Privacy-Preserving Network Flow Recording

By

Bilal Shebaro and Jedidiah Crandall

Presented At

The Digital Forensic Research Conference

DFRWS 2011 USA New Orleans, LA (Aug 1st - 3rd)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

Privacy-Preserving Network Flow Recording

Bilal Shebaro (Computer Science-UNM)
Jedidiah R. Crandall (Computer Science-UNM)

1996...

2011...

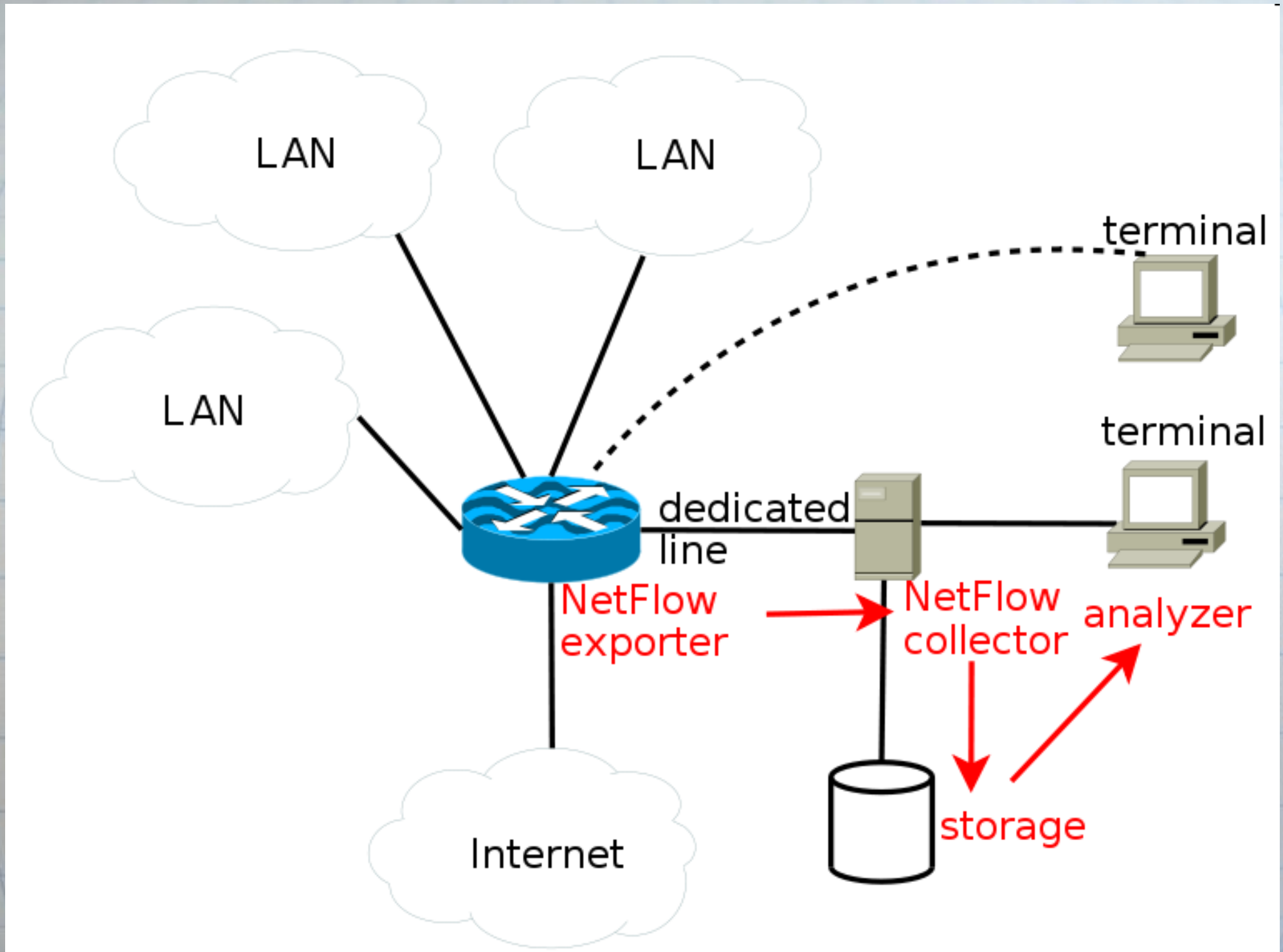

1

Outline

- Background info
- Requirements
- Threat model and challenges
- Prototype implementation
- Experimental setup and results

NetFlow

- Network protocol developed by Cisco
  - Cisco IOS, NXOS such as Juniper routers, Enterasys Switches, Linux, FreeBSD, NetBSD and OpenBSD.


