清華大學
Tsinghua University

Network Forensics

Presenter

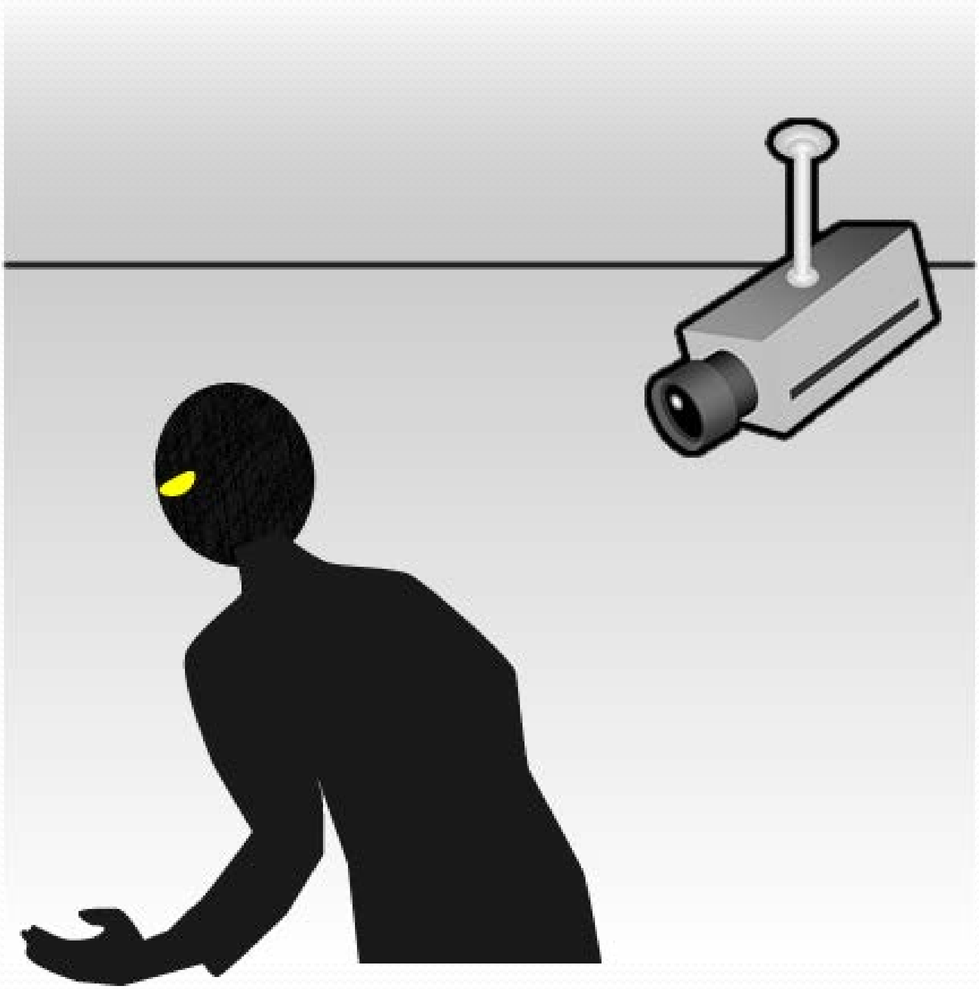
Ali（阿里）

30 November 2016

Introduction

Introduction – Contd.

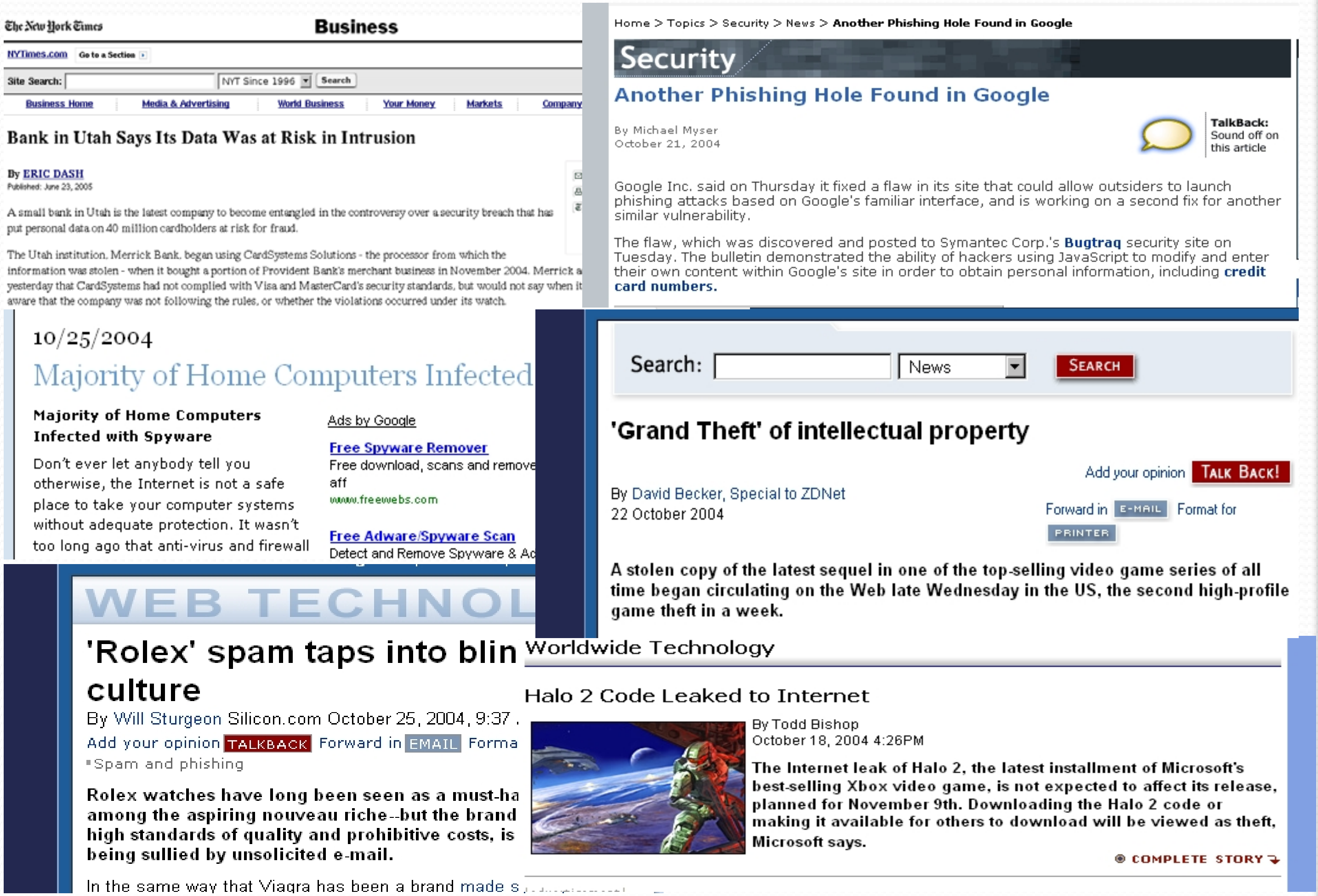
Introduction – Contd.



In Computer Networks

Problem

- ❑ IDS/IPS
- ❑ UTM
- ❑ DPI
- ❑ Firewall
- ❑ …



The New York Times **Business**

NYTimes.com Go to a Section

Site Search: NYT Since 1996 Search

Business Home | Media & Advertising | World Business | Your Money | Markets | Company

Bank in Utah Says Its Data Was at Risk in Intrusion

By ERIC DASH
Published: June 23, 2005

A small bank in Utah is the latest company to become entangled in the controversy over a security breach that has put personal data on 40 million cardholders at risk for fraud.

The Utah institution, Merrick Bank, began using CardSystems Solutions - the processor from which the information was stolen - when it bought a portion of Provident Bank's merchant business in November 2004. Merrick a yesterday that CardSystems had not complied with Visa and MasterCard's security standards, but would not say when it aware that the company was not following the rules, or whether the violations occurred under its watch.

Home > Topics > Security > News > **Another Phishing Hole Found in Google**

Security

Another Phishing Hole Found in Google

By Michael Myser
October 21, 2004

TalkBack: Sound off on this article

Google Inc. said on Thursday it fixed a flaw in its site that could allow outsiders to launch phishing attacks based on Google's familiar interface, and is working on a second fix for another similar vulnerability.

