# Large-scale Security Analysis of the Web: Challenges and Findings

by Tom Van Goethem



DistriNet
iMinds
KU LEUVEN

# About me

- Tom Van Goethem
- PhD student at University of Leuven
  - Large-scale web security experiments
- Security researcher
- Blog: *https://vagosec.org*
-  : *@tomvangoethem*

# Contents

- **Chapter 1: The Experiment**
  - Research goal
  - Experiment setup
  - Security scoring system
- **Chapter 2: Security on the Web - the Good, Bad and Ugly**
  - Security features
  - Experiment findings
- **Chapter 3: Lessons learned**
  - Limitations & challenges
  - Conclusion

# Chapter 1: The Experiment

# Chapter 1: The Experiment

- **Research goal**
- Experiment setup
- Security scoring system

# Research goal

- How secure is the web?
- Are websites from one country more secure than from another?
- How does one quantify security?
- Where do things go wrong?

# Chapter 1: The Experiment

- Research goal
- **Experiment setup**
- Security scoring system

# Experiment setup

- 28 EU member states
  - different demographics
- 1,000 websites per country (ccTLD)
  - 22,851 websites
- up to 200 pages
  - in total: 3 million webpages

# Experiment setup

- Distributed crawler using PhantomJS
  - 60 machines; 5 days
- Check presence of security features and weaknesses
  - 8 security mechanisms
  - 10 vulnerabilities and weaknesses

# Chapter 1: The Experiment

- Research goal
- Experiment setup
- **Security scoring system**

# Security scoring system

- Give a *security score* to a website
  - based on Common Weakness Scoring System (CWSS)
  - for each weakness: score related to business impact, technical impact, likelihood of discovery & exploitability, ...
  - for security mechanisms: calculate score of vulnerabilities it prevents
  - website positive score: sum of security mechanism scores
  - website negative score: sum of weakness scores

# Security scoring system

| Defense mechanism | Score |
|---|---|
| Content Security Policy | 58.93 |
| X-Frame-Options | 45.21 |
| HTTP Strict-Transport-Security | 33.52 |
| X-Content-Type-Options | 8.02 |

| Vulnerability/weakness | Score |
|---|---|
| Vulnerable remote JS inclusion | 67.50 |
| Sensitive files | 41.81 |
| Outdated CMS | 18.30 |
| Information leakage | 9.44 |

# Chapter 2: Security on the web the Good, Bad and Ugly

# Chapter 2: Security on the web

- **Security features**
  - **Defence mechanisms**
  - Weaknesses & vulnerabilities
- Findings

# Defence mechanisms

- 8 defence mechanisms

HTTP Strict-Transport-Security
Secure cookie
→ Transport layer security

Content Security Policy
HttpOnly cookie
X-Content-Type-Options
→ Cross-Site Scripting

