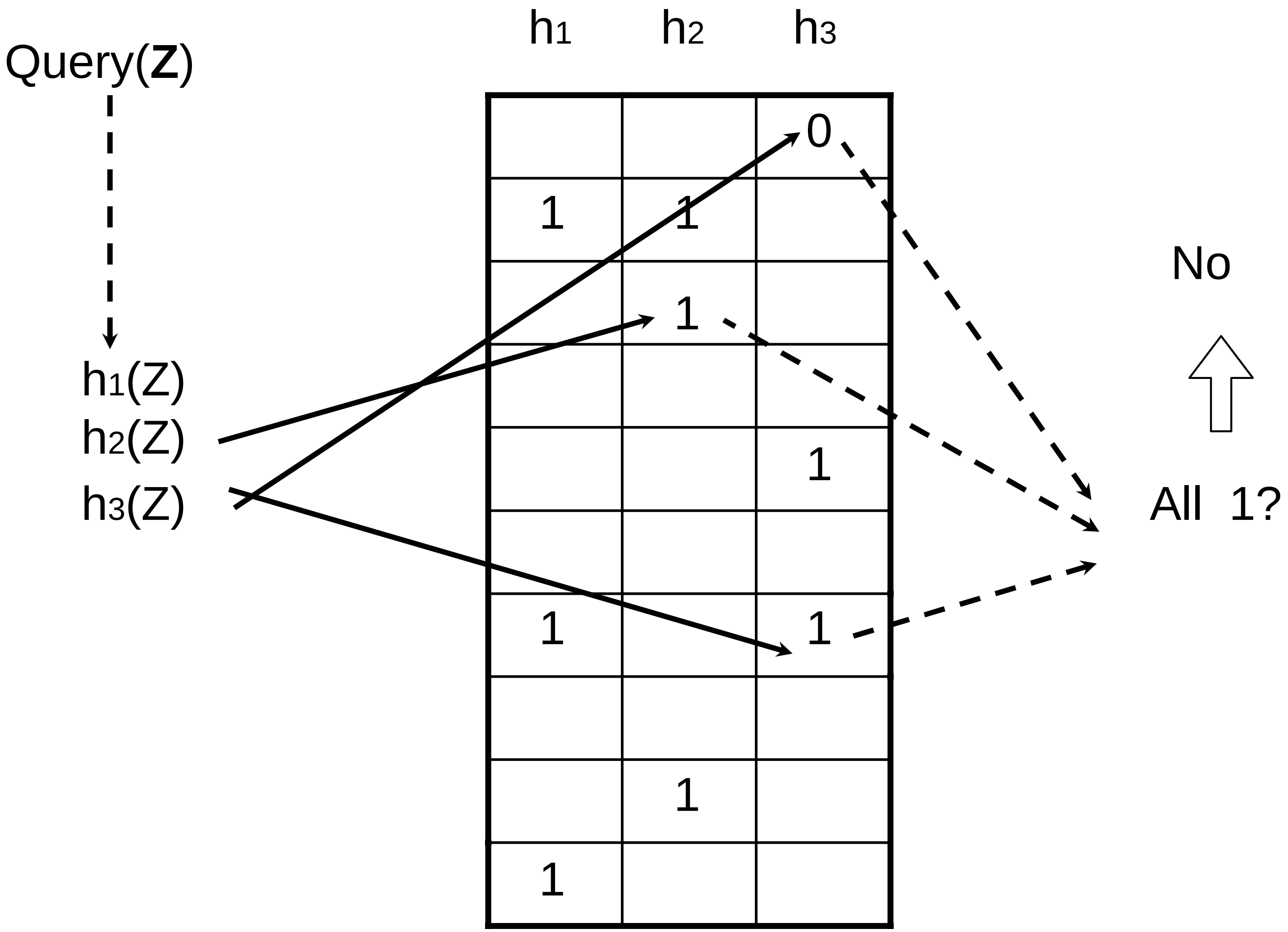
Bloom filter (insert)



Bloom filter (query)



Bloom filter (query)

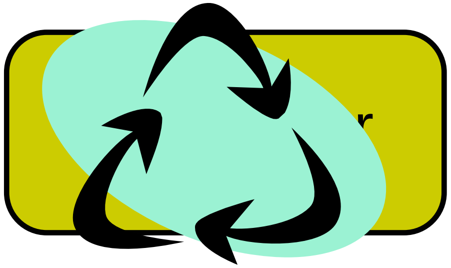


h_1	h_2	h_3
		0
1	1	
	1	
		1
1		1
	1	
1		

Proposed Solution (State after preprocessing)

