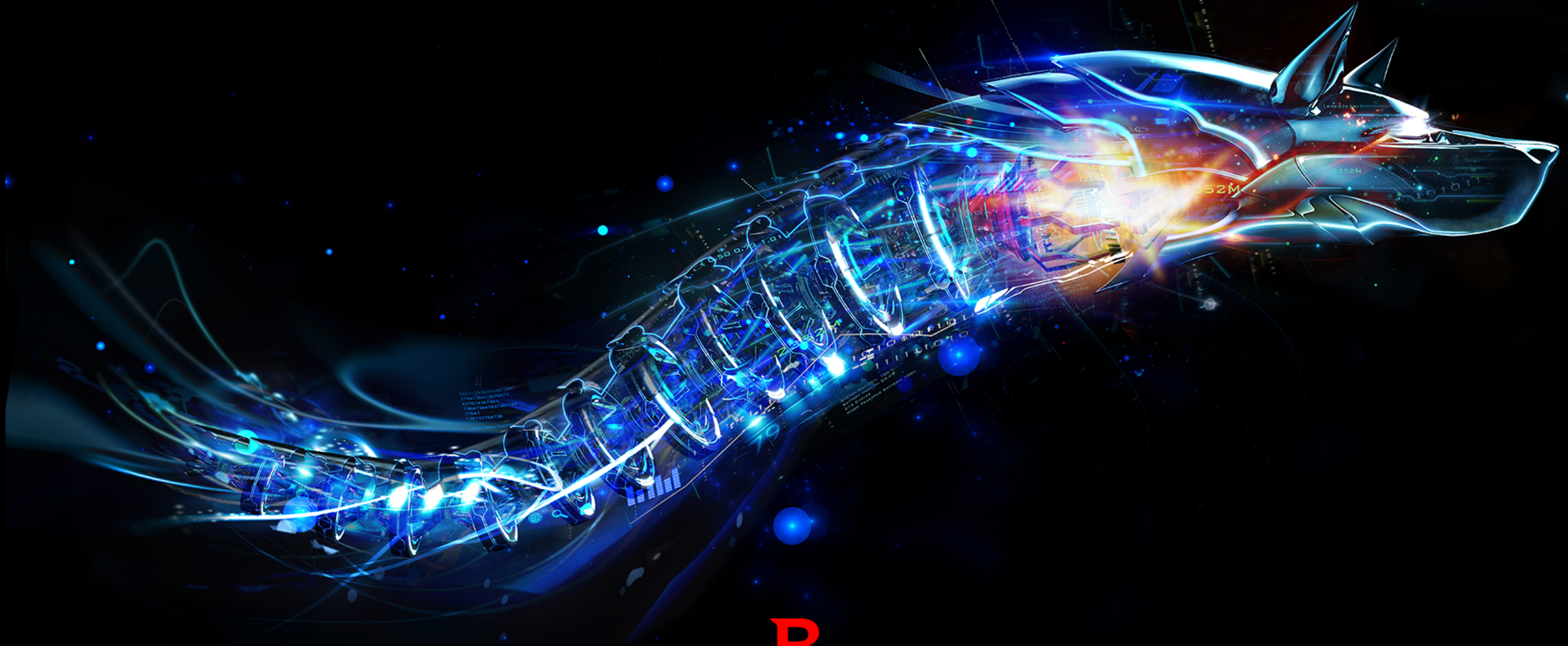


PROTECTING **500 MILLION USERS** WORLDWIDE

Labeling spam through the analysis of protocol patterns

Andrei Hușanu, Alexandru Trifan

Bitdefender®

Outline

- Introduction: about TCP/IP and MSS
- Inception: the initial observations
- The experiment setup
- Protocol patterns
- Pitfalls
- Conclusion: future work

