black hat ASIA 2016

Practical New Developments in the BREACH Attack

Dimitris Karakostas
Dionysis Zindros

erc

UBM

Who are we?

Dimitris Karakostas & Dionysis Zindros

Researchers at Security & Crypto lab
University of Athens, Greece

HTTPS is **broken**

- BREACH broke HTTPS + RC4 in 2013
- People upgraded to AES – thought they were safe

Today...

- We show TLS + AES is **still broken**
- **HTTPS can be decrypted**
- We launch **open source tool** to do it here in Singapore

Overview

- BREACH review
- Our contributions
- Statistical attacks
- Attacking block ciphers
- Attacking noise
- Optimization techniques
- Our tool: Rupture
- Mitigation recommendations

Original BREACH research

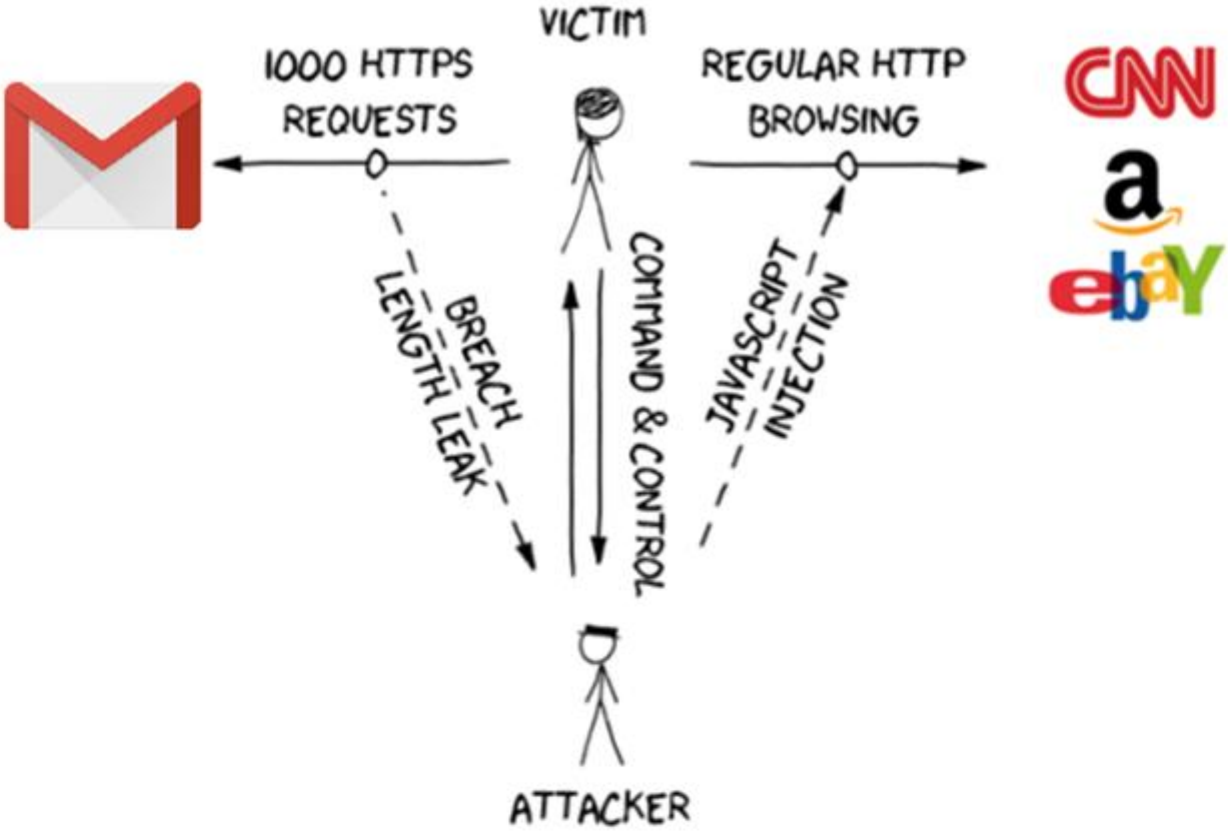
Introduced in Black Hat USA 2013

Angelo Prado

Neal Harris

Yoel Gluck

BREACH attack anatomy


VICTIM
1000 HTTPS REQUESTS
REGULAR HTTP BROWSING
BREACH LENGTH LEAK
COMMAND & CONTROL
JAVASCRIPT INJECTION
ATTACKER

Original BREACH assumptions

Target website:

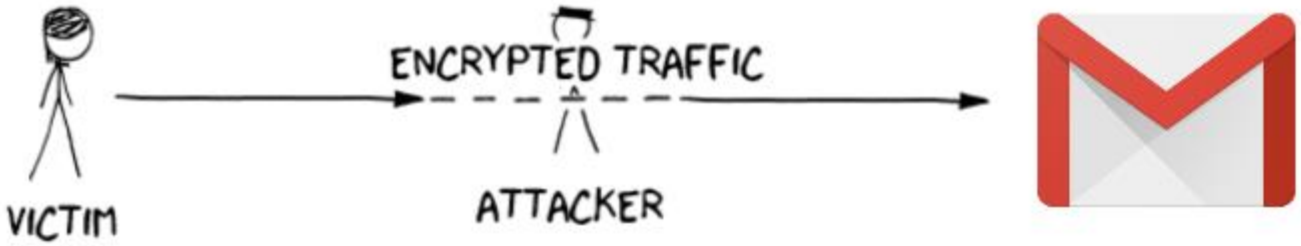
- Uses **HTTPS**
- Compresses response using **gzip**
- Uses **stream cipher**
- Response has **zero** noise
- Contains end-point that **reflects** URL parameter