LAN
LAN
LAN
terminal
terminal
dedicated
line
NetFlow
exporter
NetFlow
collector
analyzer
storage
Internet

What's in a NetFlow record?

1. Source IP address
2. Destination IP address
3. Source port for UDP or TCP
4. Destination port for UDP or TCP
5. IP protocol
6. Ingress interface
7. IP Type of Service
8. Bytes transferred

Identity Based Encryption (IBE)

- Proposed by Shamir, 1984
- Realized by Boneh and Franklin, 2001
- Nice properties
  - Hierarchical
  - Can split the key
  - Identity can contain timestamps

Requirements

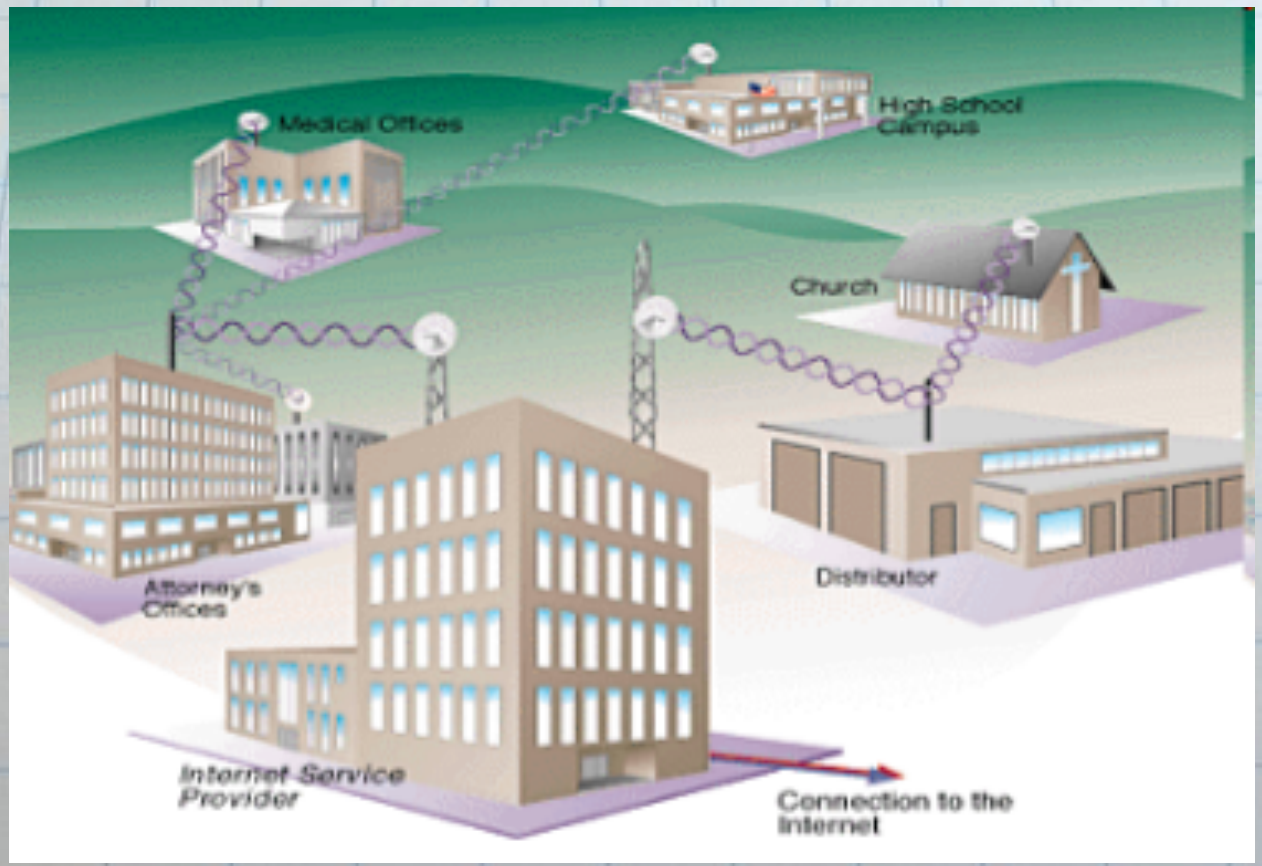
- Uses of NetFlow
  - Network debugging
  - Law enforcement
  - Statistics
    - Billing
    - Network planning
- Privacy
  - Separation of duty
  - Aggregate vs. individual
- 1-10 GBit rates for /24, /22, /20
- Transparent to users