The flaw, which was discovered and posted to Symantec Corp.'s **Bugtraq** security site on Tuesday. The bulletin demonstrated the ability of hackers using JavaScript to modify and enter their own content within Google's site in order to obtain personal information, including **credit card numbers.**

10/25/2004

Majority of Home Computers Infected

Majority of Home Computers Infected with Spyware

Don't ever let anybody tell you otherwise, the Internet is not a safe place to take your computer systems without adequate protection. It wasn't too long ago that anti-virus and firewall

Ads by Google

Free Spyware Remover
Free download, scans and remove aff
www.freewebs.com

Free Adware/Spyware Scan
Detect and Remove Spyware & Ac

Search: News SEARCH

'Grand Theft' of intellectual property

Add your opinion TALK BACK!

By David Becker, Special to ZDNet
22 October 2004

Forward in E-MAIL Format for PRINTER

A stolen copy of the latest sequel in one of the top-selling video game series of all time began circulating on the Web late Wednesday in the US, the second high-profile game theft in a week.

WEB TECHNOL

'Rolex' spam taps into blin culture

By Will Sturgeon Silicon.com October 25, 2004, 9:37
Add your opinion TALKBACK Forward in EMAIL Forma
▪ Spam and phishing

Rolex watches have long been seen as a must-ha among the aspiring nouveau riche--but the brand high standards of quality and prohibitive costs, is being sullied by unsolicited e-mail.

In the same way that Viagra has been a brand made s

Worldwide Technology

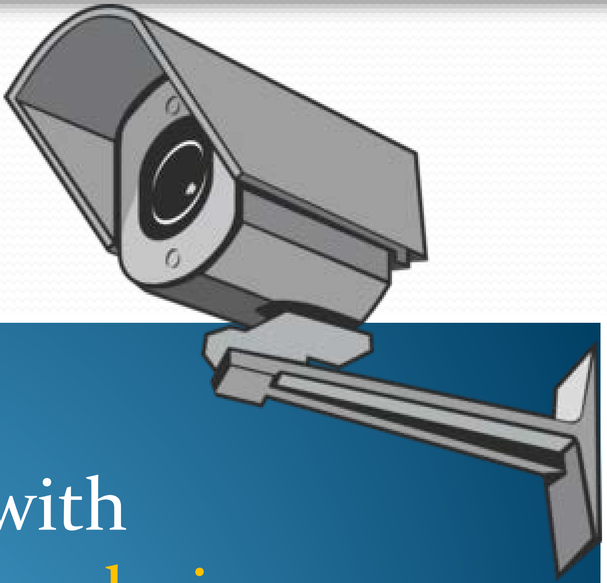
Halo 2 Code Leaked to Internet

By Todd Bishop
October 18, 2004 4:26PM

The Internet leak of Halo 2, the latest installment of Microsoft's best-selling Xbox video game, is not expected to affect its release, planned for November 9th. Downloading the Halo 2 code or making it available for others to download will be viewed as theft, Microsoft says.

◉ COMPLETE STORY

Network Forensics



A science that deals with
capture, recording, and analysis
of network traffic
for detecting and investigating intrusions.

Outline

- Introduction to Network Forensics
- Payload Attribution Systems
  - Bloom Filters
  - BBF & HBF
  - Others
  - Payload Attribution via Character Dependent Multi-Bloom Filters

Introduction to NF

Goal

- ❑ Who
- ❑ How
- ❑ When
- ❑ What reason

Challenges

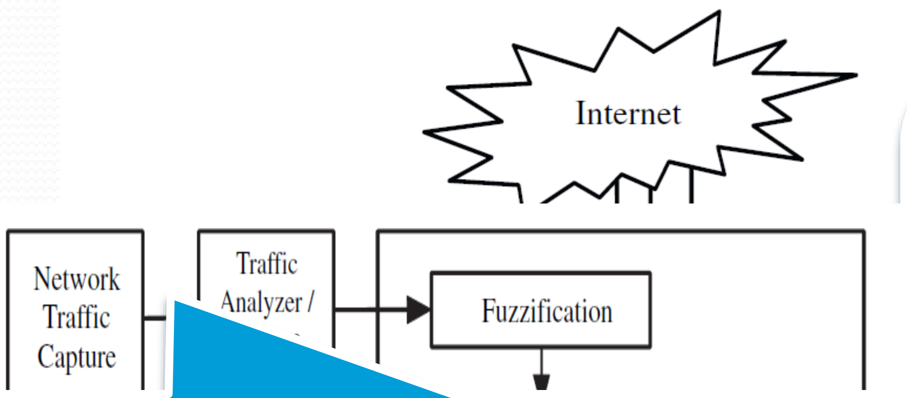
- Large amount of data
  - Storage
  - Search
- Different types of network protocol

Classification

- ❑ Purpose
  - ✓ General Network Forensics
  - ✓ Strict Network Forensics
- ❑ Collection of traffic
  - ✓ Catch it as you can
  - ✓ Stop look and listen
- ❑ Nature
  - ✓ Hardware and pre-installed software
  - ✓ Software tool

Different Works

- Distributed systems based frameworks



eworks

Shanmugasundaram et al. (2003)

Wang and Daniels (2008)

Yas 02)
Th 8)

Rekhis et al. (2008)

8)

Honeypot frameworks are used to attract the attackers

To observe the methodology of the att

To improve defense mechanisms

Almulhem and Traore (2005)
Nikkel (2006)
Vandenberghe (2008)

Payload Attribution Systems

Packet Digests

- Compute and store Synopsis

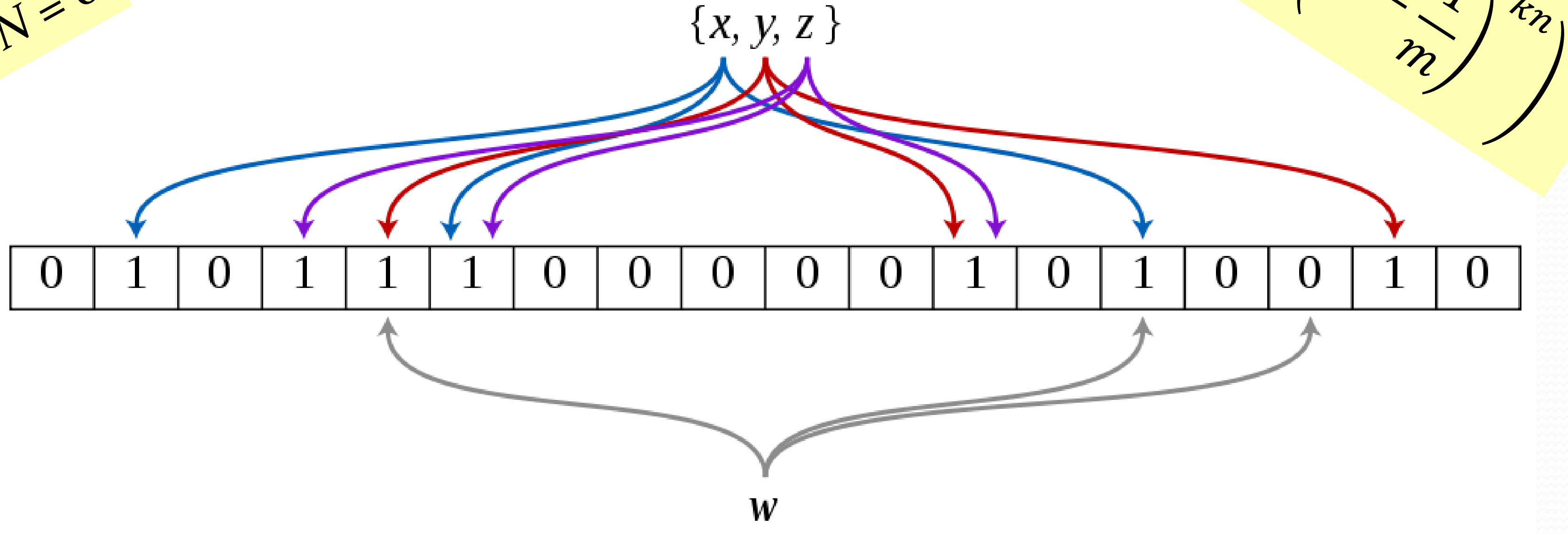
Representing a set of elements succinctly with predefined loss in information and has the ability to recall the original set of elements with a preset accuracy