Original BREACH target

1. Steal **secret** in HTTPS response (CSRF tokens)
2. Use CSRF to impersonate victim client to victim server

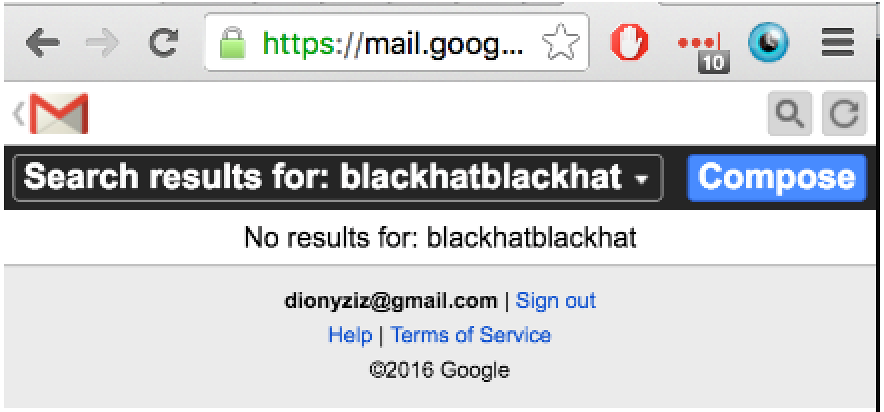
Length leaks

$$|E(A)| < |E(B)| \Leftrightarrow |A| < |B|$$



Let’s attack Gmail

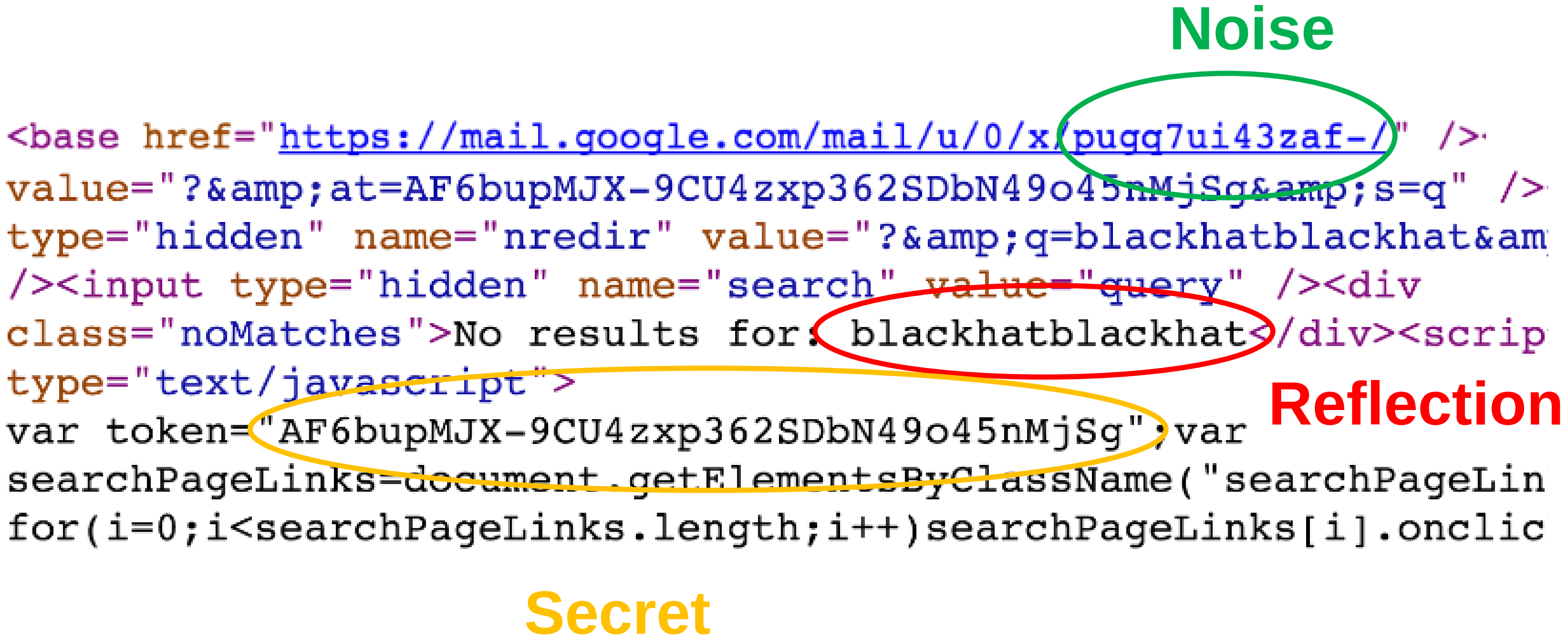
- **m.gmail.com** mobile Gmail view
- Mobile search functionality uses HTTP POST – but HTTP GET still works :)
- CSRF token included in response – valid for all of Gmail