X-Frame-Options
Iframe sandboxing
CSRF tokens
→ Miscellaneous

# Defence mechanisms

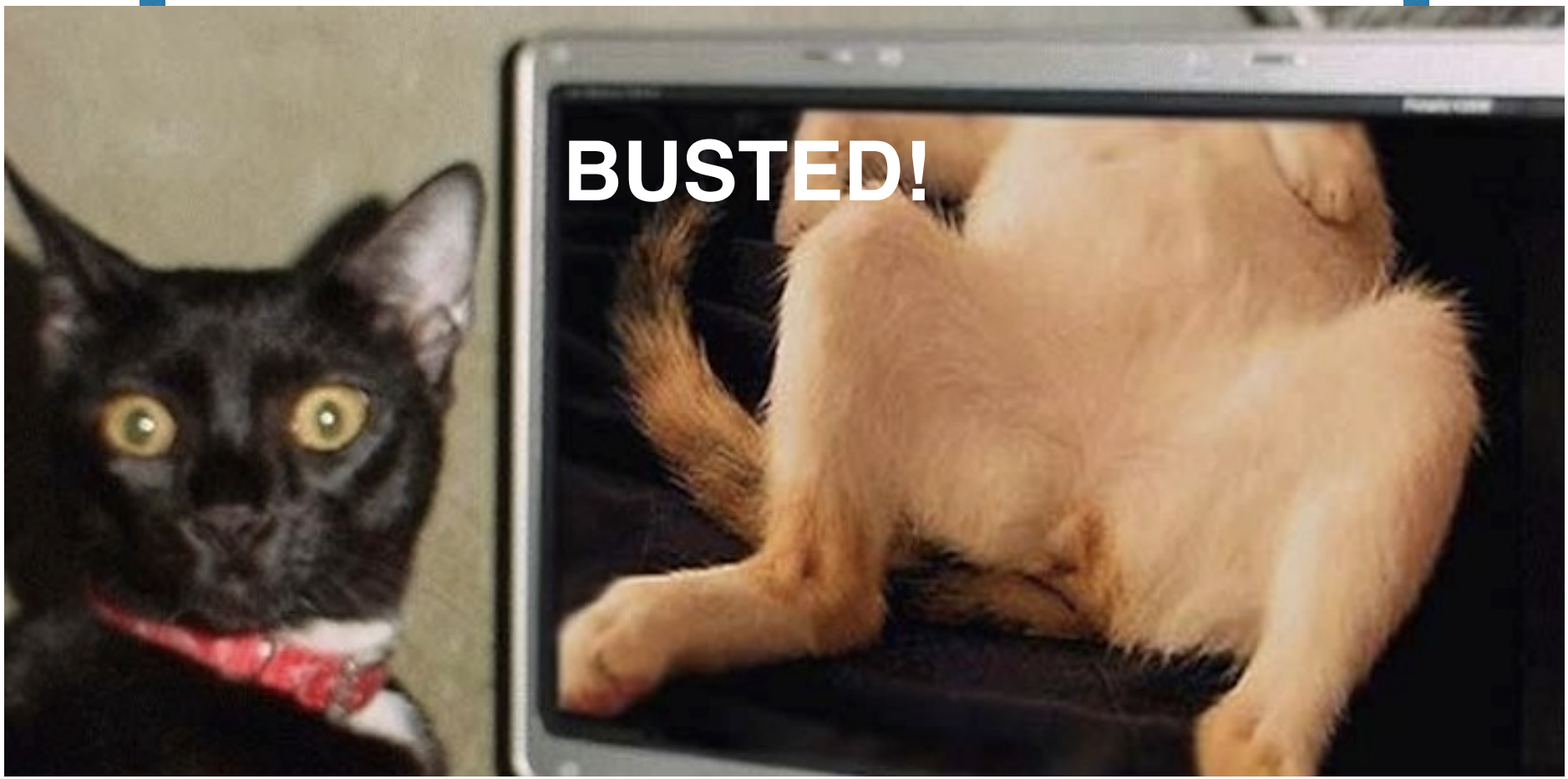
HTTP Strict Transport Security


User
sexy-cats.com