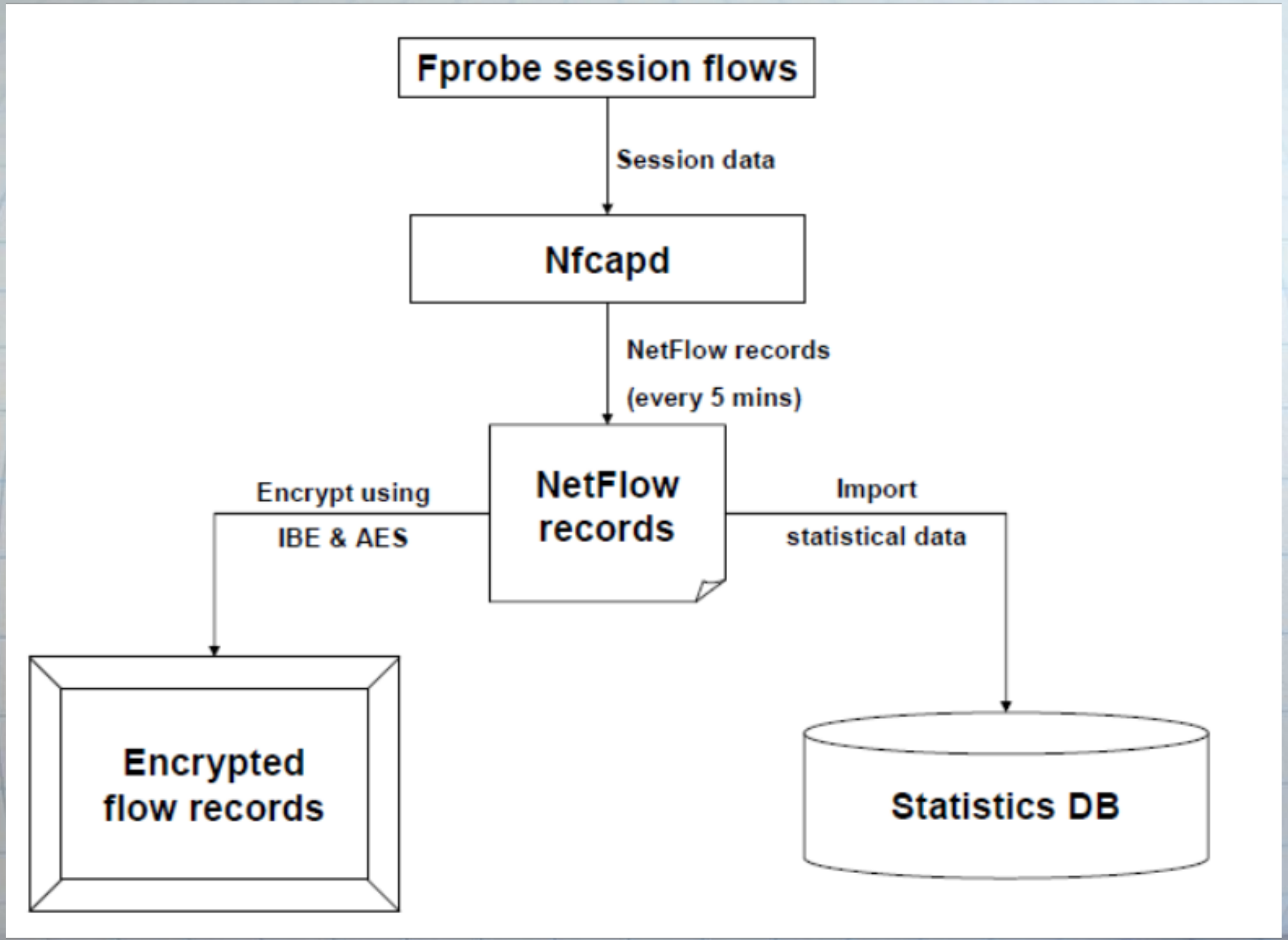
Threat model & challenges

- Confidentiality
  - Employees that don't understand the policy
  - Not following proper legal processes
  - Rogue employees (backward security)
  - Compromised systems (backward security)
- Must have basic statistics readily available
- Challenges
  - PKI
  - Scalability
  - Statistics