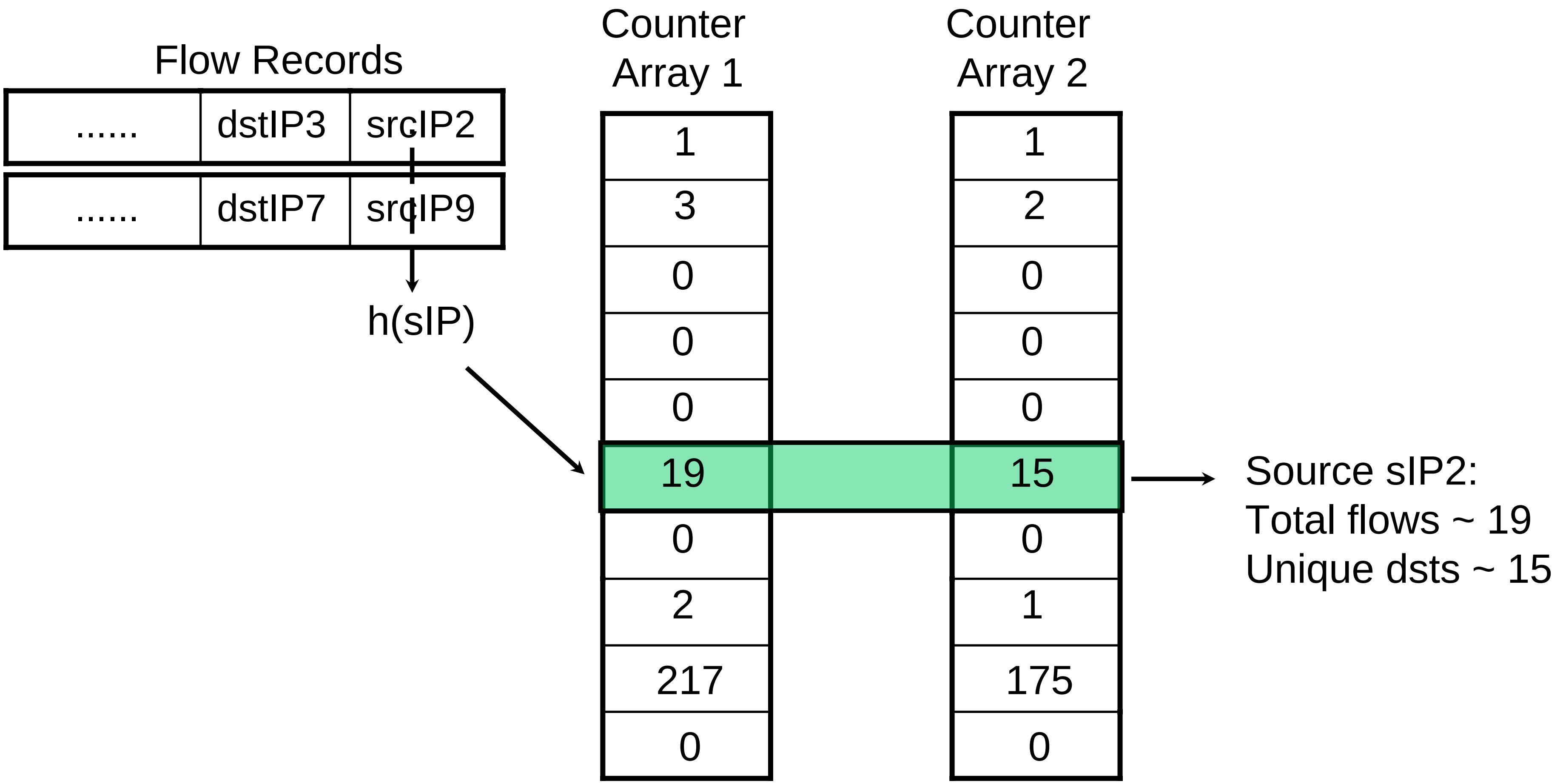
Flow Records

......	dstIP3	srcIP2
......	dstIP7	srcIP9



Counter Array 1	Counter Array 2
1	1
3	2
0	0
0	0
0	0
19	15
0	0
2	1
217	175
0	0

Proposed Solution (Query processing)