User
sexy-cats.com
1. visit **http**://sexy-cats.com

**User** **sexy-cats.com**

1. visit **http**://sexy-cats.com



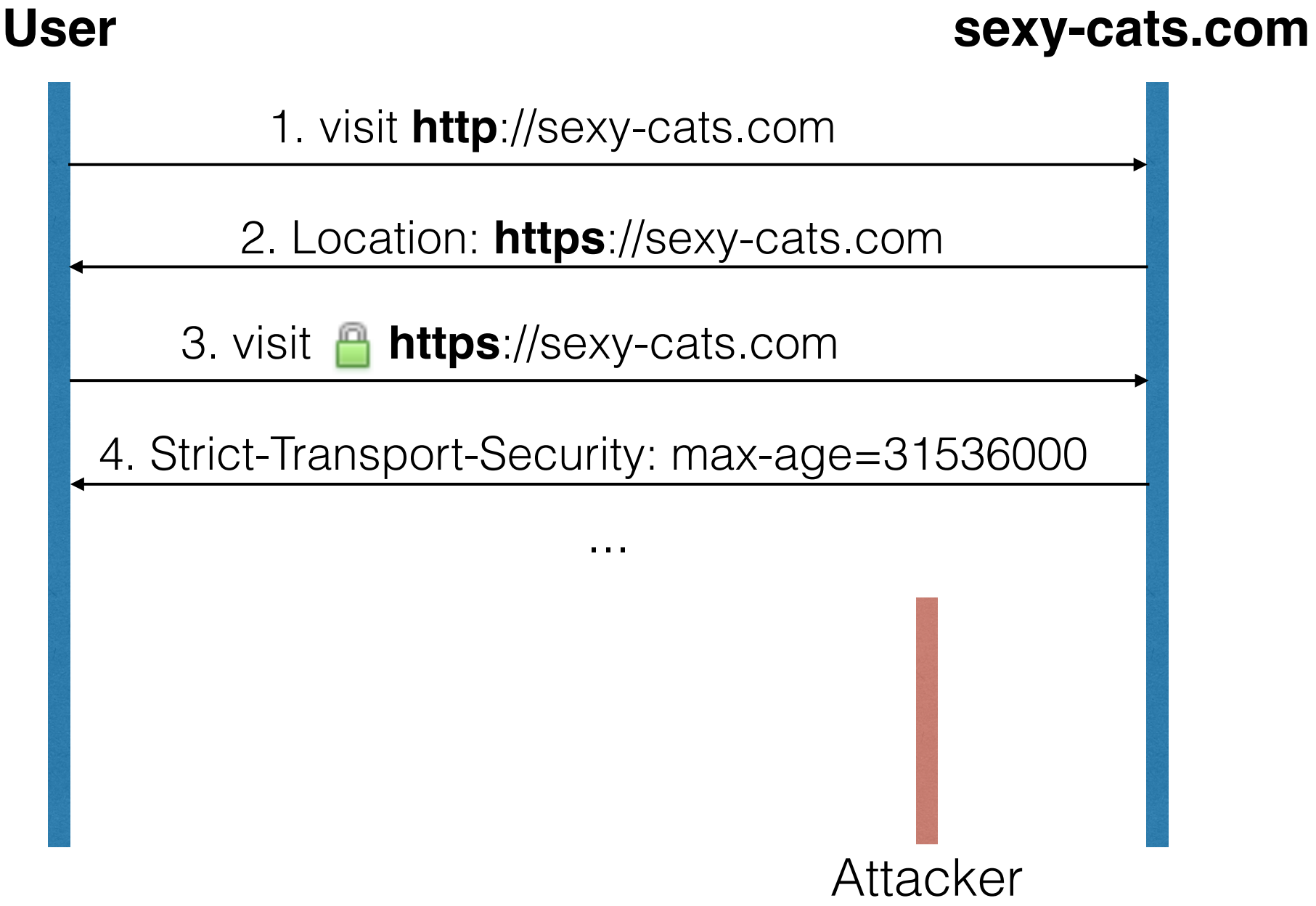

User
sexy-cats.com