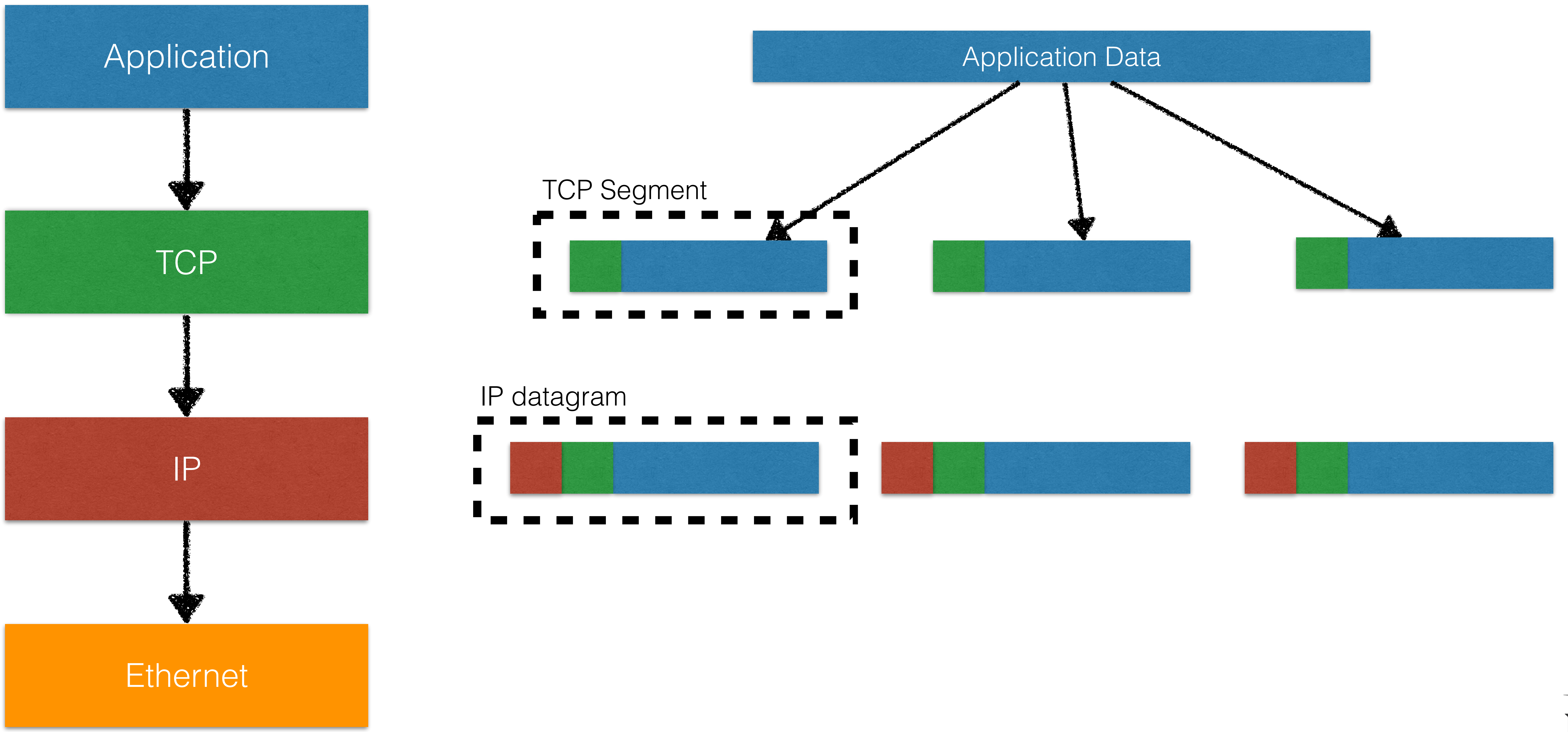
Introduction

- In how many ways can an e-mail message be sent?
- Differences show up at connection level
- The TCP/IP transmission can say a lot about the sender

The TCP/IP protocol

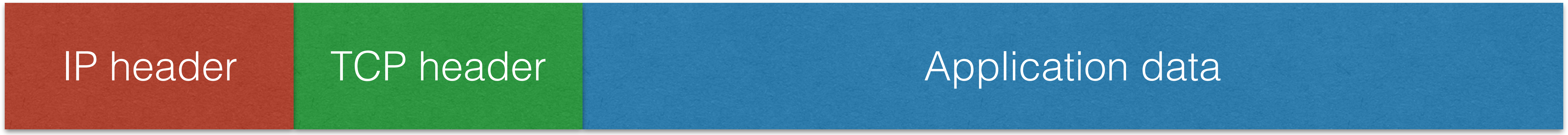
- RAS syndrome
- Redundant Acronym Syndrome … syndrome
- Other notable cases:
    - ATM machine
    - LCD display
    - PIN number