Bloom Filter

- Randomized data structure

$FN = 0$

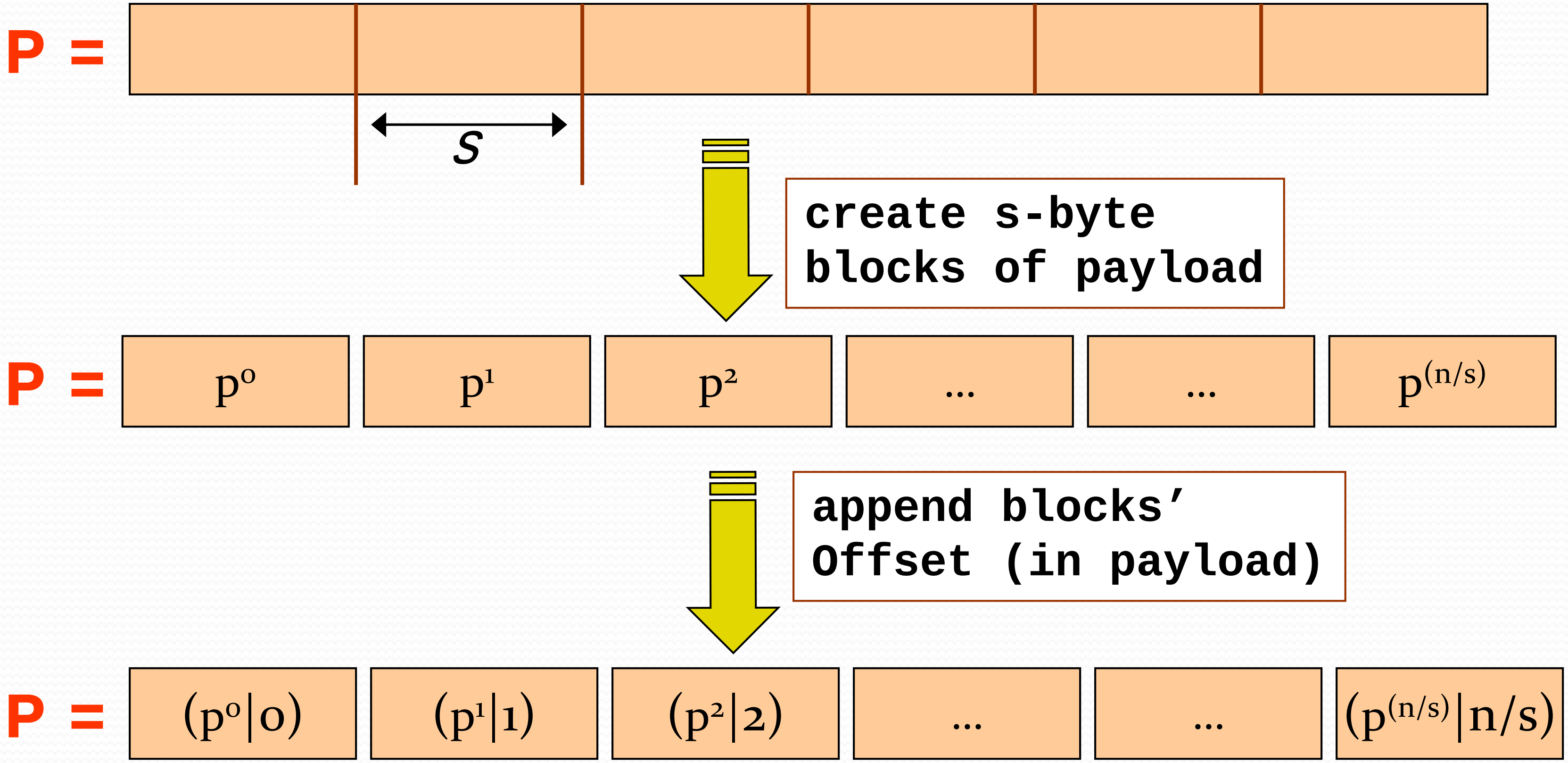
$$FP = \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^{k}$$



Bloom Filter – Contd.

- $FP = \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^{k}$
- $k = \frac{m}{n} \ln 2$
- $\ln(FP) = -\frac{m}{n} (\ln 2)^2$
- $m = -\frac{n \ln p}{(\ln 2)^2}$

Block-based Bloom Filter

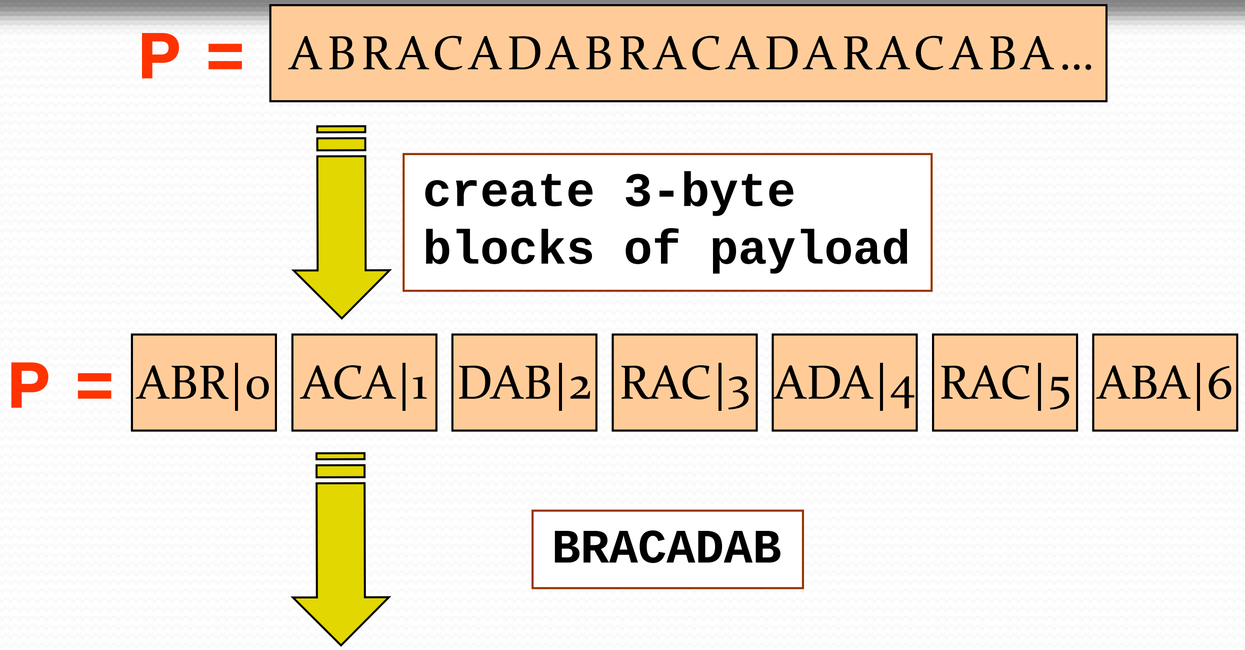


create s-byte blocks of payload

append blocks’ Offset (in payload)

Insert each block into a Bloom Filter

Block-based Bloom Filter – contd.



P = ABRACADABRACADARACABA…

create 3-byte blocks of payload

P = ABR|0 ACA|1 DAB|2 RAC|3 ADA|4 RAC|5 ABA|6

BRACADAB

	Offset							
Alignment	0	1	2	3	4	5	6	…
BRA CAD AB	X	X	X	X	X	X		
B RAC ADA B		X	X	YY	X	YX		
BR ACA CAB		YY	X	X	X	X		

Block-based Bloom Filter – contd.

P1 =

A	B	R	A	C	A

P2 =

C	D	A	B	R	A

BBF =

(A|0)	(B|1)	(R|2)	(A|3)	(C|4)	(A|5)
(C|0)	(D|1)	(A|2)	(B|3)	(R|4)	(A|5)