Flow Records

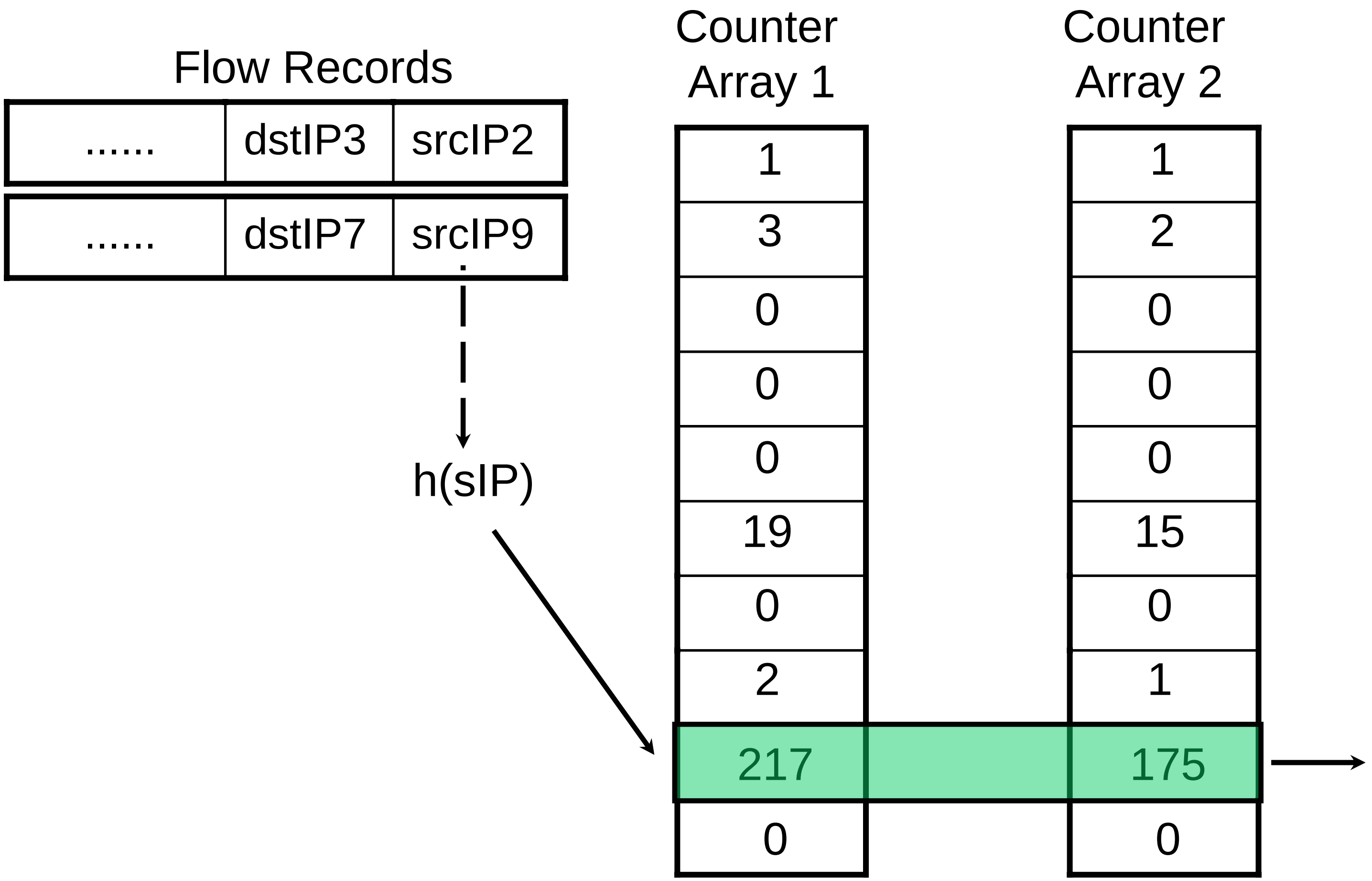
……	dstIP3	srcIP2
……	dstIP7	srcIP9

h(sIP)

Counter Array 1	Counter Array 2
1	1
3	2
0	0
0	0
0	0
19	15
0	0
2	1
217	175
0	0

Source sIP2:
Total flows ~ 19
Unique dsts ~ 15

Proposed Solution (Query processing)