TCP/IP and MSS



Bitdefender

TCP/IP and MSS



```
MSS = MTU - sizeof( TCPHDR ) - sizeof( IPHDR )
```

- MSS - maximum segment size
- Too small: overhead
- Too large: IP fragmentation
- Adjusted dynamically with network

Inception

The four stages of discovery:

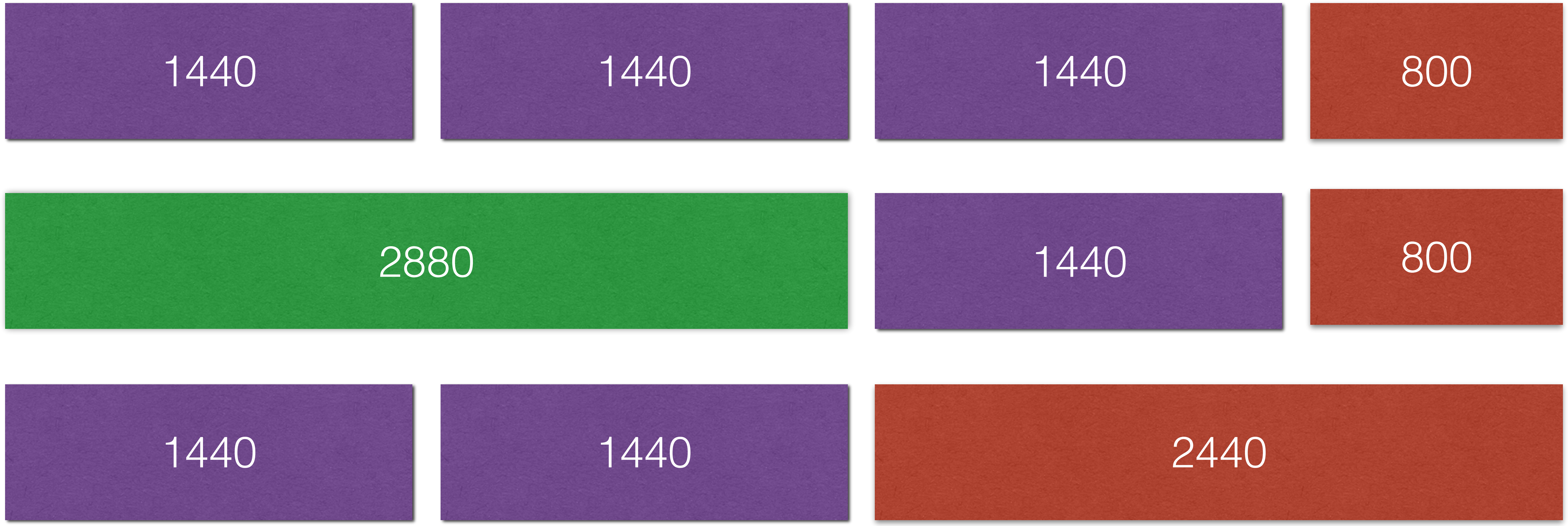
1. Hmmm?
   - Lucky grep through GBs of logs.
2. Hmmm …
   - Notice recurring pattern, figure out the reason.
3. Whoah!
   - Infer similar patterns, confirm through actual data.
4. Let’s write a paper for VB 2014!

The setup

- To better understand TCP/IP and MSS behavior
  - TCP/IP server - East Coast, USA
  - small client - Bitdefender HQ (Romania, Europe)
- To analyze and extract protocol patterns in spam
  - node.js event-oriented SMTP server