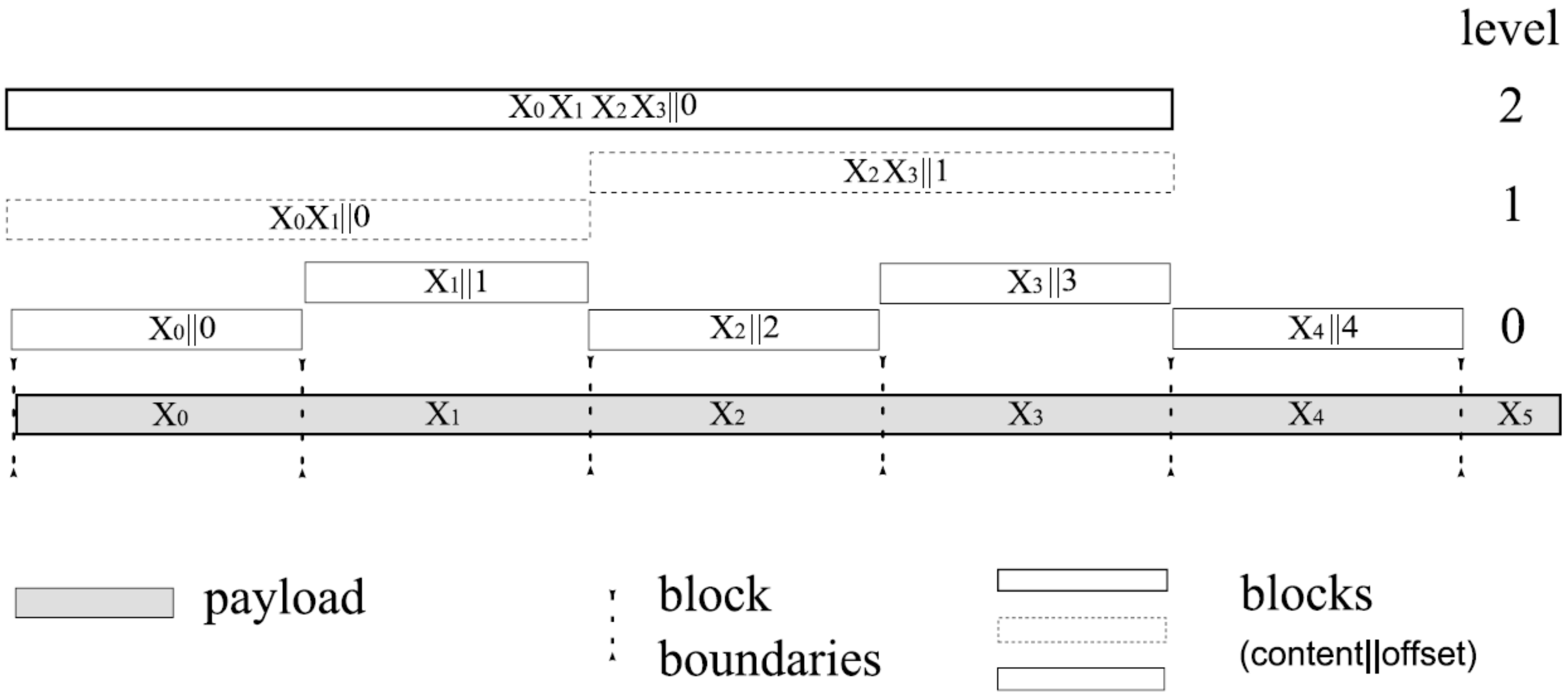
"Offset Collisions"

(A|0)	(B|1)	(R|2)	(A|3)	(C|4)	(A|5)
(C|0)	(D|1)	(A|2)	(B|3)	(R|4)	(A|5)

For query strings: "AD", "CB", "DR", "AA" etc. BBF falsely identifies them as seen in the payload!

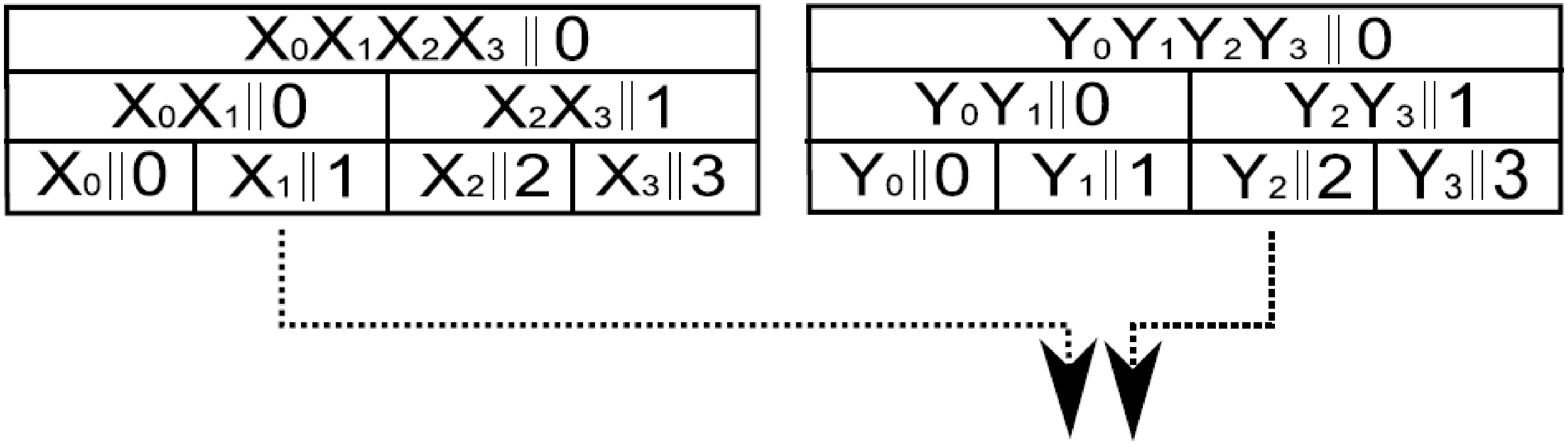
Because BBF cannot distinguish between P1 and P2

Hierarchical Bloom Filter



Hierarchical Bloom Filter – Contd.