1. visit **http**://sexy-cats.com

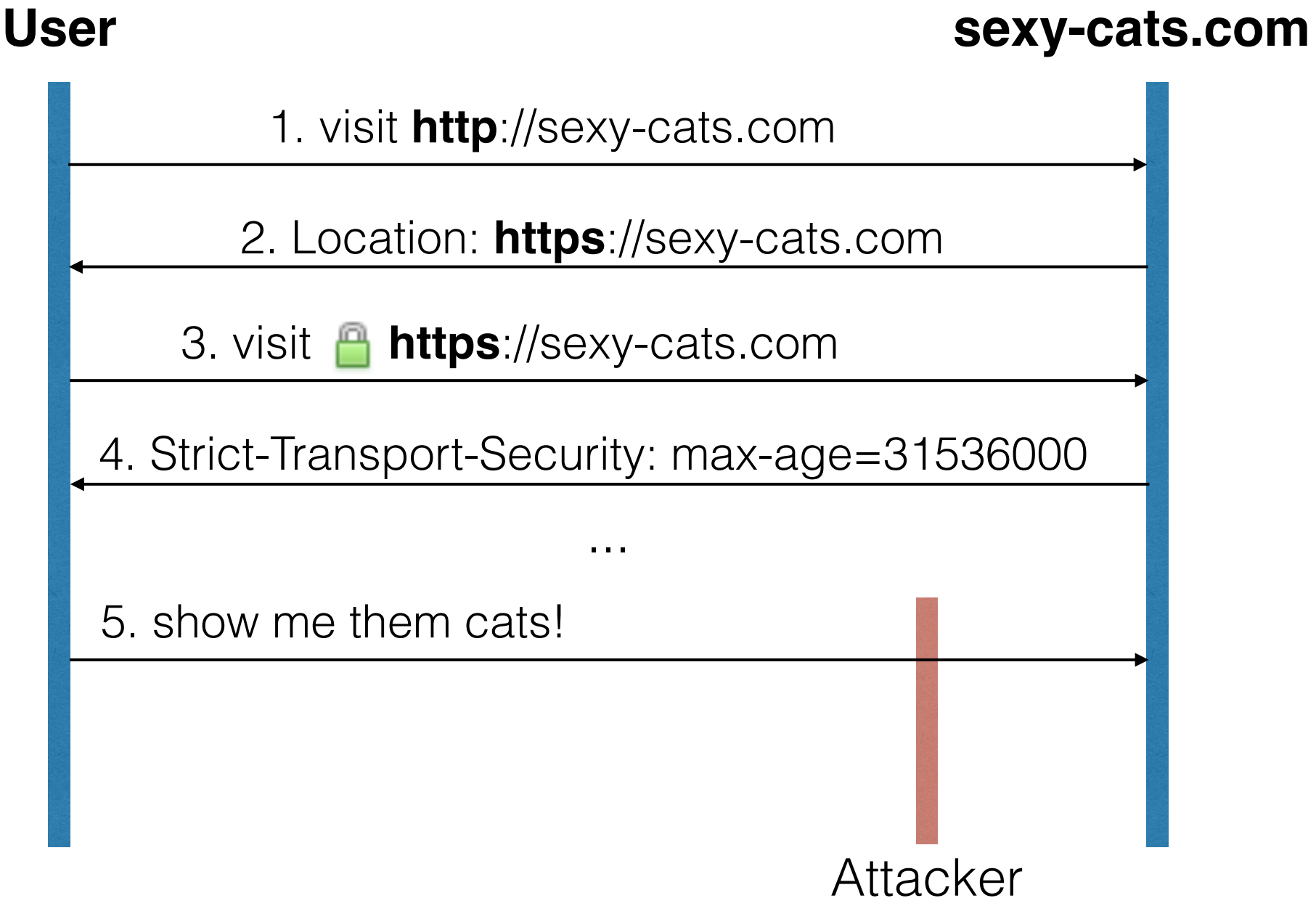

User
sexy-cats.com
1. visit **http**://sexy-cats.com
2. Location: **https**://sexy-cats.com