Flow Records

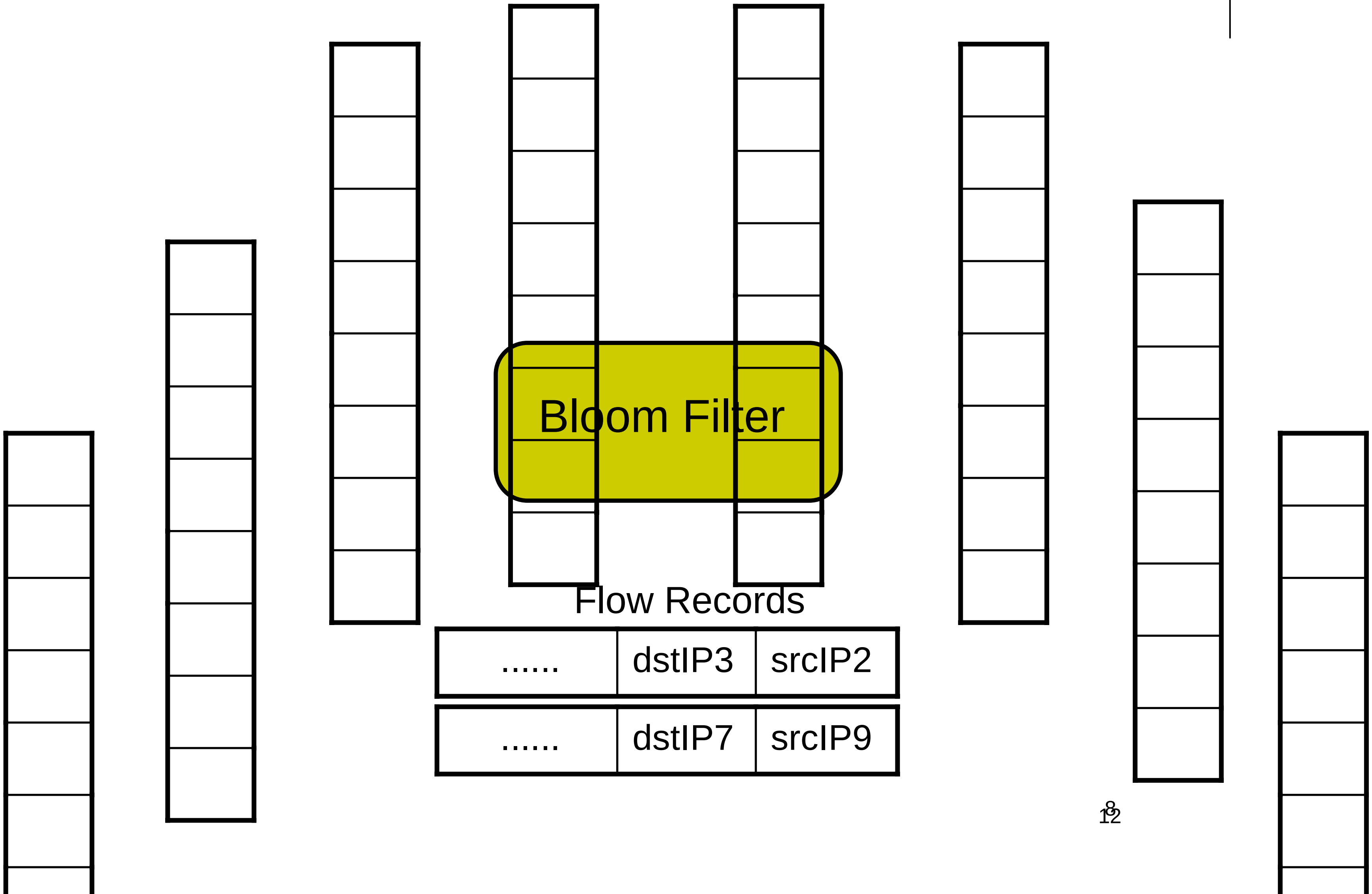
......	dstIP3	srcIP2
......	dstIP7	srcIP9

Counter Array 1	Counter Array 2
1	1
3	2
0	0
0	0
0	0
19	15
0	0
2	1
217	175
0	0

Source sIP2:
Total flows ~ 19
Unique dsts ~ 15

Source sIP9:
Total flows ~ 217
Unique dsts ~ 175

Can we build a more comprehensive system ?