Search results for: blackhatblackhat

No results for: blackhatblackhat

dionyziz@gmail.com | Sign out
Help | Terms of Service
©2016 Google



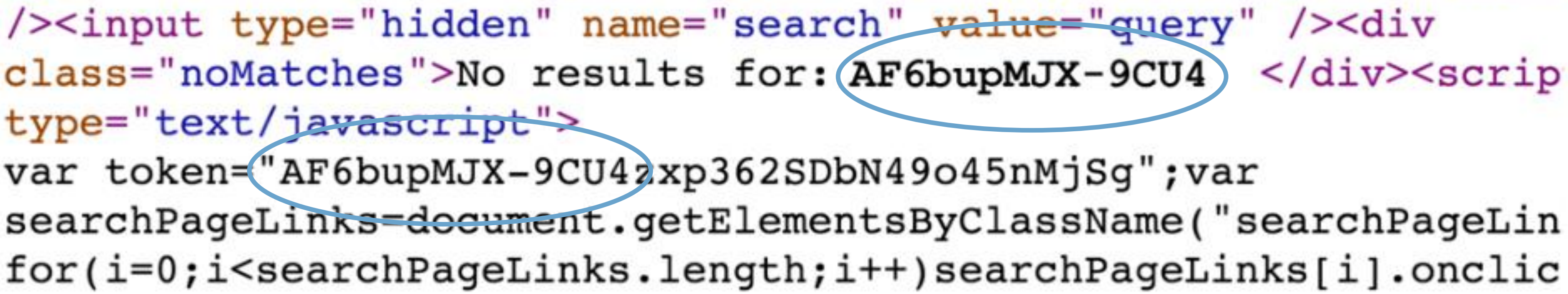
```
<base href="https://mail.google.com/mail/u/0/x/pugq7ui43zaf-/" />
value="?&amp;at=AF6bupMJX-9CU4zxp362SDbN49o45nMjSg&amp;s=q" />
type="hidden" name="nredir" value="?&amp;q=blackhatblackhat&am
/><input type="hidden" name="search" value="query" /><div
class="noMatches">No results for: blackhatblackhat</div><scrip
type="text/javascript">
var token="AF6bupMJX-9CU4zxp362SDbN49o45nMjSg";var
searchPageLinks=document.getElementsByClassName("searchPageLin
for(i=0;i<searchPageLinks.length;i++)searchPageLinks[i].onclic
```

Noise

Reflection

Secret

- Attacker **guesses part of secret**
- Uses it in **reflection**
- Compressed/encrypted response **is shorter** if right!



```
/><input type="hidden" name="search" value="query" /><div
class="noMatches">No results for: AF6bupMJX-9CU4 </div><scrip
type="text/javascript">
var token="AF6bupMJX-9CU4zxp362SDbN49o45nMjSg";var
searchPageLinks=document.getElementsByClassName("searchPageLin
for(i=0;i<searchPageLinks.length;i++)searchPageLinks[i].onclic
```

Original BREACH methodology

- **Guess part of secret and insert into reflection**
- **Match**? → **Shorter** length due to compression
- **No match**? → **Longer** length
- **Bootstrap** by guessing 3-byte sequence
- Extend **one character** at a time
- O(n|Σ|) complexity
  - **n**: length of secret
  - **Σ**: alphabet of secret

Our contributions

Our contributions

We extend the BREACH attack

1. Alternative secrets
2. Attack **noisy** end-points
3. Attack **block cipher** end-points
4. **Optimize** attack
5. Novel **mitigation** techniques

Alternative secrets

- Not only CSRF tokens can be stolen
- Gmail email bodies
- Facebook chat messages
- Anything!
- Masking CSRF tokens is not enough

Statistical methods

Statistical methods

- We can attack **noisy** end-points
- Multiple requests per alphabet symbol
- Take **mean response length**
- **m**-sized noise → attack works in $O(n|\Sigma|\sqrt{\mathbf{m}})$
  - m = (max response size) - (min response size)
- Length converges to correct results (LLN)