HBF hierarchies for two payloads



$X_0X_1X_2X_3 \| 0$

$X_0X_1\|0$ $X_2X_3\|1$

$X_0\|0$ $X_1\|1$ $X_2\|2$ $X_3\|3$

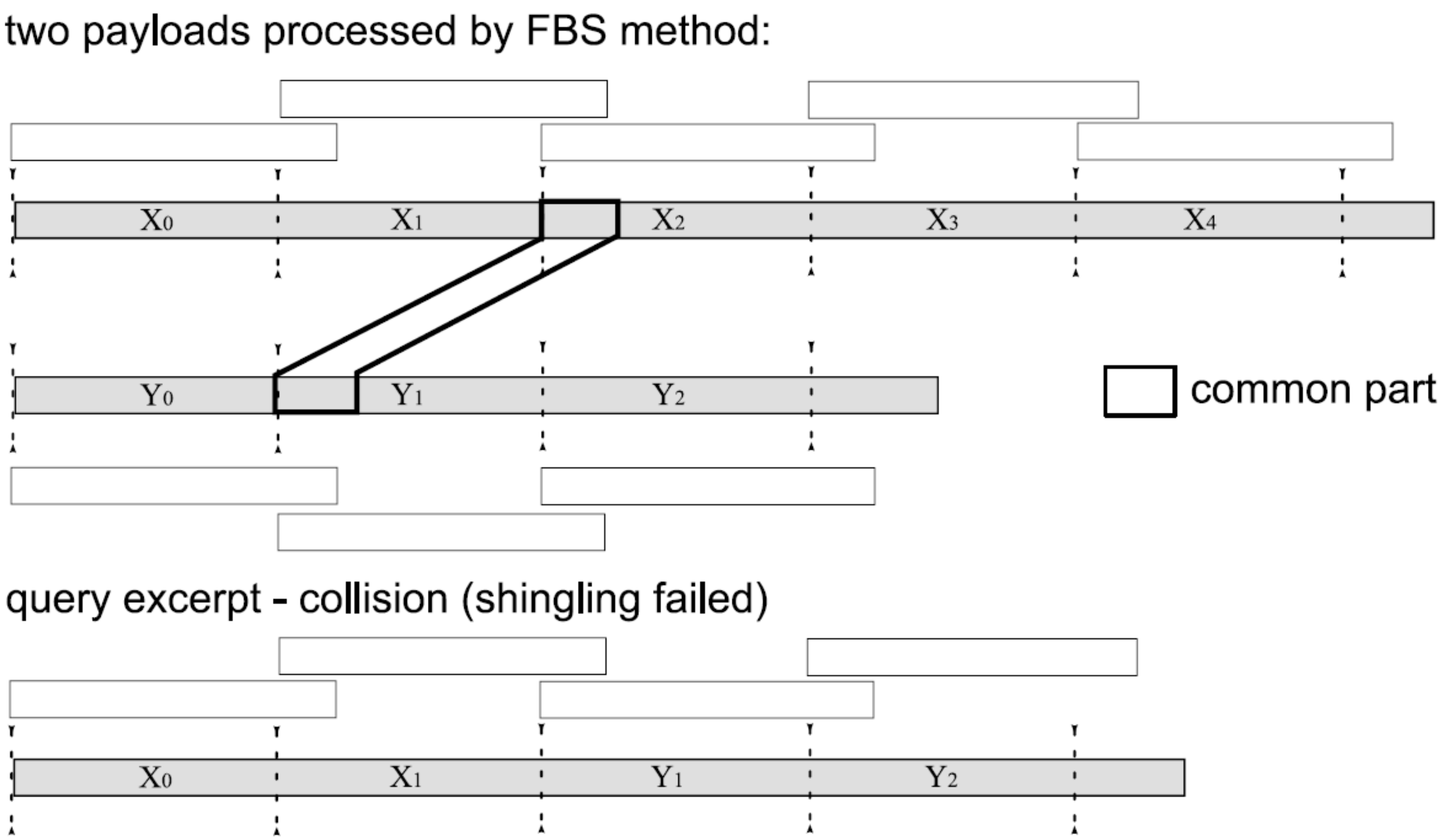
$Y_0Y_1Y_2Y_3 \| 0$

$Y_0Y_1\|0$ $Y_2Y_3\|1$

$Y_0\|0$ $Y_1\|1$ $Y_2\|2$ $Y_3\|3$

false match for an excerpt X_1Y_2

Others – FBS



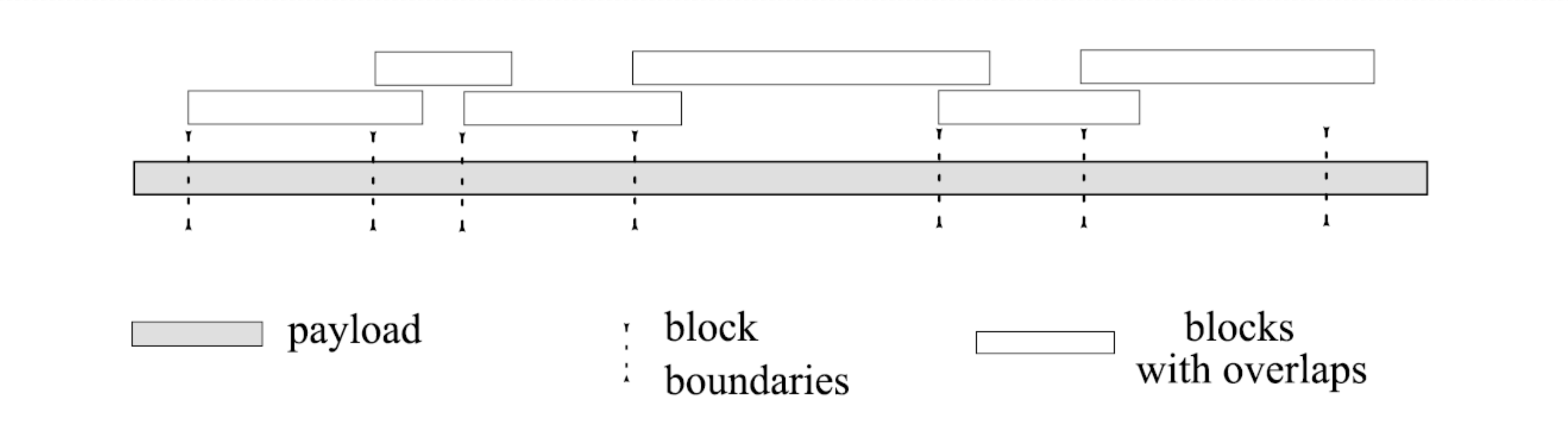
two payloads processed by FBS method:

query excerpt - collision (shingling failed)

Others – VBS

- slide a window of size *w* bytes through the whole payload

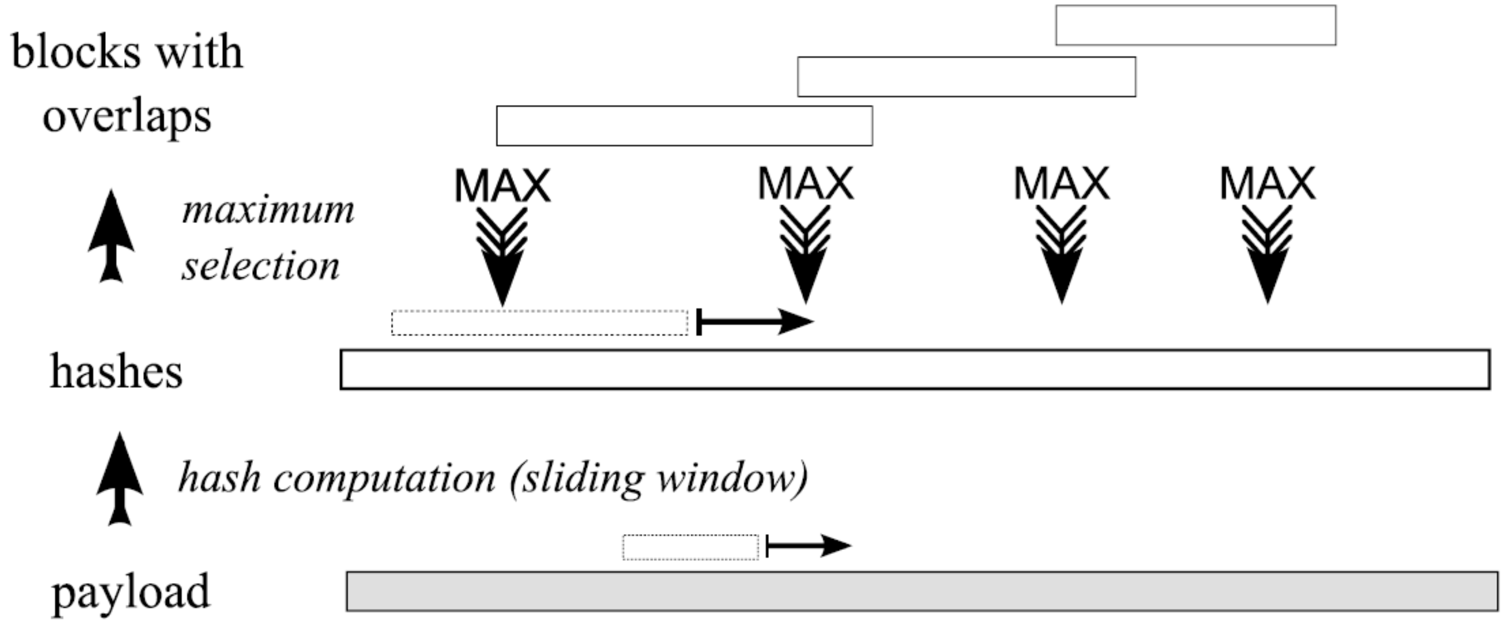
- Compute fingerprint
  - $F(c_1,\ldots, c_w) = (c_1 p^{w-1} + c_2 p^{w-2} + \ldots + c_w) \bmod M$
  - $F(c_2,\ldots, c_{w+1}) = (pF(c_1,\ldots, c_w) + c_{w+1} - c_1 p^w) \bmod M$



Others – WBS

- Compute the fingerprint of payload like VBS method
- Get an array of hashes
  - i-th element is the hash of bytes $c_i,\ldots,c_{i+w-1}$
  - c_i is the i-th byte of the payload
- Slide a winnowing window of size ww through the array
- put a boundary immediately before the position of the maximum hash value for each position of the winnowing window

Others – WBS (Contd.)