Scenario

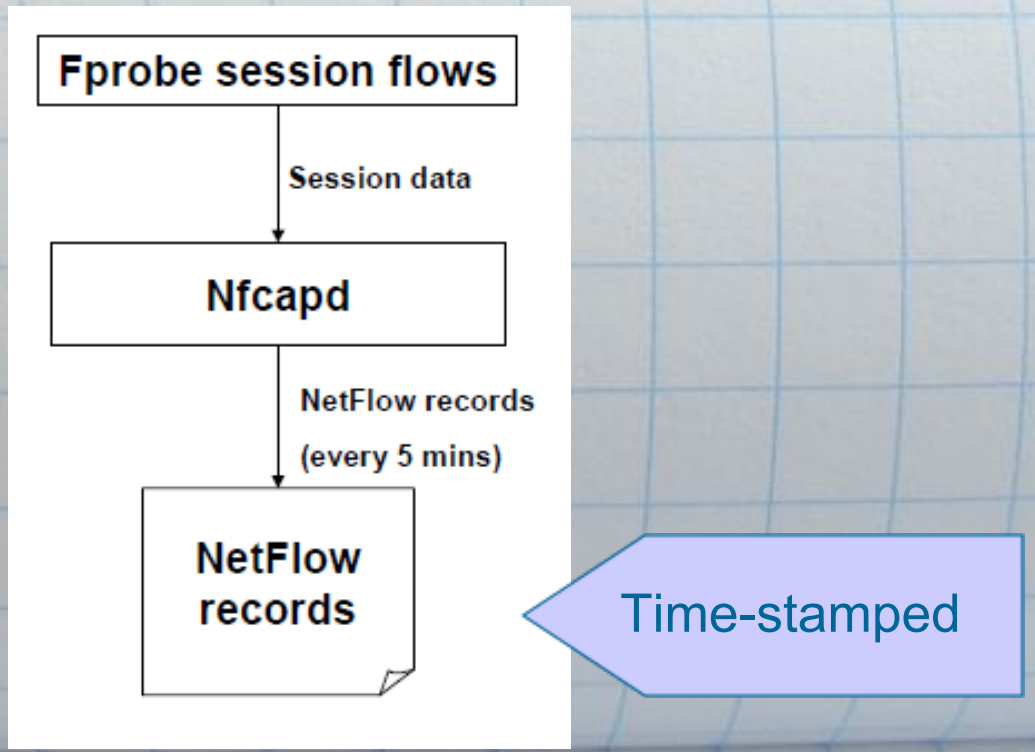
Scenario


Medical Offices
High School Campus
Church
Attorney's Offices
Distributor
Internet Service Provider
Connection to the Internet


Fprobe session flows
Session data
Nfcapd
NetFlow records
(every 5 mins)
NetFlow records
Encrypt using
IBE & AES
Import
statistical data
Encrypted flow records
Statistics DB

Step 0: Data Collection

- Fprobe 1.1 running
- Nfcapd collects the flow and does file rotation every 5 minutes (configured)

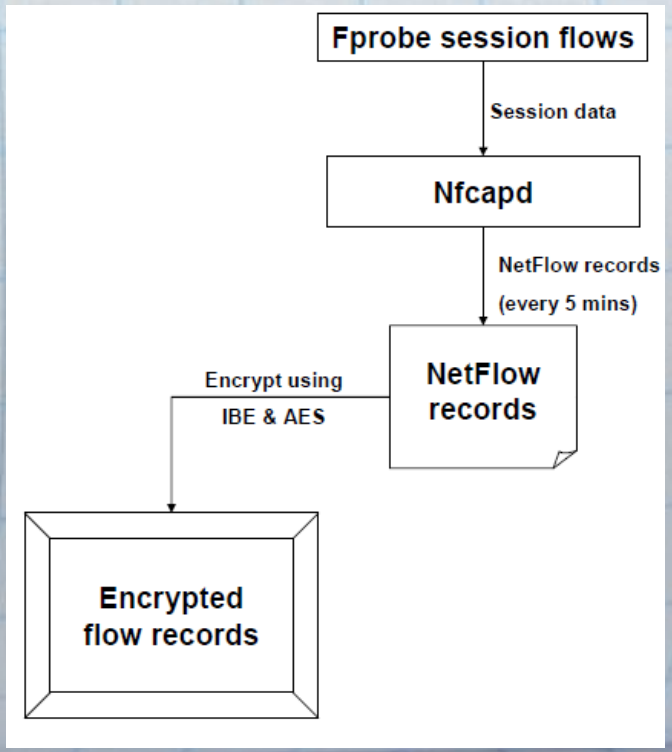


Step 1: Flow Encryption

- Flows are combined per IP
- AES (128 key size) encrypts the flow
- Identity Based Encryption (IBE) encrypts AES key using:
  - Corresponding IP address
  - Corresponding file timestamp

IP, IBE(AES-key), AES(flow record)

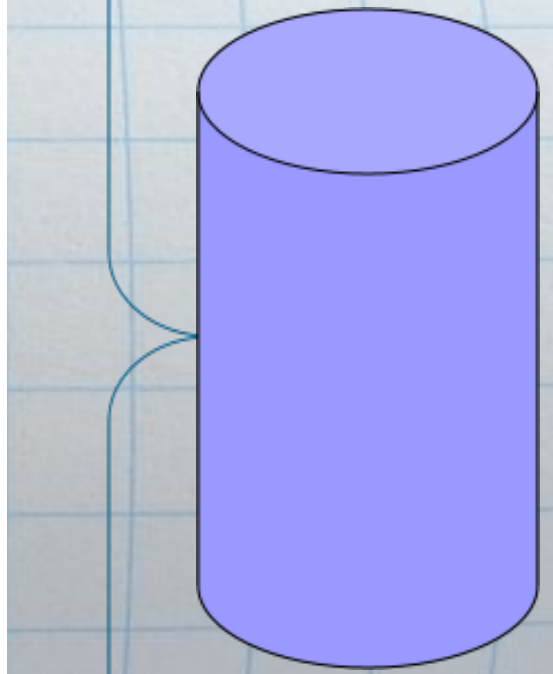
.
.
.
.