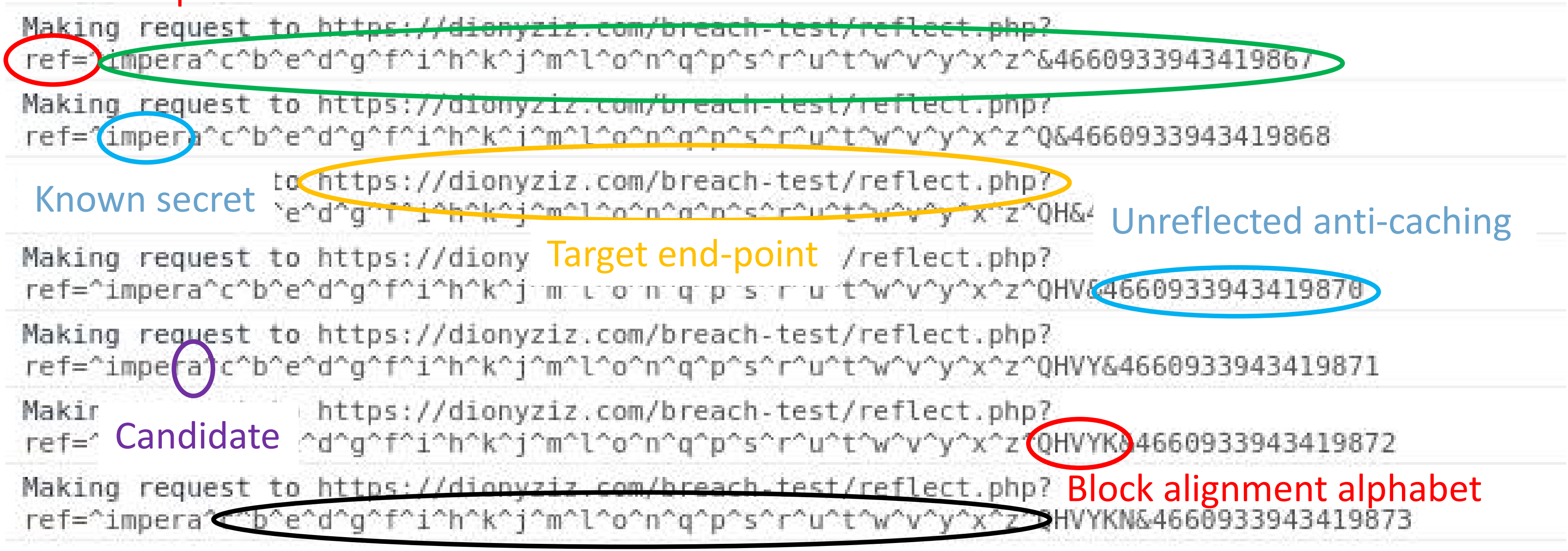
Statistical methods against block ciphers

- Everyone uses block ciphers
- Statistical methods break them
- We introduce **artificial noise**
- Block ciphers round length to 128-bits
- In practice **16x more requests**
- Blocks aligned → Length difference measurable

Block alignment with artificial noise

- For each candidate, send 16 requests
- Pad each request with **artificial noise**
- **0…15** additional random bytes in reflection
- This will cross a **block boundary**
- Ideally, symbols that don’t appear elsewhere

One sampleset in a batch: A single candidate (‘a’)



```
Making request to https://dionyziz.com/breach-test/reflect.php?
ref=^impera^c^b^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^&466093394341986
Making request to https://dionyziz.com/breach-test/reflect.php?
ref=^impera^c^b^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^Q&466093394341986
                 to https://dionyziz.com/breach-test/reflect.php?
                 ^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^QH&4
Making request to https://diony                  /reflect.php?
ref=^impera^c^b^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^QHV&466093394341987
Making request to https://dionyziz.com/breach-test/reflect.php?
ref=^impera^c^b^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^QHVY&466093394341987
Makir                  https://dionyziz.com/breach-test/reflect.php?
ref=^                  ^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^QHVYK&466093394341987
Making request to https://dionyziz.com/breach-test/reflect.php?
ref=^impera^c^b^e^d^g^f^i^h^k^j^m^l^o^n^q^p^s^r^u^t^w^v^y^x^z^QHVYKN&466093394341987
```