Others – WMH

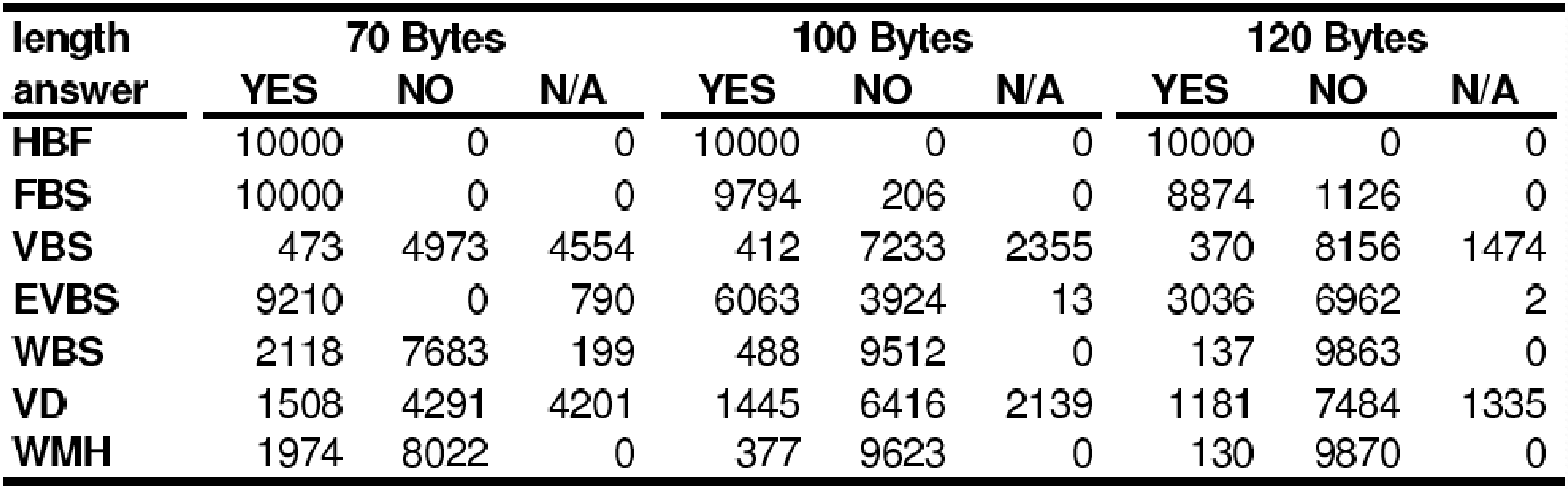
- use multiple instances of WBS

- Differ in hash functions

- reduce the probability of false positives

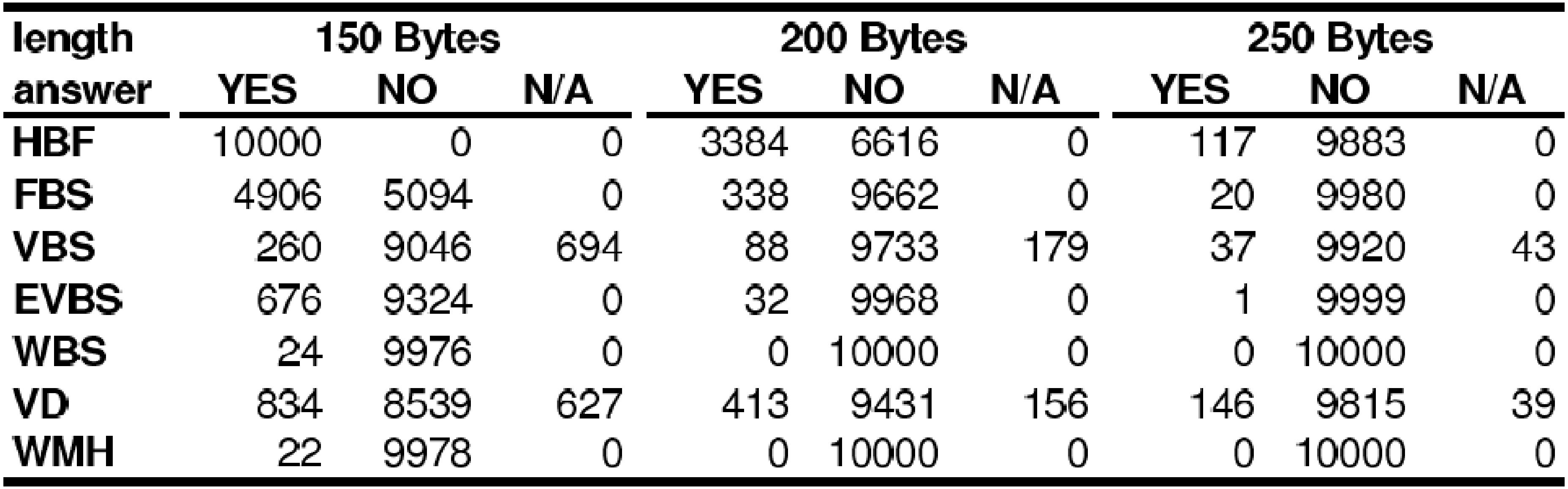
- Final answer for *t* instances is positive only if all *t* answers are positive.

Results



length	70 Bytes			100 Bytes			120 Bytes		
answer	YES	NO	N/A	YES	NO	N/A	YES	NO	N/A
HBF	10000	0	0	10000	0	0	10000	0	0
FBS	10000	0	0	9794	206	0	8874	1126	0
VBS	473	4973	4554	412	7233	2355	370	8156	1474
EVBS	9210	0	790	6063	3924	13	3036	6962	2
WBS	2118	7683	199	488	9512	0	137	9863	0
VD	1508	4291	4201	1445	6416	2139	1181	7484	1335
WMH	1974	8022	0	377	9623	0	130	9870	0

Results – Contd.



length	150 Bytes			200 Bytes			250 Bytes		
answer	YES	NO	N/A	YES	NO	N/A	YES	NO	N/A
HBF	10000	0	0	3384	6616	0	117	9883	0
FBS	4906	5094	0	338	9662	0	20	9980	0
VBS	260	9046	694	88	9733	179	37	9920	43
EVBS	676	9324	0	32	9968	0	1	9999	0
WBS	24	9976	0	0	10000	0	0	10000	0
VD	834	8539	627	413	9431	156	146	9815	39
WMH	22	9978	0	0	10000	0	0	10000	0

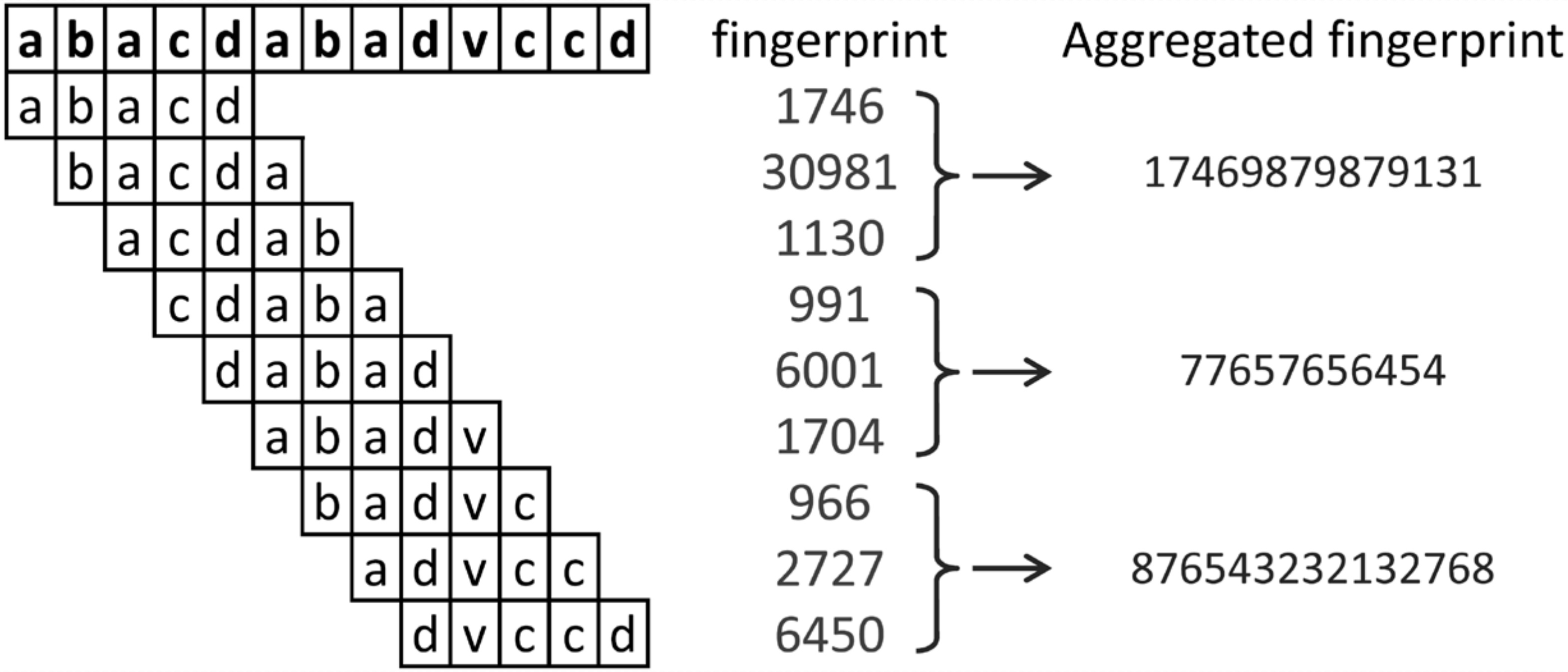
Payload Attribution via Character Dependent Multi-Bloom Filter

Goals & Structure

- Support RegEx Queries
  - $ABC\ldots\ldots DEF$
- Better Data Reduction Ratio
- Use 256 Bloom Filters