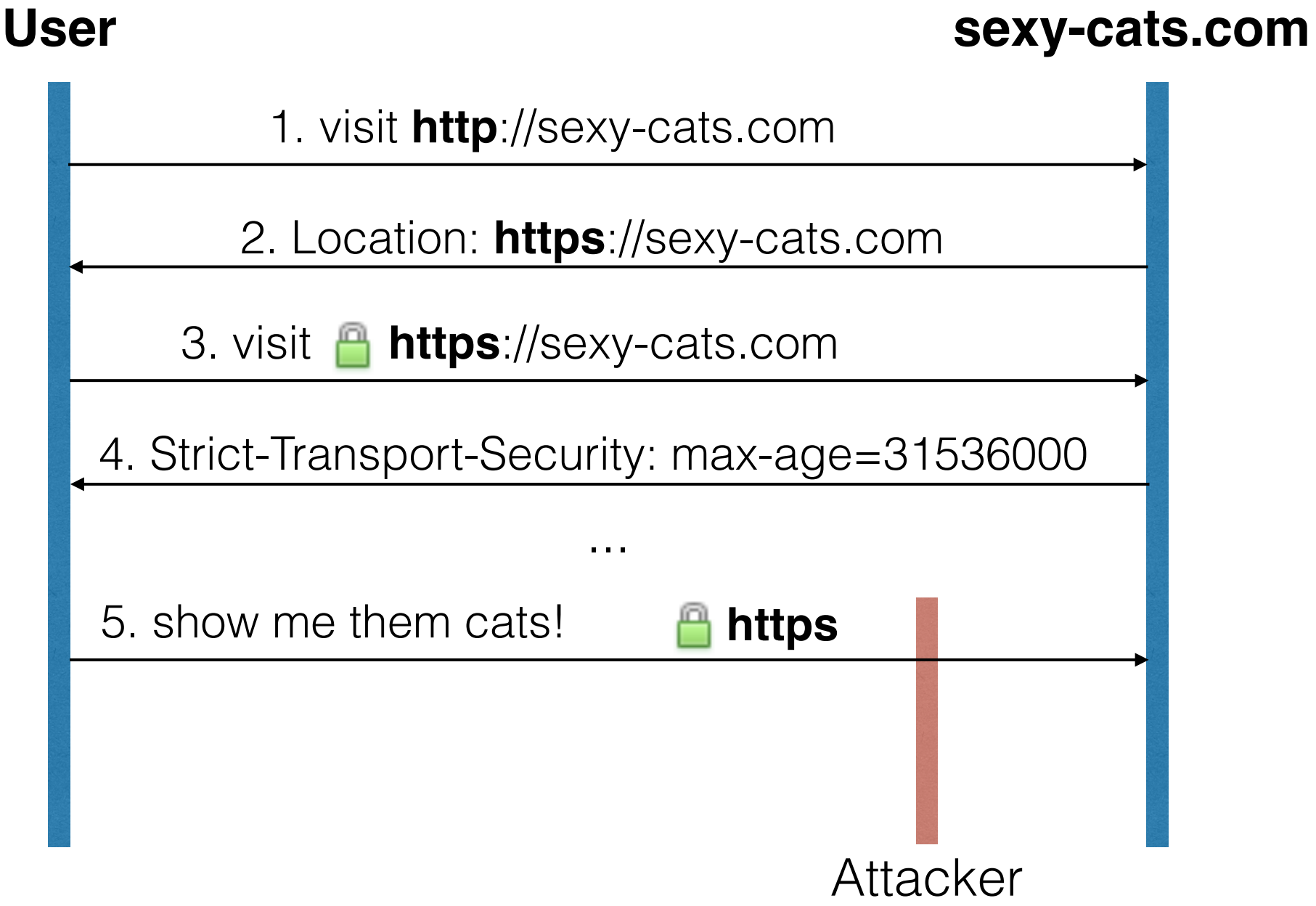

User
sexy-cats.com
1. visit **http**://sexy-cats.com
2. Location: **https**://sexy-cats.com
3. visit **https**://sexy-cats.com