Reflected parameter

Reflected value

Known secret

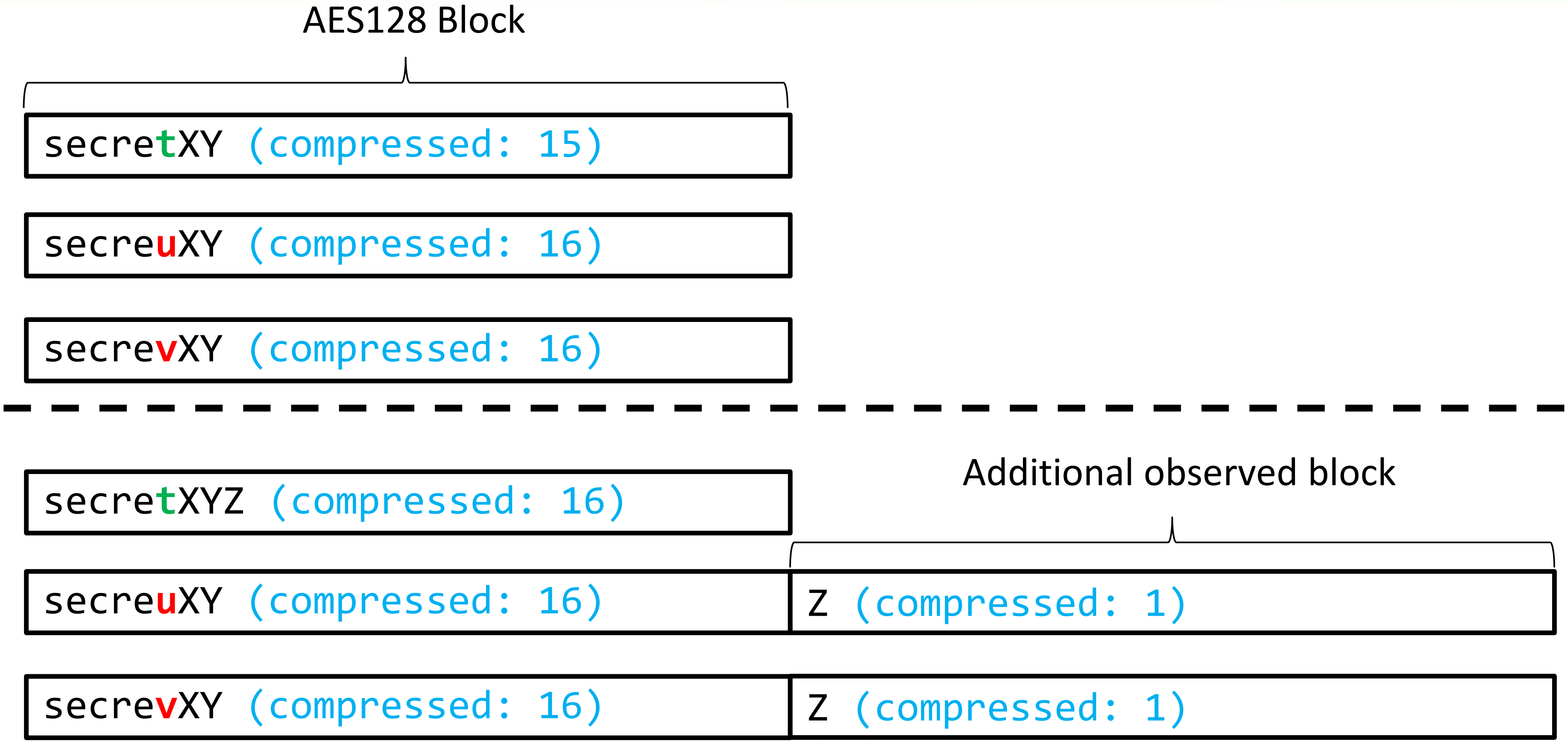
Target end-point

Unreflected anti-caching

Candidate

Block alignment alphabet

Huffman pool



AES128 Block

secretXY (compressed: 15)

secreuXY (compressed: 16)

secrevXY (compressed: 16)

secretXYZ (compressed: 16)

Additional observed block

secreuXY (compressed: 16) | Z (compressed: 1)

secrevXY (compressed: 16) | Z (compressed: 1)

Experimental results

- **AES_128 is vulnerable**
- Popular web services are vulnerable:
  - Gmail
  - Facebook
  - etc.

Noise generators

- Noise = Response part that changes per request

- Web app noise: Timestamps, random token
- Connection: close / keep-alive
- Huffman header encoding
  - Huffman tree changes due to block alignment padding :(
  - We can’t predict how it changes – plaintext unknown
- Content-encoding: chunked – boundaries may change

Optimizations

Optimizations overview

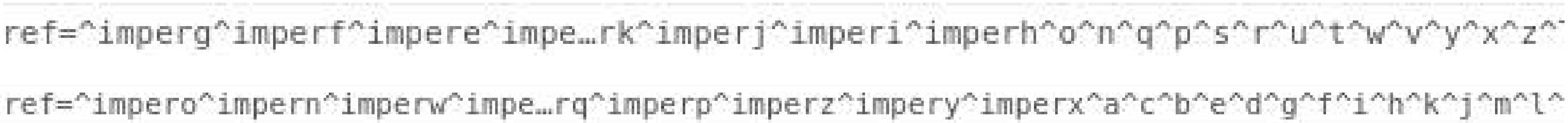
Block ciphers cause min 16x slowdown. We need to optimize.

- **Divide and conquer**: 6x speed-up
- **Request soup**: 16x speed-up
- **Browser parallelization**: 6x speed-up

Total ~ 500x speed-up!