CMBF

$$fingerprint(c_i, c_{i+1}, \ldots, c_{i+w-1}) =$$
$$(c_i \bmod q) \times p_{w-1} + (c_{i+1} \bmod q) \times p_{w-2} + \cdots + (c_{i+w-1} \bmod q) \times p_0$$



a b a c d a b a d v c c d

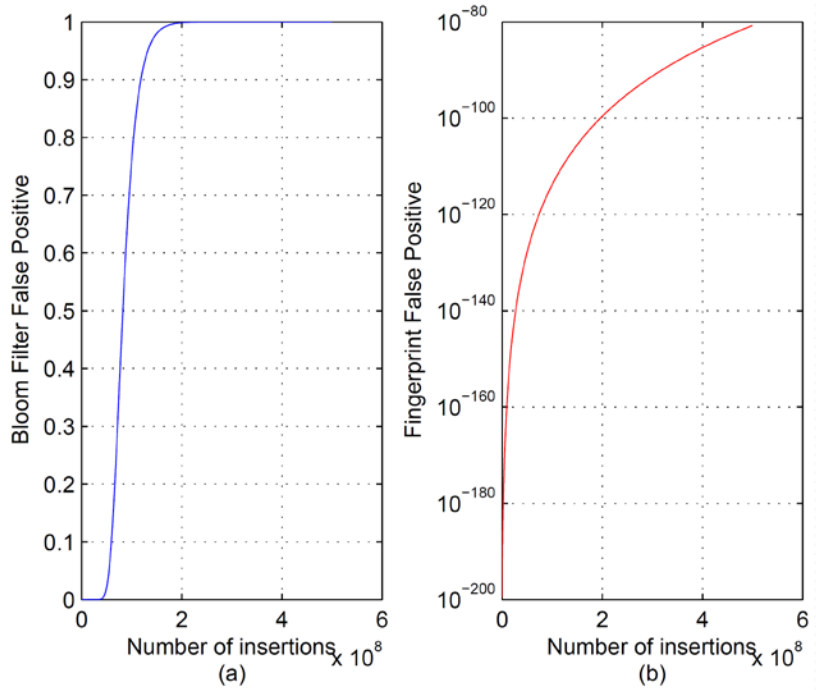
	fingerprint	Aggregated fingerprint
a b a c d	1746	17469879879131
b a c d a	30981	
a c d a b	1130	
c d a b a	991	77657656454
d a b a d	6001	
a b a d v	1704	
b a d v c	966	876543232132768
a d v c c	2727	
d v c c d	6450	

CMBF – Contd.

- Find the corresponding bloom filter
  - The first byte that was involved in the aggregated fingerprint calculation
- Store the aggregated fingerprints

Theoretical Analysis

- Bloom Filter Collision
- Fingerprint Collision



Theoretical Analysis

- ❑ $FP = \frac{\left(1+255\times\sqrt[g]{a}\right)^{l}-1}{256^{l}}$

- ❑ $a = \left(1 - \left(1 - \frac{1}{m}\right)^{\frac{n}{256g}}\right)$

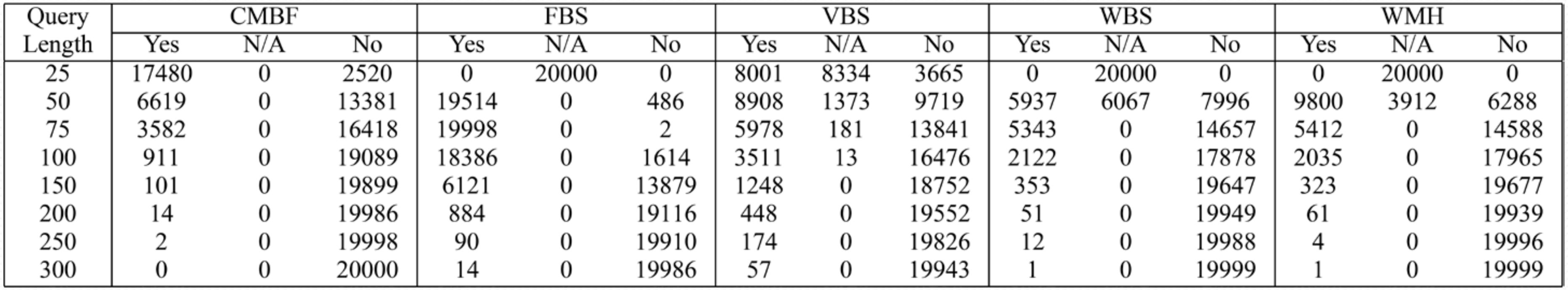
- ❑ $g = \frac{n}{256\times m\times\ln 2}$

- ❑ $m = -\frac{n\times\ln\left(\frac{\sqrt[l]{256^{l}\times FP+1}-1}{255}\right)}{256\times(\ln 2)^{2}}$

Results

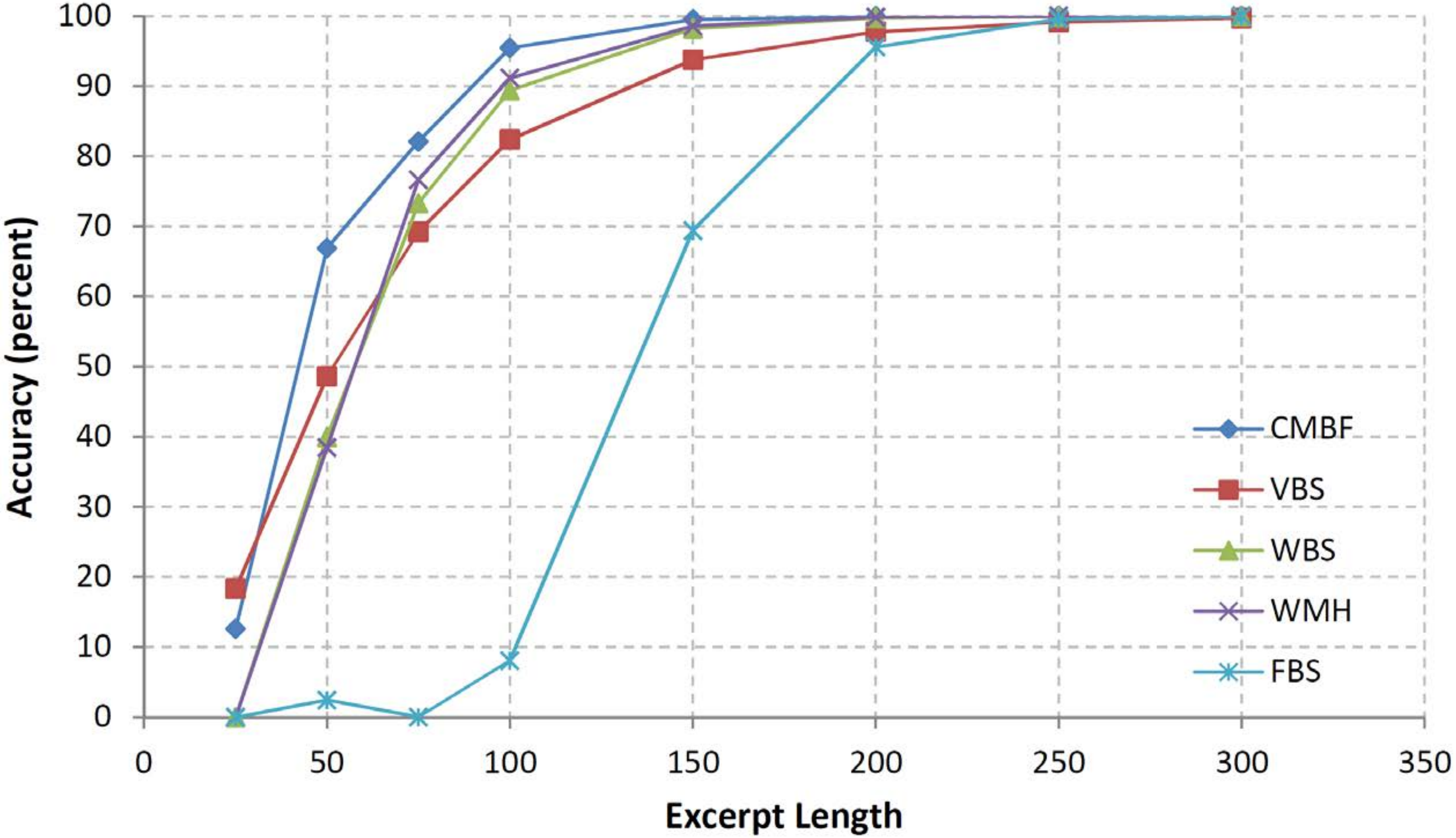
- Querying “S_1…….S_2”
  - CMBF
    - less that 1 second
  - Previous works
    - Estimate more than 4500 years!

Results – Contd.