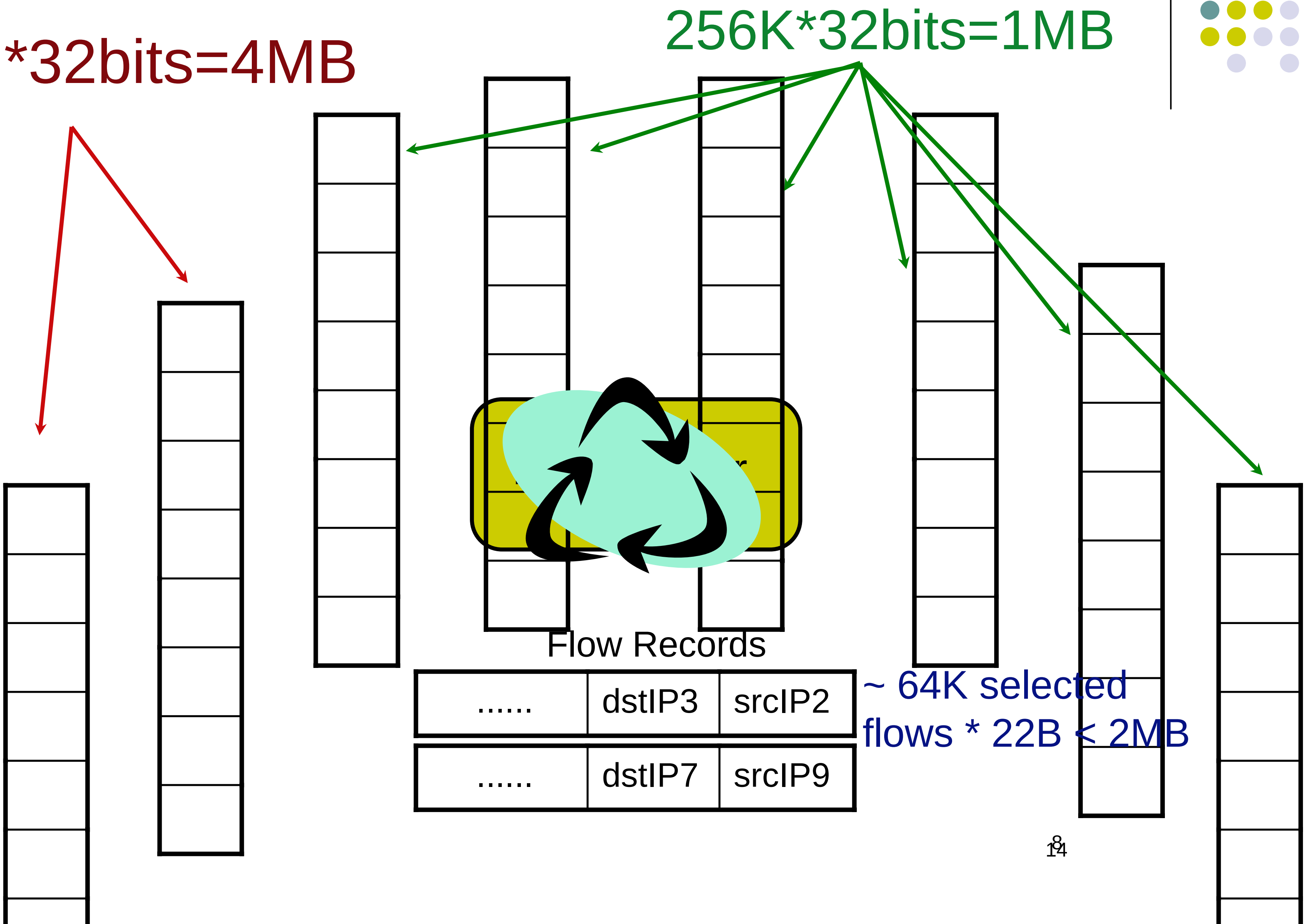
Flow Records

……	dstIP3	srcIP2
……	dstIP7	srcIP9

What will it track ?

Metric	Key Field(s)	Aggregate Queries	Distributional queries	Idenity Queries
Bytes	5-tuple	Total Bytes, Flows	Flows with x<bytes<y	Large Flows
Packets	5-tuple	Total Pkts, Flows	Flows with x Pkts	Large flows by pkts
Total Flows	Source IP	Total sources	Sources sending x flows	Sources sending many flows (> Threshold)
Unique Destinations	Source IP	Total sources	Sources contacting x destinations	Sources contacting many destinations
Total Flows	Dest IP	Total Destinations	Destinations receiving x flows	Destinations receiving many flows
Unique Sources	Dest IP	Total Destinations	Destinations contacted by x sources	Destinations contacted by many sources
Total Flows	<dIP, dPort, proto>	Total 3-tuples	3-tuples receiving x flows	3-tuples receiving many flows
Unique Sources	<dIP, dPort, proto>	Total 3-tuples	3-tuples contacted by x sources	3-tuples contacted by many sources

How much space for 60 M flows?