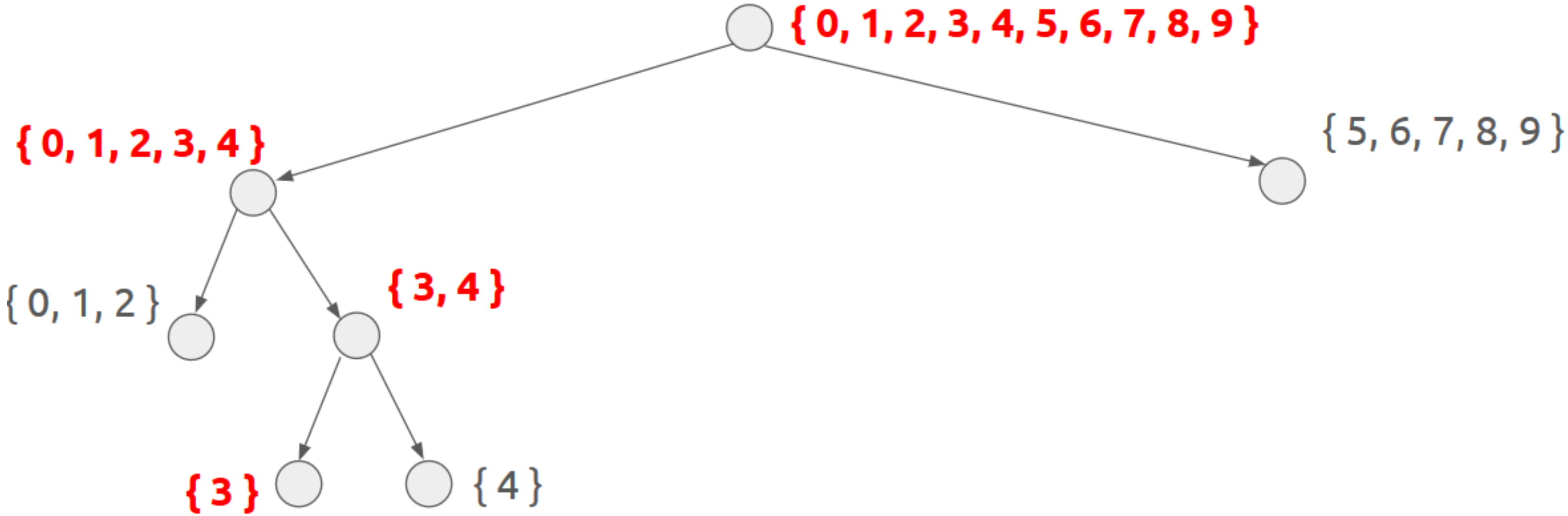
Optimization: Divide & Conquer

- Each request tries multiple candidates from alphabet
- Partition alphabet using divide-and-conquer
- Binary search on alphabet partitions
- Reduces attack complexity from **$O(n|\Sigma|)$** to **$O(n \lg|\Sigma|)$**
- Practically this gives **6x speed-up**



```
ref=^imperg^imperf^impere^impe…rk^imperj^imperi^imperh^o^n^q^p^s^r^u^t^w^v^y^x^z^
ref=^impero^impern^imperw^impe…rq^imperp^imperz^impery^imperx^a^c^b^e^d^g^f^i^h^k^j^m^l^
```

Binary search in alphabet space



Optimization: Request soup

Problem:

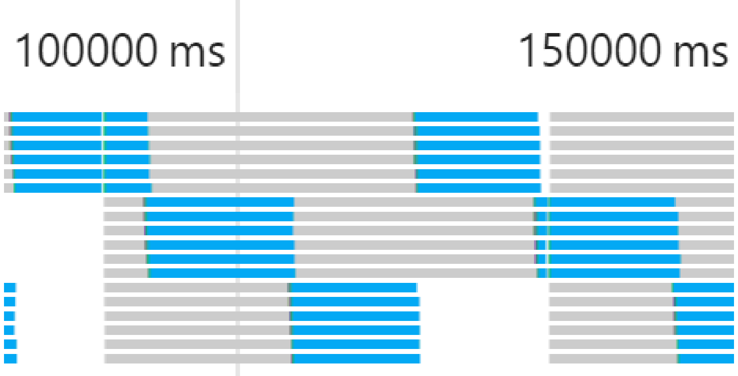
- Need 16x samples for block ciphers
- But we only need the ***length mean***

Solution:

- Responses come pipelined, can’t tell them apart
- We don’t care! Measure total length
- Divide by amount, extract mean

Optimization: Browser parallelization

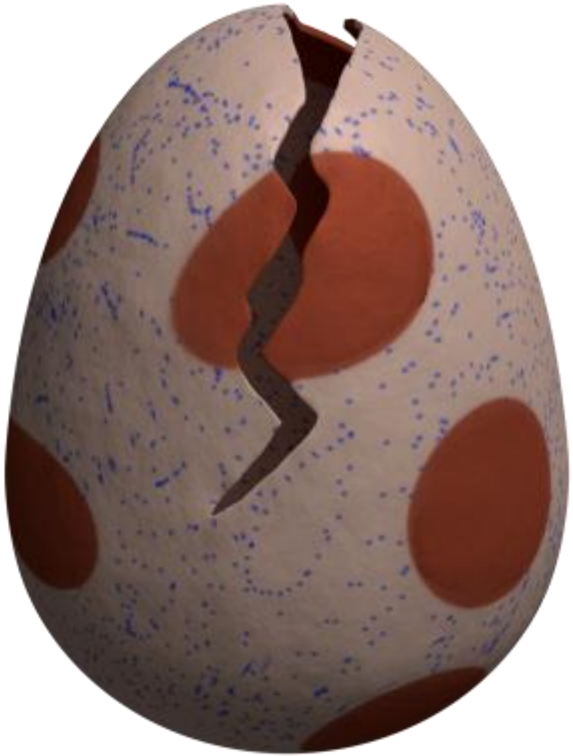
- Do 6x parallel requests; browsers support it
- Each parallel request cannot adapt based on previous
- But we need many samples of same candidates anyway
- No need to adapt before we collect enough