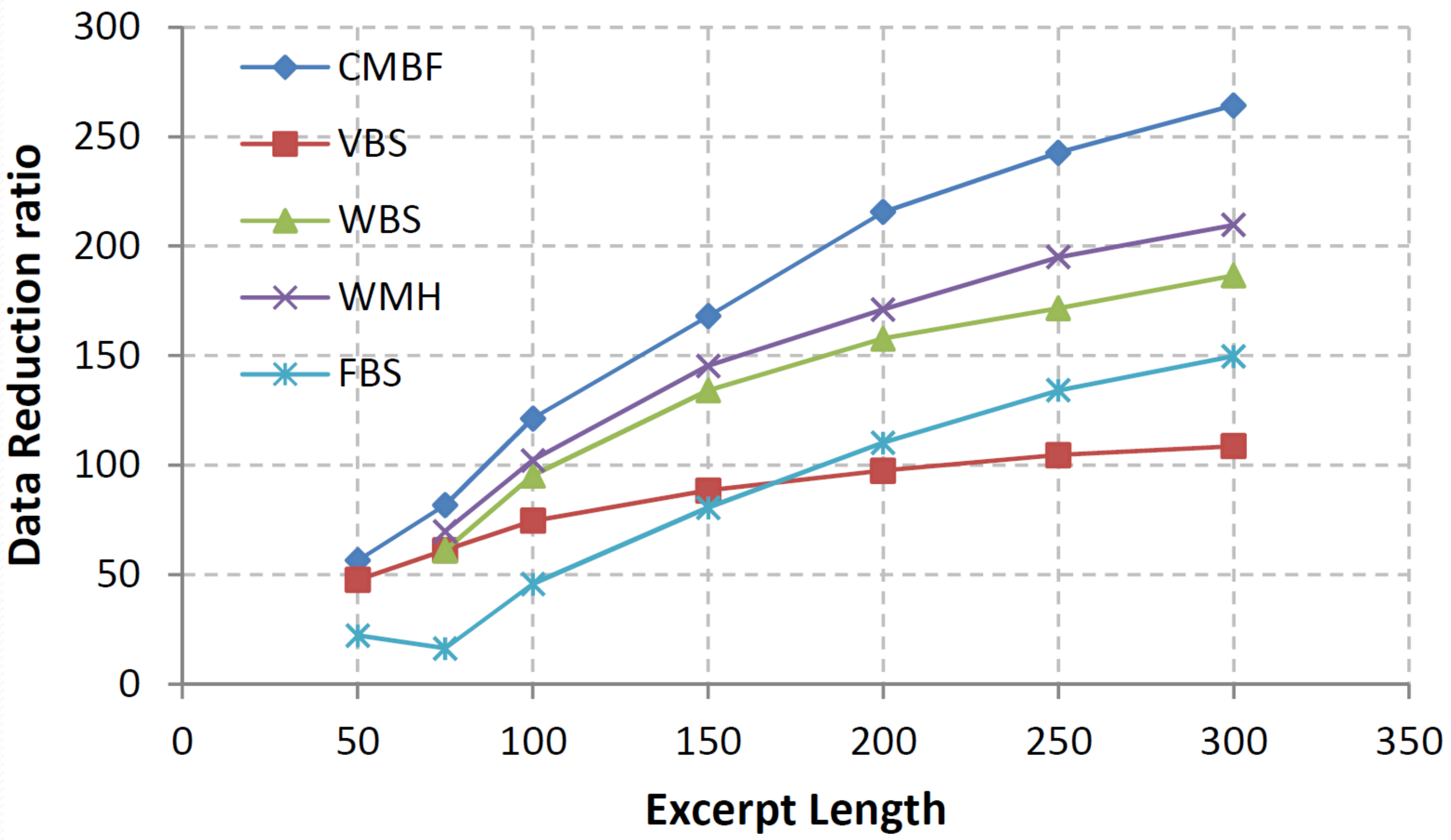


Query	CMBF			FBS			VBS			WBS			WMH		
Length	Yes	N/A	No	Yes	N/A	No	Yes	N/A	No	Yes	N/A	No	Yes	N/A	No
25	17480	0	2520	0	20000	0	8001	8334	3665	0	20000	0	0	20000	0
50	6619	0	13381	19514	0	486	8908	1373	9719	5937	6067	7996	9800	3912	6288
75	3582	0	16418	19998	0	2	5978	181	13841	5343	0	14657	5412	0	14588
100	911	0	19089	18386	0	1614	3511	13	16476	2122	0	17878	2035	0	17965
150	101	0	19899	6121	0	13879	1248	0	18752	353	0	19647	323	0	19677
200	14	0	19986	884	0	19116	448	0	19552	51	0	19949	61	0	19939
250	2	0	19998	90	0	19910	174	0	19826	12	0	19988	4	0	19996
300	0	0	20000	14	0	19986	57	0	19943	1	0	19999	1	0	19999

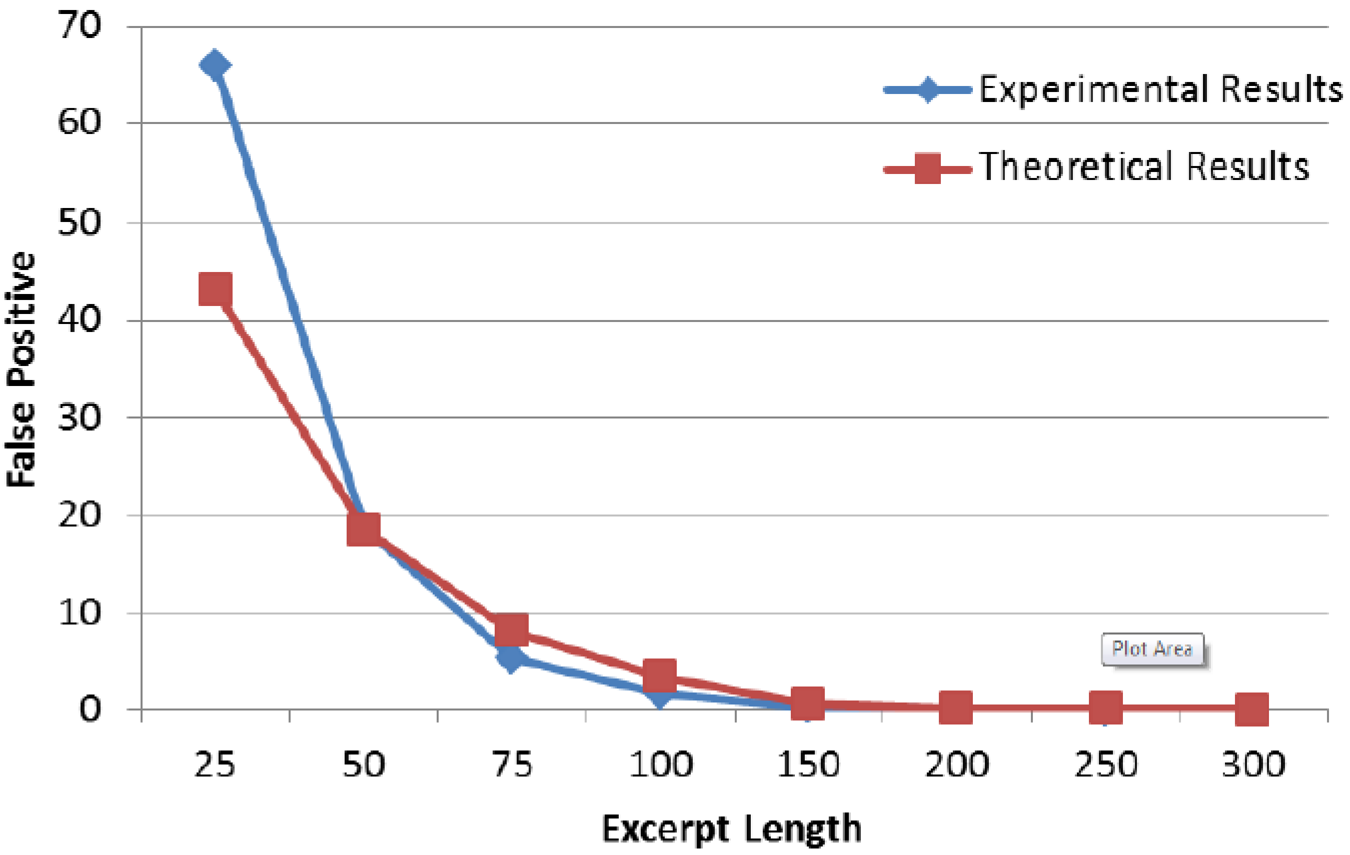
Results – Contd.



Results – Contd.



Results – Contd.



Summary & Discussion

- ❑ Network Forensics
- ❑ Payload Attribution Systems
- ❑ CMBF
- ❑ Just Yes/No answers
  - Who?

Thanks for your attention!