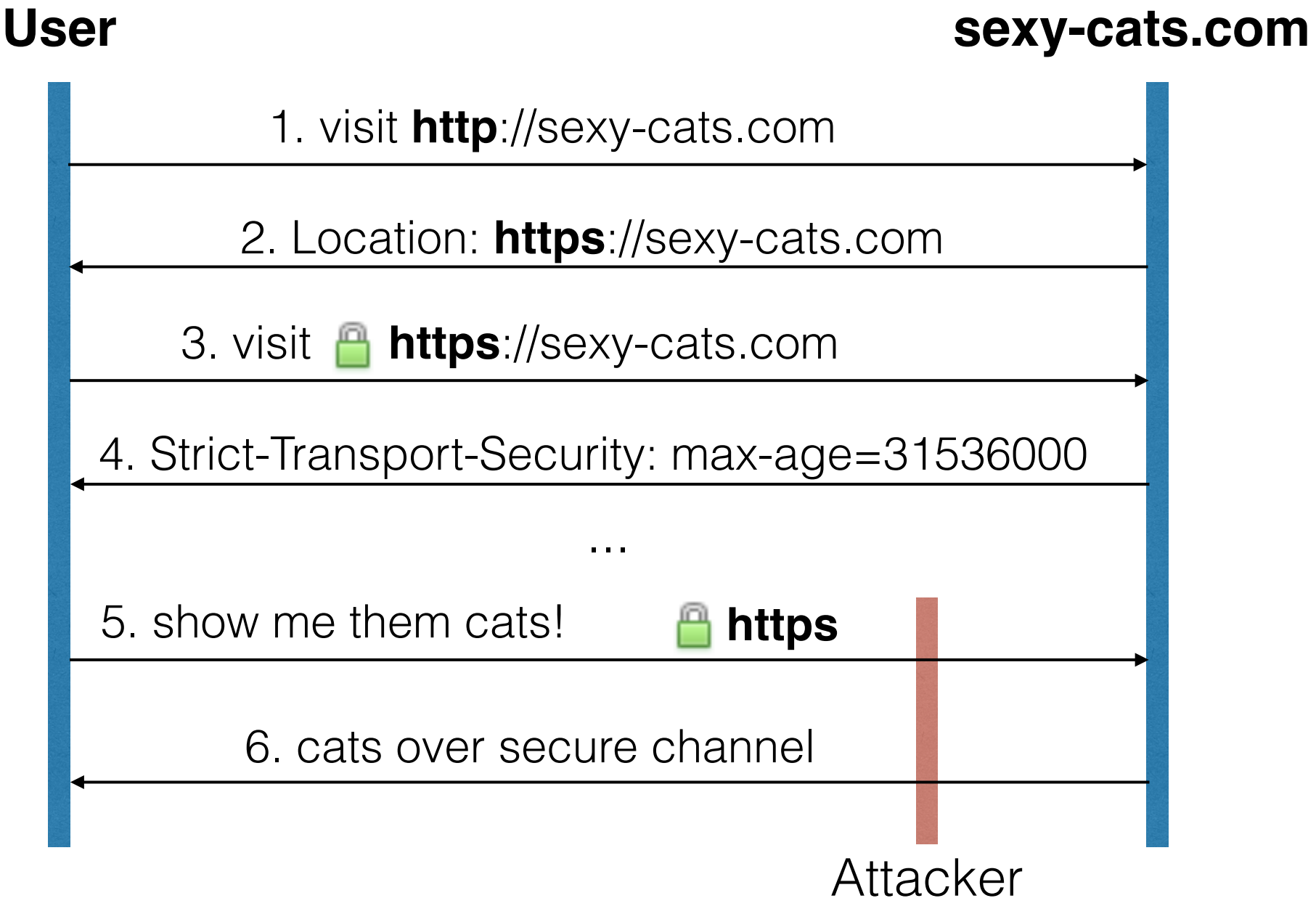

User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000


User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000
...


User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000
...
Attacker


User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000
...
5. show me them cats!
Attacker


User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000
...
5. show me them cats!
https
Attacker


User
sexy-cats.com
1. visit http://sexy-cats.com
2. Location: https://sexy-cats.com
3. visit https://sexy-cats.com
4. Strict-Transport-Security: max-age=31536000
...
5. show me them cats!
https
6. cats over secure channel
Attacker

# Defence mechanisms

## HTTP Strict Transport Security

- prevents SSL-stripping attacks in MitM scenario
- `Strict-Transport-Security: max-age=31536000`
- Needs to be sent over HTTPS
- All subsequent requests over HTTPS

# The Good

# Chapter 2: Security on the web

- **Security features**
  - Defence mechanisms
  - **Weaknesses & vulnerabilities**
- Findings

# Weaknesses & vulnerabilities

- 10 vulnerabilities and weaknesses

Mixed content inclusion
SSL-stripping
Insecure SSL implementation
→ Transport layer security

Vulnerable remote JS inclusion
Weak browser XSS protection
HTTP Parameter Pollution
→ Cross-Site Scripting

Outdated server software
Outdated CMS
Information leakage
Sensitive files
→ Miscellaneous

# Weaknesses & vulnerabilities

## Vulnerable remote JS inclusion

```
<script type="text/javascript"
src="http://register-me.com/
script.js"> </script>
```

read more:

"You Are What You Include" - Nikiforakis

# Weaknesses & vulnerabilities

## HTTP Parameter Pollution

Demo-time!

# The Bad

# Chapter 2: Security on the web

- Security features
  - Defence mechanisms
  - Weaknesses & vulnerabilities
- **Findings**