Request soup + browser parallelization:

16 requests in 1.5 sec

(in good network)



Rupture

Today, we make BREACH easy

- Over the past months, we’ve developed **Rupture**
- Today in Black Hat Asia 2016, we launch it
- **Open source:** MIT licensed

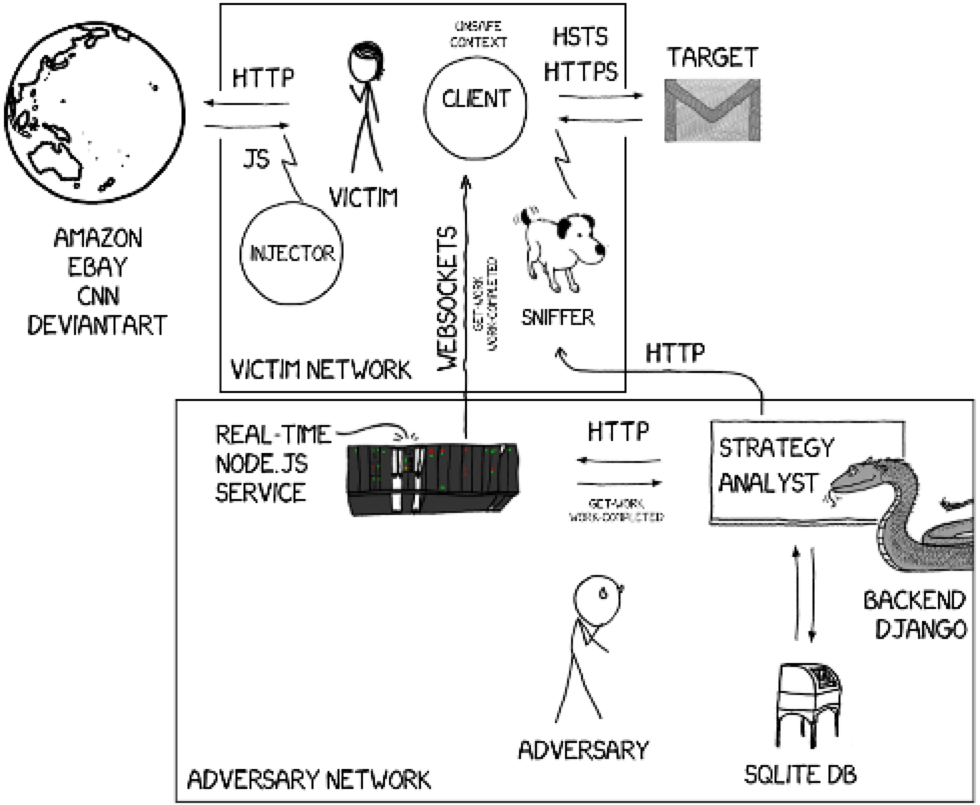
https://github.com/dionyziz/rupture

ruptureit.com

Rupture

- Extensible
  - Modular analysis / optimizations / strategies
  - Experiment with your own
- General web attack framework
  - Can be adapted to work for CRIME, POODLE, …
  - Persistent command & control channel
- Scalable architecture: Multiple attacks simultaneously

RUPTURE ARCHITECTURE



Robust, persistent command & control

- Automatically inject JS to HTTP
- All plaintext connections infected
- One tab at a time gets work from C&C server
- User closes tab? **Different tab** starts attacking
- User switches browsers? Works on **different browser**
- Data collection failed for a sample? Sample **recollected**
- User reboots computer? **Attack continues**

Persistent attack data storage

- Collected data processed by Django middleware
- Attack historical data **stored permanently** in SQLite db
- Future analysis with new techniques possible

Rupture demo

Statistically expected* runtime

- Assuming limited noise:
- Using sequential technique: 3 min / byte
    - 3 batches per candidate
- Using divide & conquer: 36 sec / byte

* Additional batches may be needed if confidence is low

Mitigation

First-party cookies

- Don’t send auth cookies cross-origin
- Backwards compatibility: Web server opts-in
- Mike West implemented it in Chrome 51
- Coming April 8th

Set-Cookie: SID=31d4d96e407aad42; **First-Party**

Future work

- Responsible disclosure:
  - Publish specific preconfigured Rupture “targets” – Gmail, Facebook, etc.
  - In coordination with web app developers
- Implement First-Party cookies in Firefox and other browsers
- Extend Rupture with other attacks: CRIME, etc.
- Implement SPDY support for Rupture
- Backtracking
- Come help us make Rupture better – many bugs on GitHub

Key takeaways

1. HTTPS + gzip = **broken**
2. Rupture framework is live – **attacks are easy**
3. Enable **first-party cookies** on your web app