TCP behavior #1

```
client.write( buff5k );
```



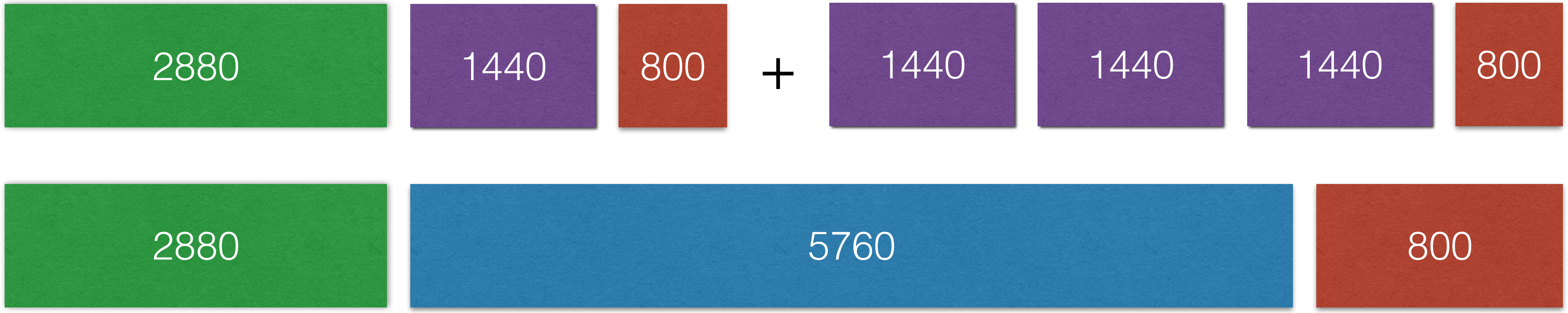
1440 1440 1440 800

2880 1440 800

1440 1440 2440

TCP behavior #2

```
client.write( buff5k );
client.write( buff5k );
```

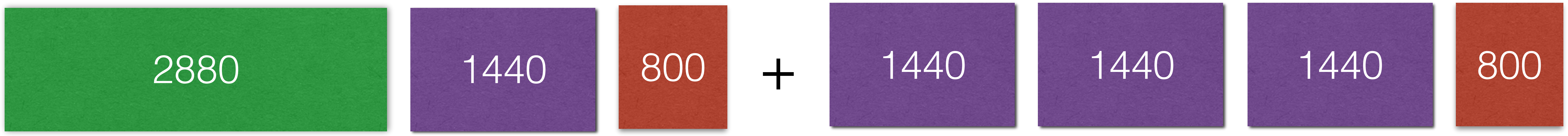


2880 1440 800 + 1440 1440 1440 800

2880 5760 800

TCP behavior #3

```
client.write( buff5k );

setTimeout( function( ) {
    client.write( buff5k );
    client.end( );
}, 100 );
```



2880 1440 800 + 1440 1440 1440 800

SMTP patterns

- Official MTAs
- SMTP relay servers
- Clients which “adjust” the message on the fly
- Other interesting types

Official MTAs

- Persistence - primary objective
- Optimized for congested environments
- Minimize syscalls
- Similar to TCP behavior #1

Official MTAs

1448	'MIME-Version: 1.0\r\nX-Received: b[...]servation League.\r\nIn 1788 a cla'
1448	'ssical theater was built under h[...]January 2007 and returned to NAS'
2896	'A in an administrative position.[...]believe in me, and if you don\'t '
2896	'it\'s too bad. You will be sendin[...]lf to be a hacker at the time. G'
138	'uam to refuel, then hit the enem[...]57969aacd855bb3383b0b9a0f--\r\n.\r\n'

SMTP relay servers

- Similar to official MTAs
- ‘Received’ header stands out

SMTP relay servers

77	`'Received: from 192.168.1.3 (HELO[...]un, 23 Feb 2014 20:37:02 +0800\r\n'`
2896	`'Date: Sun, 23 Feb 2014 20:37:02 [...]IgaHJlZj0iaHR0cDovL2RldGFpbC50\r\n'`
138	`'bWFsbC5jb20vaXRlbS5odG0/c3BtPW[...]xMDU3NTI5ODY3\r\nLmpwZyIgLz4=\r\n.\r\n'`

Spam bots

- Messages generated from templates
- They “adjust” the message on the fly
- Clear correlation with spamming techniques

Spam bots #1

- Headers - in a distinct TCP sequence
- Probably forged

Spam bots #1

520	`'Received: from PC2014021309BEH[1[...]t\r\nContent-Disposition: inline\r\n'`
1440	`'\r\n<!DOCTYPE HTML PUBLIC "-//W3C/[...]</P>\r\n<P style="MARGIN: 0cm 0cm '`
1440	`'0pt" class=MsoNormal><SPAN style[...]NA <o:p></o:p></SPAN></P>\r\n<P st'`
1345	`'yle="MARGIN: 0cm 0cm 0pt" class=[...]al></SPAN></P></BODY></HTML>\r\n\r\n'`
5	`'\r\n.\r\n'`