# Findings

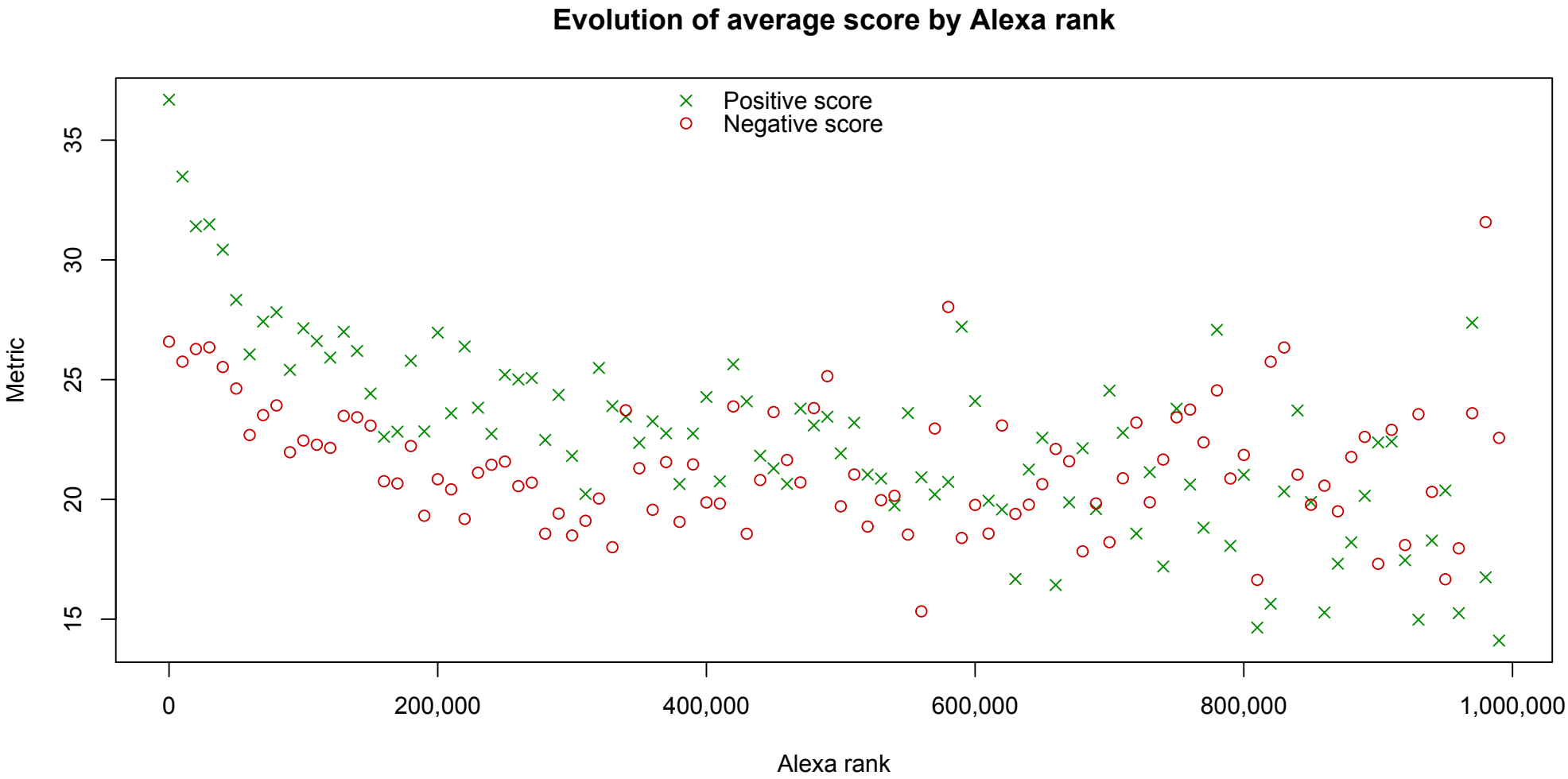
- General
  - 46.12% enabled at least 1 security mechanism
    - Most popular: *`HttpOnly`* (33.51%)
  - 56.39% contained at least 1 vulnerability/weakness
    - Most common: Outdated server software (28.06%)
    - For SSL websites: 80.32% had mixed content or bad SSL implementation
    - 15.24% of sites tested for HPP were vulnerable; 75% of those: vulnerable to XSS

# Findings

- Incorrect security-header usage
  - *`X-Frame-Options: SAME-ORIGIN`* (vs `SAMEORIGIN`)
  - *`Strict-Transport-Security: max-age=0`*
    - 12.93% of HSTS adopters failed to do it properly

# Findings

- Security by Alexa rank
  - high rank ~ high positive score
  - negative score unrelated to popularity