Step 2: Statistical Reports

- Records are filtered out into:

- IP Address
- TP: Time Period (time-stamped)
- TTI: Total TCP bytes In
- TTO: Total TCP bytes Out
- TUI: Total UDP bytes In
- TUO: Total UDP bytes Out
- LPI: List of Ports In
- LPO: List of Ports Out
- BI: Bytes In
- BO: Bytes Out
- PI: Packets In
- PO: Packets Out



Query examples

(Link utilization)

$$Q1 : Sum[BI, (TP \geq \alpha) \bullet IP] \ \ \& \ result \geq \beta$$

$$Q2 : Sum[BO, (TP \geq \alpha) \bullet IP] \ \ \& \ result \geq \beta$$

$$Q3 : Sum[BI + BO, (TP \geq \alpha) \bullet IP] \ \ \& \ result \geq \beta$$

Query examples

(Applications being used)

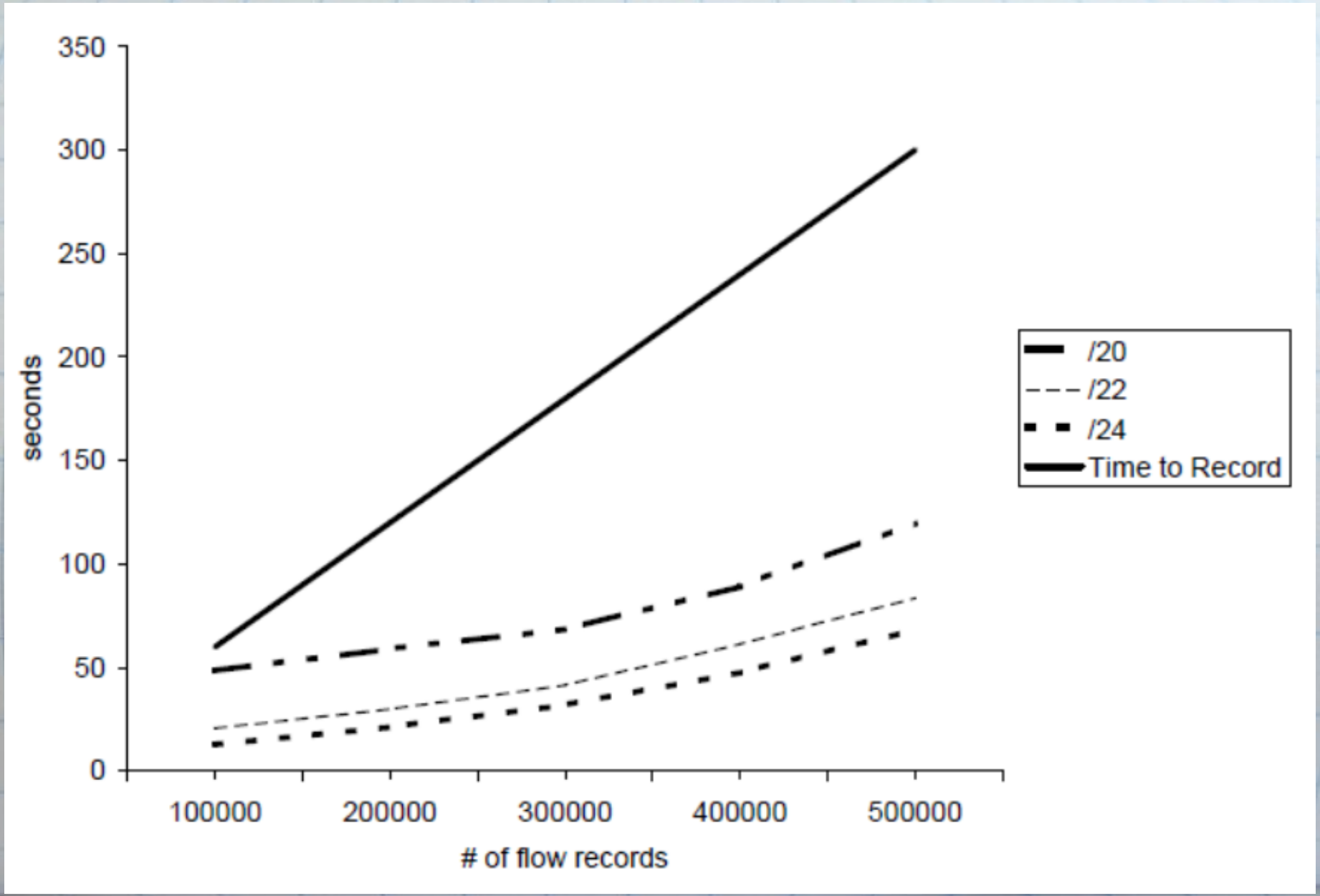
$$Q5 : list[LPI, (TP \geq \alpha) \bullet IP_i]$$
$$+ \quad list[LPO, (TP \geq \alpha) \bullet IP_i]$$