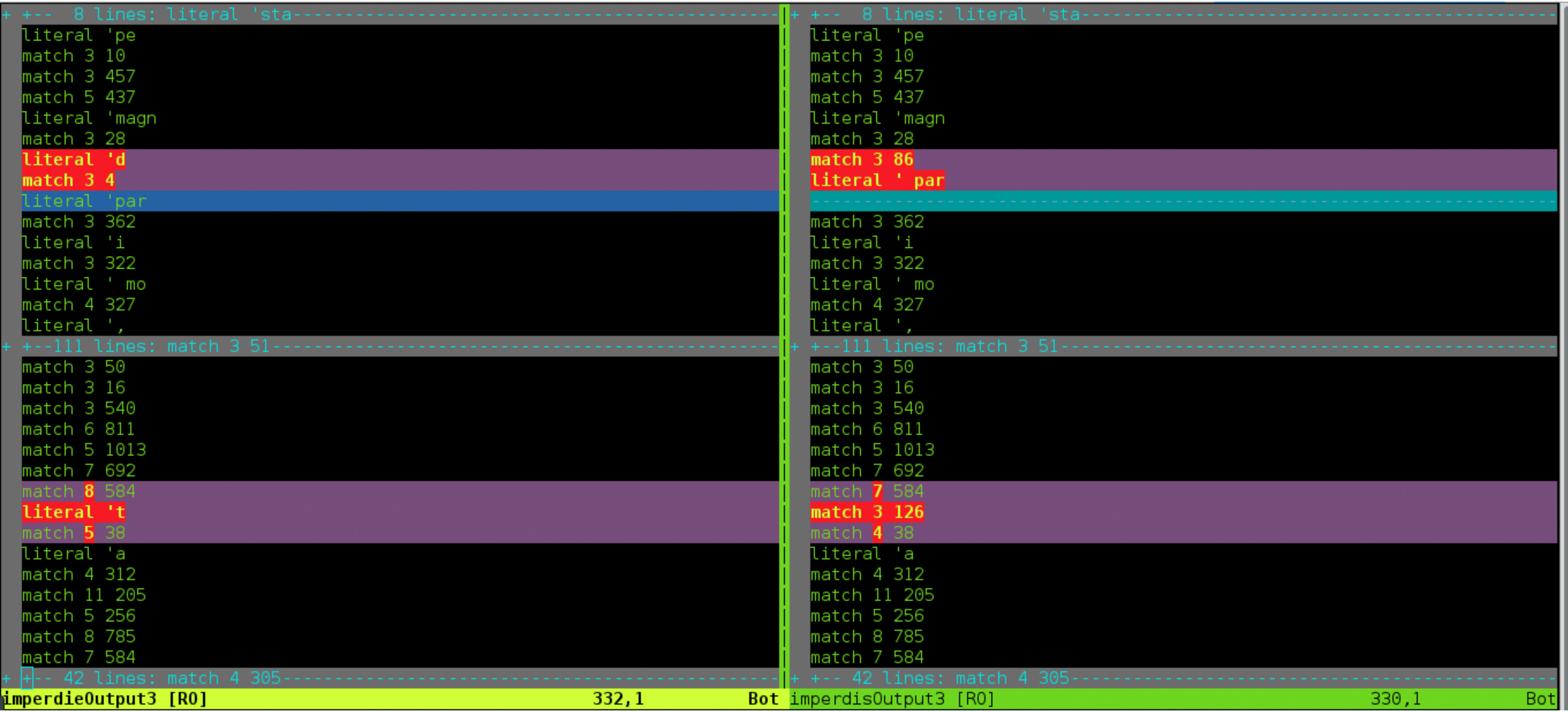
Thank you! Questions?

twitter.com/dionyziz

45DC 00AE FDDF 5D5C B988 EC86 2DA4 50F3 AFB0 46C7

github.com/dimkarakostas

DF46 7AFF 3398 BB31 CEA7 1E77 F896 1969 A339 D2E9



```
+ +--  8 lines: literal 'sta-------------------------------------------
  literal 'pe
  match 3 10
  match 3 457
  match 5 437
  literal 'magn
  match 3 28
  literal 'd
  match 3 4
  literal 'par
  match 3 362
  literal 'i
  match 3 322
  literal ' mo
  match 4 327
  literal ',
+ +--111 lines: match 3 51---------------------------------------------
  match 3 50
  match 3 16
  match 3 540
  match 6 811
  match 5 1013
  match 7 692
  match 8 584
  literal 't
  match 5 38
  literal 'a
  match 4 312
  match 11 205
  match 5 256
  match 8 785
  match 7 584
+ +-- 42 lines: match 4 305--------------------------------------------
imperdieOutput3 [RO]                                  332,1          Bot
```

```
+ +--  8 lines: literal 'sta-------------------------------------------
  literal 'pe
  match 3 10
  match 3 457
  match 5 437
  literal 'magn
  match 3 28
  match 3 86
  literal ' par
  ---------------------------------------------------------------------
  match 3 362
  literal 'i
  match 3 322
  literal ' mo
  match 4 327
  literal ',
+ +--111 lines: match 3 51---------------------------------------------
  match 3 50
  match 3 16
  match 3 540
  match 6 811
  match 5 1013
  match 7 692
  match 7 584
  match 3 126
  match 4 38
  literal 'a
  match 4 312
  match 11 205
  match 5 256
  match 8 785
  match 7 584
+ +-- 42 lines: match 4 305--------------------------------------------
imperdisOutput3 [RO]                                  330,1          Bot
```