Spam bots #2

- Custom tailored headers in distinct TCP sequences
- Typical spam bot behavior

Spam bots #2

399	'MIME-Version: 1.0\r\nX-Received: b[...]RPYHYekEni@xxxxx.xxx>\r\nSubject:'
29	'Alert: Best Stock to Buy Now'
184	'\r\nFrom: Alberta Hendrix <stolber [...]a6d6aa2e90de7b468e09e64e2\r\n\r\n'
1440	'--a643d6daa2e90de7b468e09e64e2\r\n[...]licopter. Picton is the only tow'
1377	'n in the Southern Hemisphere whe[...]3d6da2e90de7b468e09e64e2--\r\n.\r\n'

Garage SMTP servers

- Uniform TCP segments (like most MTAs)
- Very small MSS value
- Consistent, does not fluctuate

Garage SMTP servers

920	`'MIME-Version: 1.0\r\nX-Received: b[...]ng. There are several fountains '`
920	`‘Of course your FB feed is more interesting than this.’`
920	`'around the base of the tower and[...]he battle party. Garson obliged '`

...

- Low-tier internet connection?
- Maybe low-reputation server? Home user?

Cool stuff

- Fluctuating MSS
- Frequently dropping to minimal values
- Cellular networks?

Mobile botnets

1400	'MIME-Version: 1.0\r\nReceived: by [...]BlbmdbmVzIGJlY2FtZSBEdXJh\r\ndGVj'
536	'cyB0b28uIFRoZXNlIHRva2VucyBhcmUg[...]IHdpdGggQWppdGggaGF2aW5nIHRvIGNo'
...	
536	'IgdGhlIGZhc2hpb24gb2YgYSBzdG9yZW[...]cyBidWlsZGluZywgYmVzdCBrbm93biBh'
1400	'cyB0aGUgSGVhbHRoIExvZGdlLCBpcyBs[...] of Popular Music. In March 2006'