$$\forall IP_i \in subnet,\ count(IP_{i_s}) > \delta$$

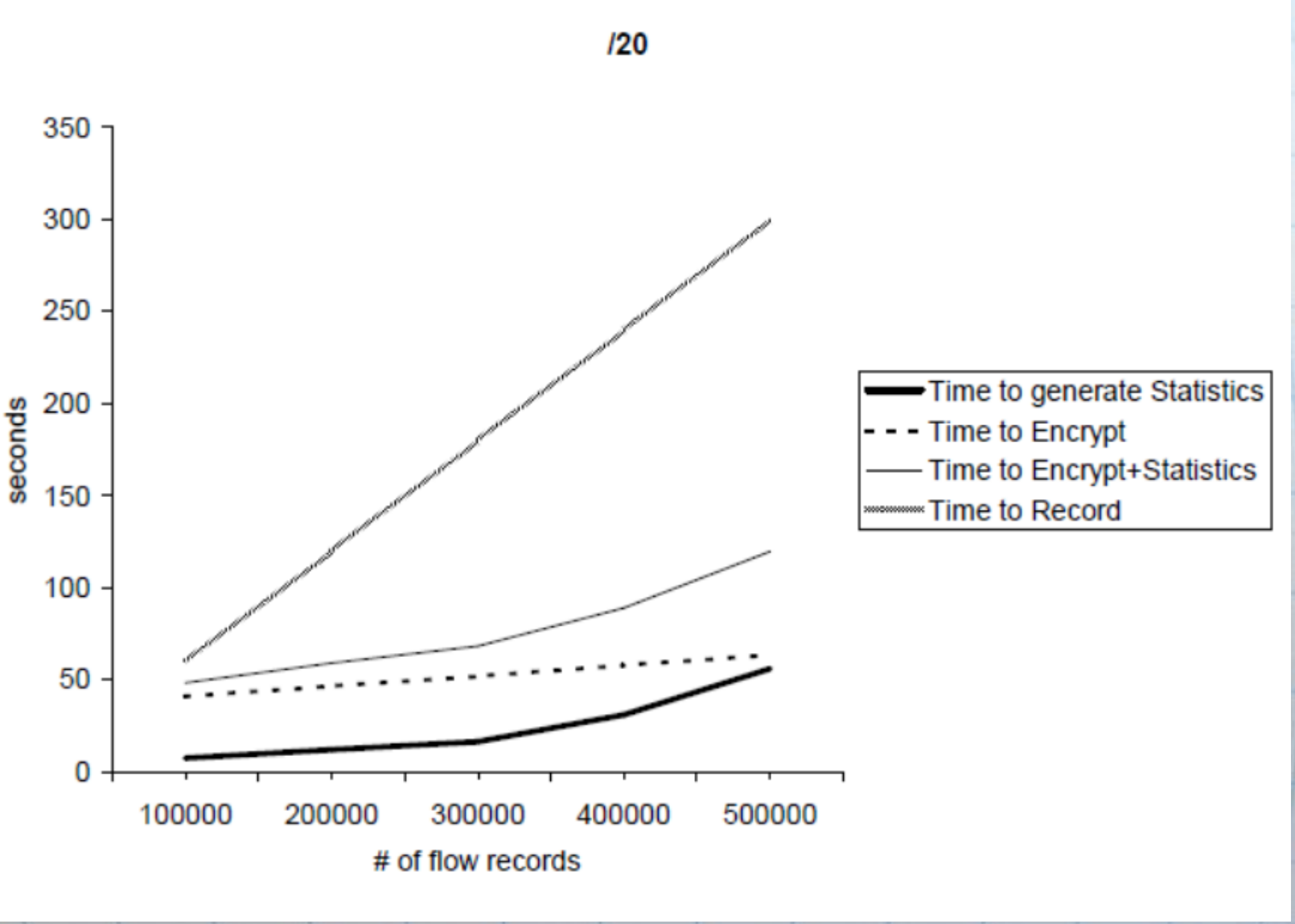
Experimental setup

- /20, /22, /24 traffic data was generated
- Core i7 X980 running at 3.33 GHz, 24 GB RAM, RAID 0 array at 3 Gbps
- Live capture experiments for 6 hours for each subnet size
  - TCP-replay
- Measured times for data recording *vs.* encryption and statistics