Evolution of average score by Alexa rank
Positive score
Negative score
Metric
35
30
25
20
15
0
200,000
400,000
600,000
800,000
1,000,000
Alexa rank

# Findings

- Novel findings
  - `Access-Control-Allow-Origin: domain.local`
  - Remote script inclusions from domains for sale
  - Remote script inclusions from stale Google Code projects
    - **3.8M requests from 1.1M unique IPs for 3,400 websites**

Google code

## Create a new project

**Instantly create your open source hosting project by filling out the form below. For your project, you'll receive:**

- Git, Mercurial, and Subversion code hosting
- Download/release hosting
- Integrated source code browsing and code review tools
- An issue tracker and project wiki

Learn more

**Note:** You can create at most 9 more projects.

**Project name**

Example: my-project-name

**Project summary**

**Description**

**Version control system**

- Git
- Mercurial
- Subversion

**Source code license**

Select a license...

**Project label(s)**

add another row

riNet KU LEUVEN


```
helpUser();
```

Delete project

Schedule this project for deletion. All project contents will be completely deleted in the near future. In the meantime, it will only be visible to project members.

**Scheduled for deletion by owner. Eligible for deletion after 29 days.**


```
compromiseUser();
```

# The Ugly

# Chapter 3: Lessons learned

- **Limitations & challenges**
- Conclusion

# Limitations & challenges

- Mainly passive analysis
  - Due to ethical considerations
  - Comparison of **known-vulnerable set** to **set of banking websites**
    - Avg. positive score: **35.21** vs. **41.62**
    - Avg. negative score: **27.33** vs. **12.80**

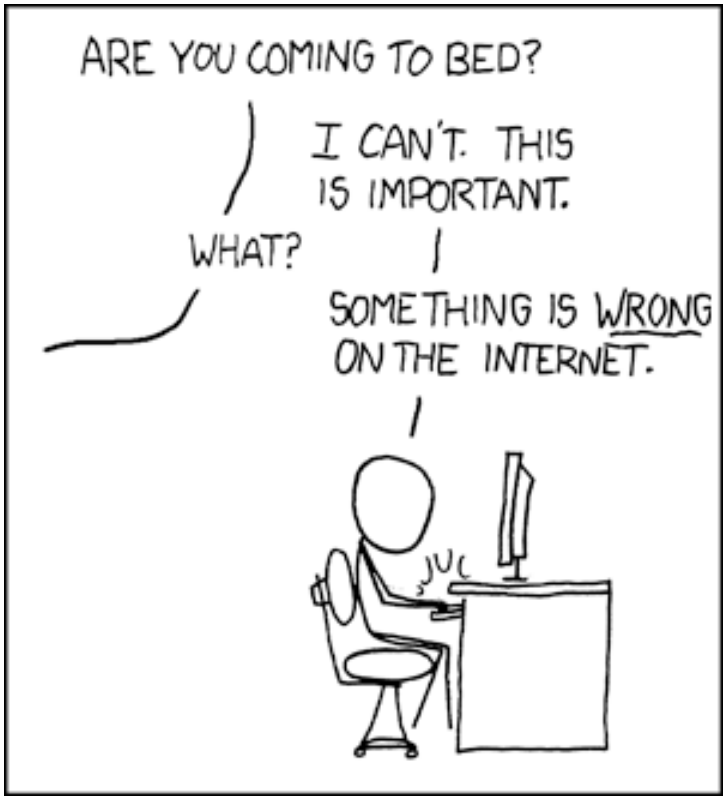
# Limitations & challenges

- Scoring system
  - Positive score on different scale than negative score
  - Limited number of features
    - ✦ Score should be seen as estimation
  - Score is subject to opinion
    - ✦ Score was calculated for a general website

# Conclusion

- Large-scale evaluation of 22,851 EU websites
- Security score based on 8 security mechanisms, 10 weaknesses
- Presence of security features related to Alexa rank
- Presence of weaknesses unrelated to Alexa rank
- Discovery of novel variants of JS inclusion attacks
- Security on the web often is ugly

# Questions?


ARE YOU COMING TO BED?
I CAN'T. THIS IS IMPORTANT.
WHAT?
SOMETHING IS WRONG ON THE INTERNET.