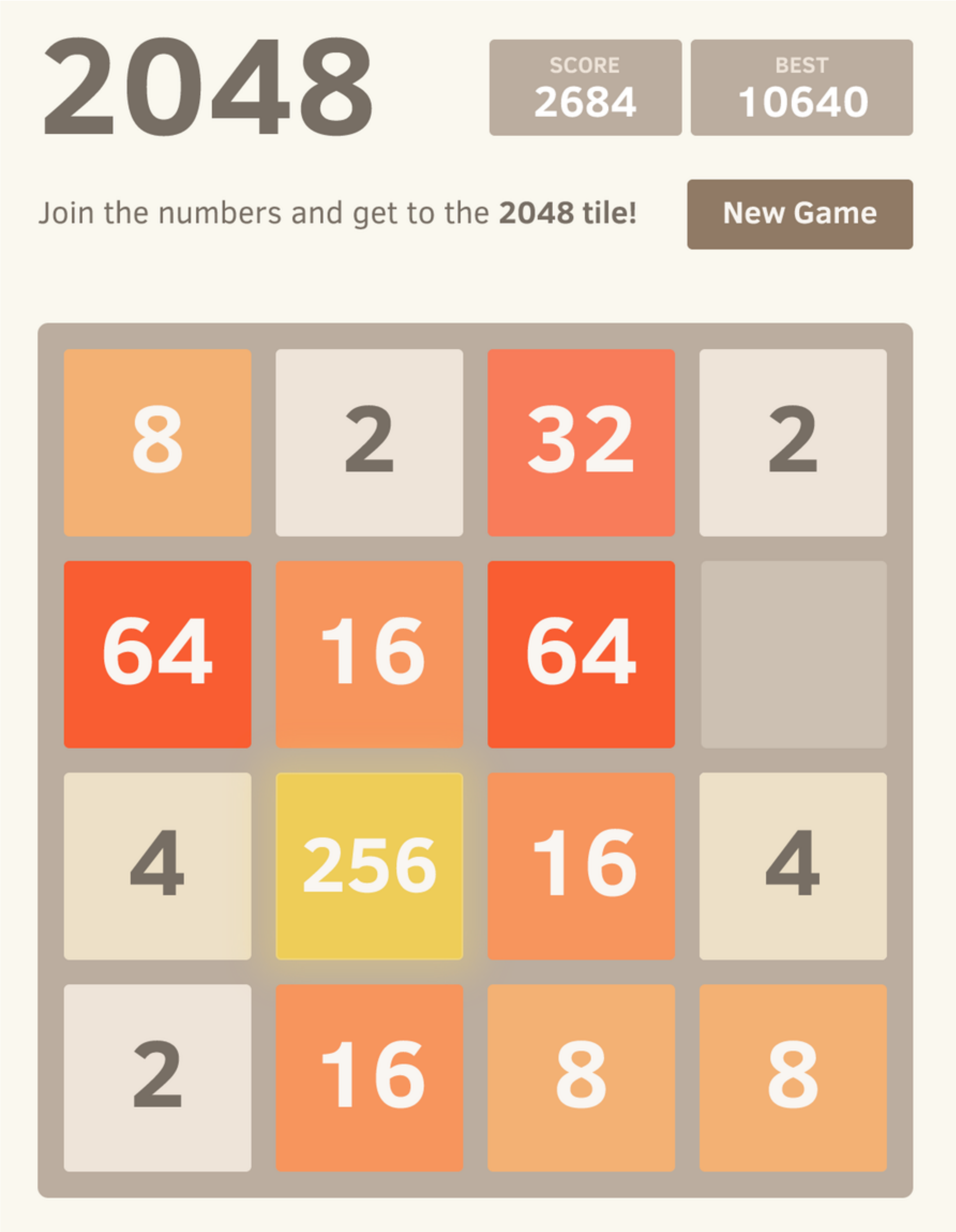
Bitdefender

Mobile botnets

- Mobile spam bots? We checked.
- 90% of IP addresses in pattern - cellular networks
- Classes of IPs from Peru, Turkey
- Open to further investigation

Funny stuff

- Notable distinct sequences
- Multiples of 1024
- Most notably 2048 and 8192



2048

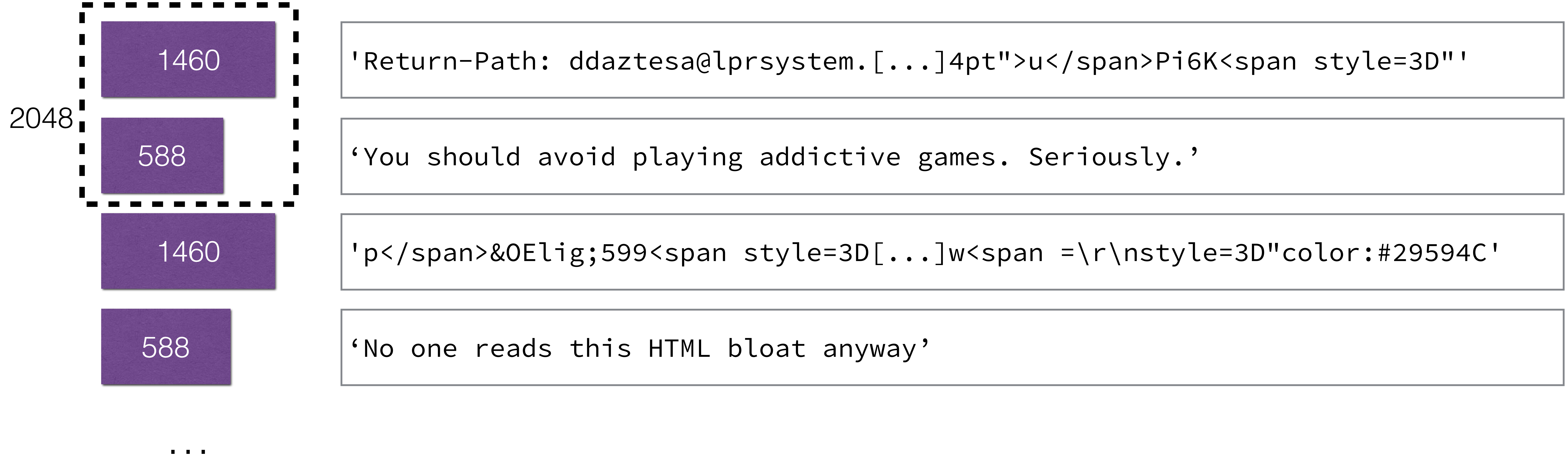
SCORE 2684

BEST 10640

Join the numbers and get to the **2048 tile!**

8	2	32	2
64	16	64	
4	256	16	4
2	16	8	8

2048 fan club



'Return-Path: ddaztesa@lprsystem.[...]4pt">u</span>Pi6K<span style=3D"'

‘You should avoid playing addictive games. Seriously.’