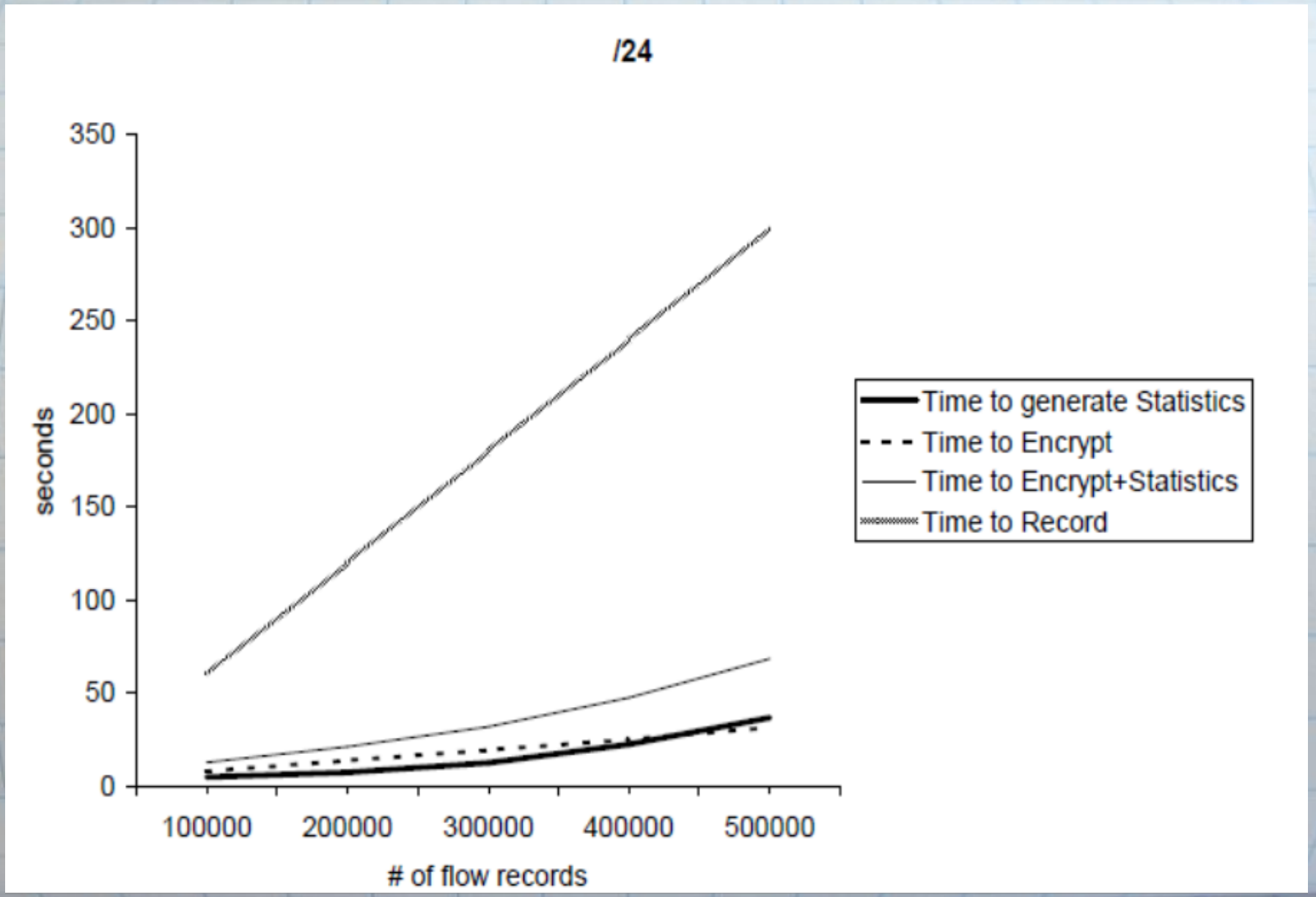
IBE scalability



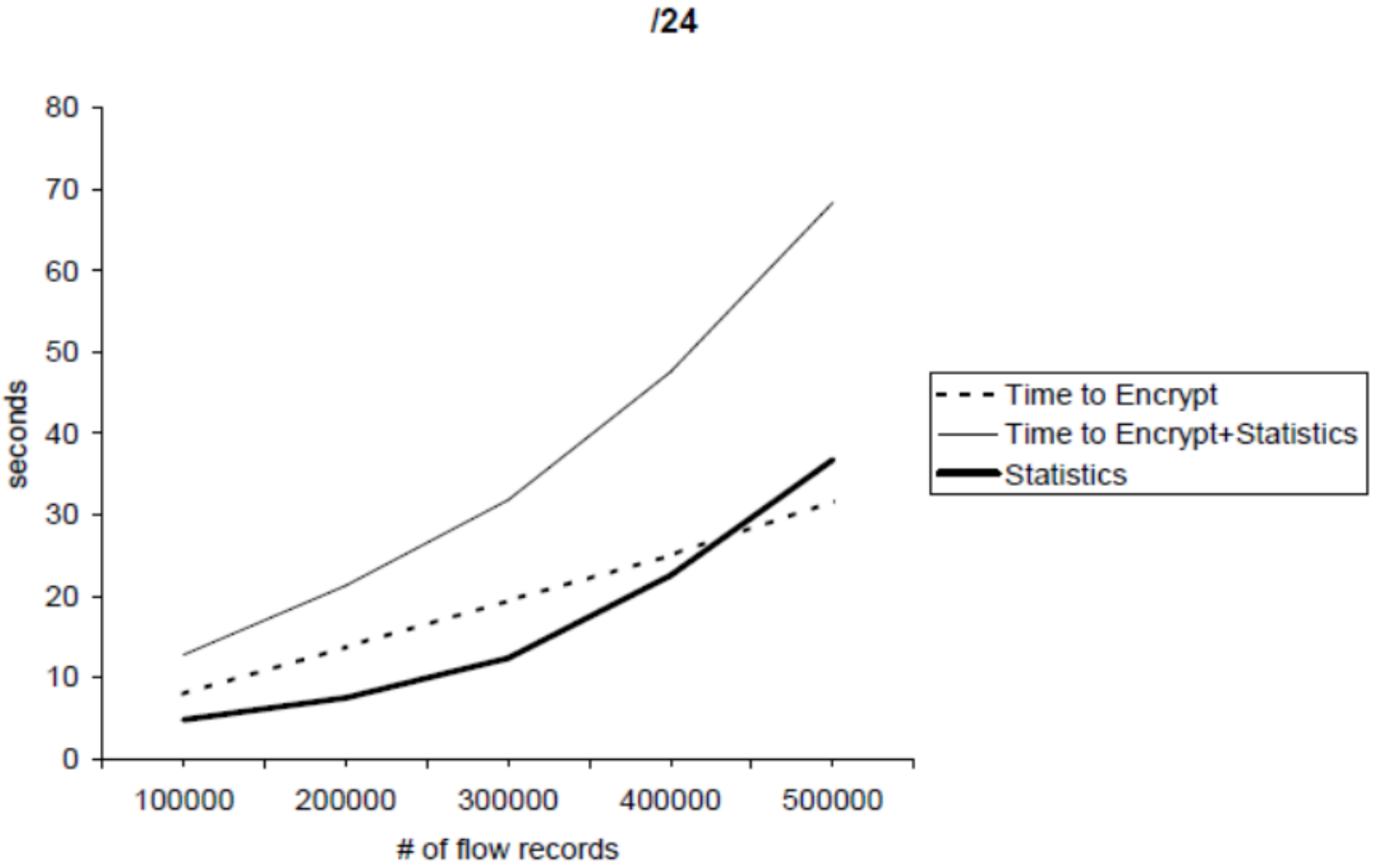
/20


/20
350
300
250
200
150
100
50
0
seconds
100000
200000
300000
400000
500000
# of flow records
Time to generate Statistics
Time to Encrypt
Time to Encrypt+Statistics
Time to Record

/24


/24
350
300
250
200
150
100
50
0
seconds
100000
200000
300000
400000
500000
# of flow records
Time to generate Statistics
Time to Encrypt
Time to Encrypt+Statistics
Time to Record

/24 (zoomed in)


/24
80
70
60
50
40
30
20
10
0
seconds
100000
200000
300000
400000
500000
# of flow records
Time to Encrypt
Time to Encrypt+Statistics
Statistics

Offline Experiments

Subnet size	Maximum rate (Gbps)
/24	23
/22	18
/20	12

Discussion

- Network size scalability acceptable
  - IBE scales as expected
  - Statistics implementation details
- Traffic rate scales to above 10 Gbps

Future work

- Implementation
  - Crypto and semantic security
  - Inference
- Differential privacy
- More statistical queries

Takeaway message...

Collection of network data doesn't need to be in conflict with privacy concerns.

Specifically, tools to help enforce policies on the data can help.

Acknowledgments

- NSF #0905177 & #0844880
- Monzy Merza
- DFWRS reviewers

"This material is based upon work supported by the National Science Foundation under Grant Nos. 0905177 and 0844880. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."