1M*32bits=4MB

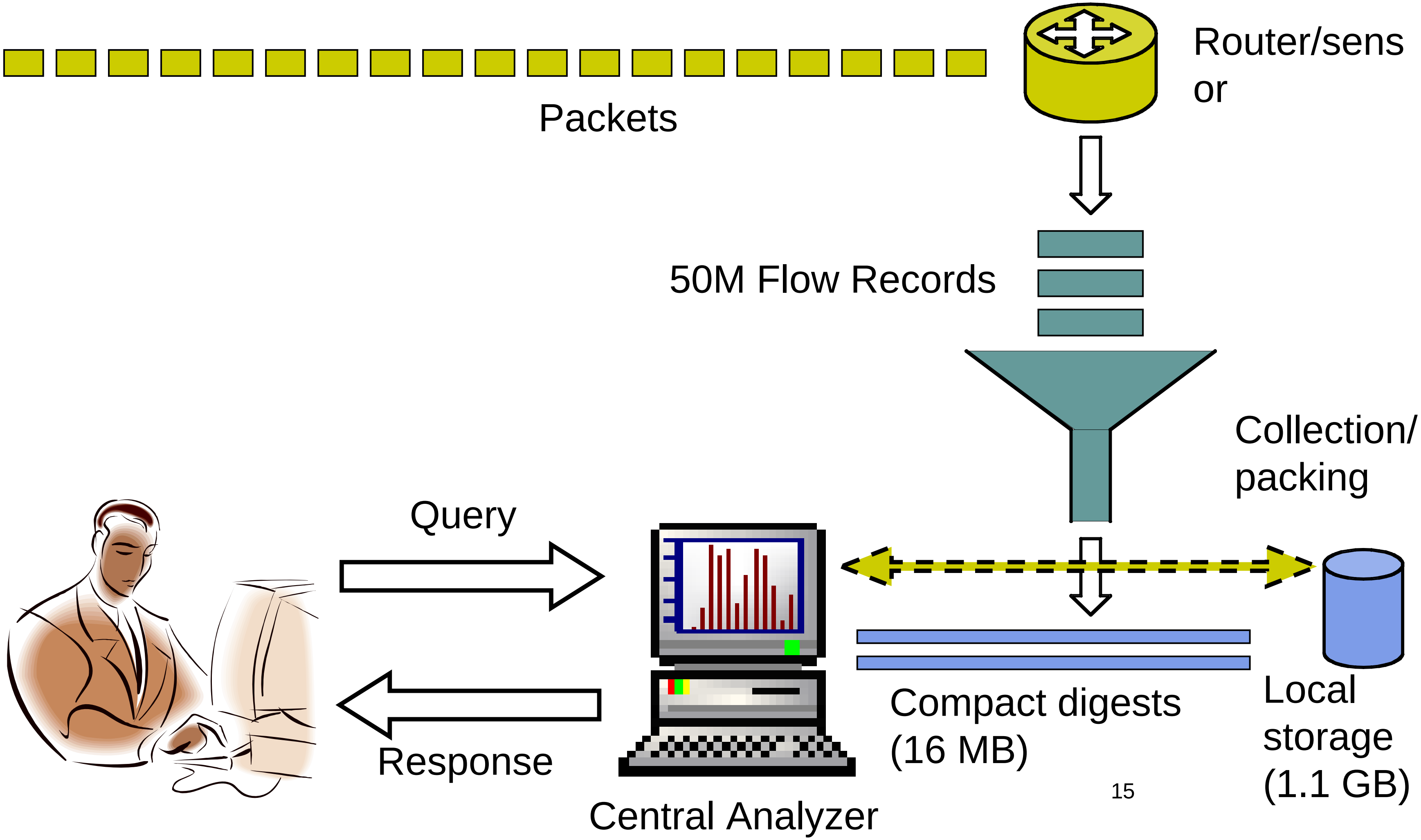
256K*32bits=1MB

Flow Records

......	dstIP3	srcIP2
......	dstIP7	srcIP9

~ 64K selected flows * 22B < 2MB

Flow Collection and Analysis Architecture



Thank you !

- Questions and comments
- Contact: akumar@cc.gatech.edu