'p</span>&OElig;599<span style=3D[...]w<span =\r\nstyle=3D"color:#29594C'

‘No one reads this HTML bloat anyway’

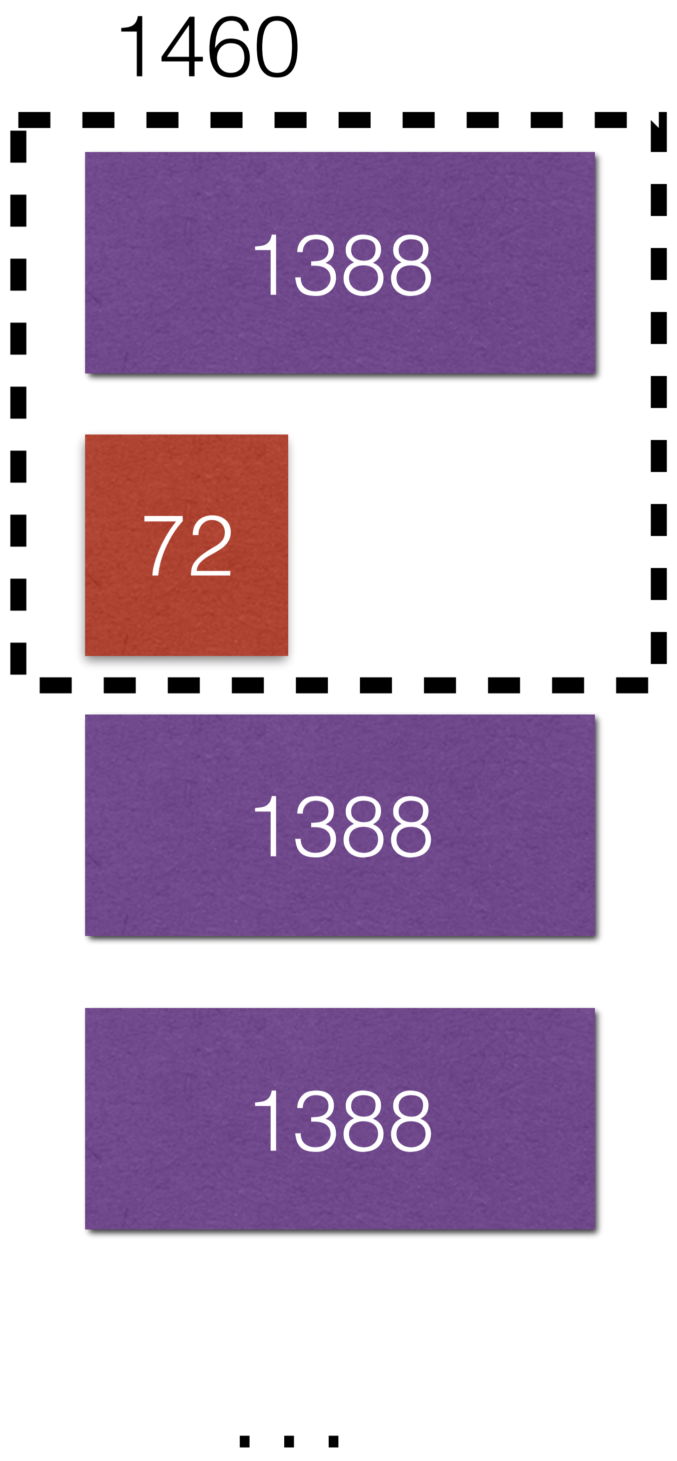
Conclusion

- Pitfalls of this method and how they affect research
- Numbers
- How it helps?

Pitfalls

- Network congestion
- Packet fragmentation
- MSS is adjusted to mitigate
- Apparent “white noise” in data collection

Pitfalls

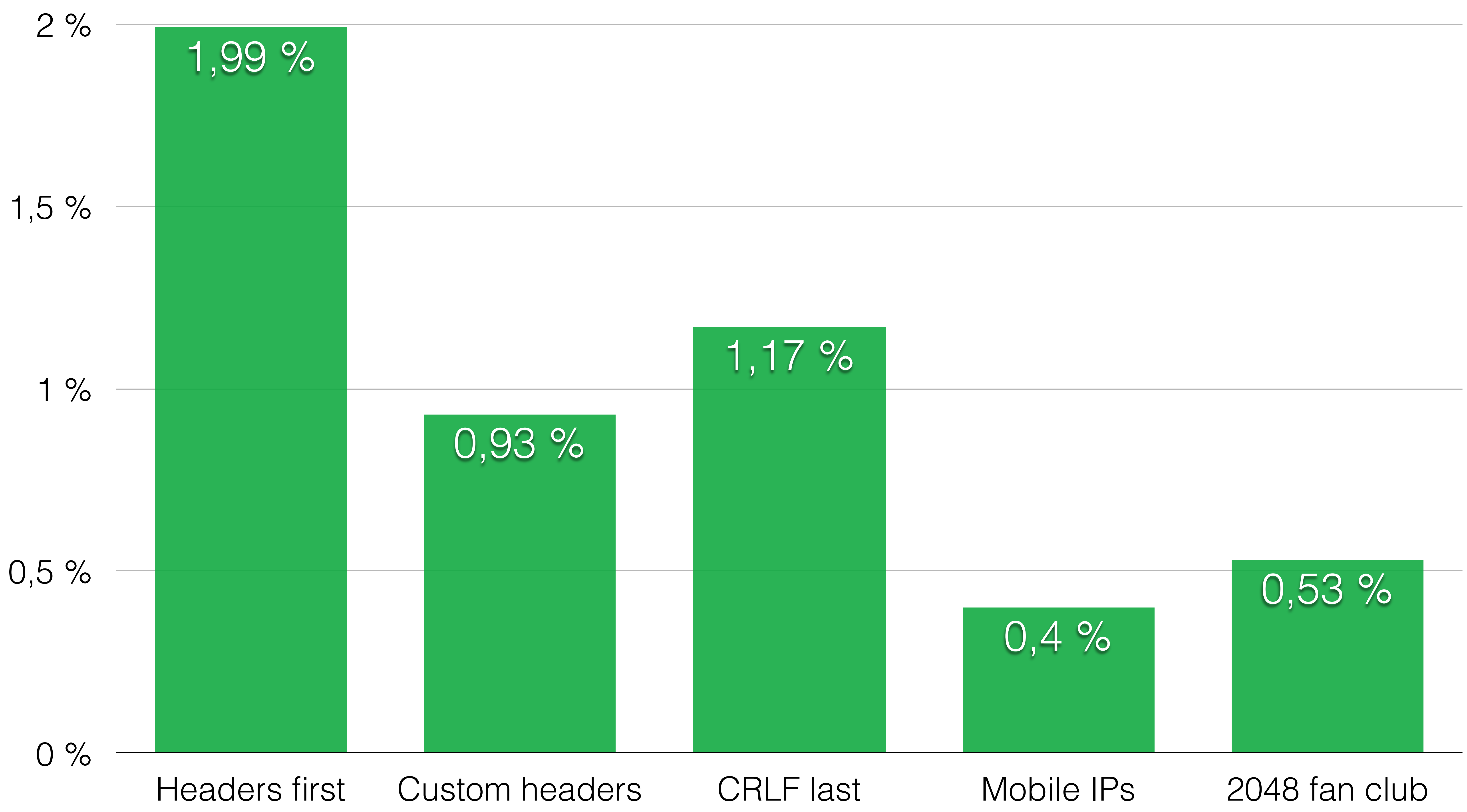


```
'MIME-Version: 1.0\r\nX-Received: b[...]hile, Zoano PharaohMan is captur'

'ing NetNavis and transforming th[...]followers of Gregar, including Z'

'oano SparkMan. Nino Valenti, and[...]an being kidnapped. He married l'

'ocal woman Marie Longley, and be[...]f you\'re tired of playing the ma'
```

Numbers

- Pattern occurence in our spam flow
- Dependent on the samples on which we counted
- 1M spam messages received over 48h
- Spam waves fluctuate

Numbers



Bitdefender

How it helps?

- Postfix: Postscreen - heuristic connection-level rules to filter out spam
- Better botnet detection
- Server / IP reputation
- Improving our image on threats from mobile devices
- Other ideas?

Alexandru Trifan - lex@bitdefender.com

Andrei Hușanu - ahusanu@bitdefender.com

Bitdefender®