



SECURITY AND PRIVACY
RESEARCH LAB

Rewriting History: Changing the Archived Web From the Present

Ada Lerner*, Tadayoshi Kohno**, and Franziska Roesner**

* Wellesley College

** Paul G. Allen School of Computer Science and Engineering, University of Washington

Web Archives

1996

A screenshot of the ACM website from December 19, 1996. The page has a blue wavy background. At the top, there's a navigation bar with a search field, a 'Go' button, and links for 'NOV', 'DEC 19 JUN', '1995', '1996', '1998', 'About this capture', and social media icons for Facebook and Twitter. Below the navigation is a banner for 'The ACM Digital Library' with the text 'Over 4200 Citations are Available On-line'. To the left of the banner is a vertical sidebar with the text 'ACM's 100-Year Journey in Computing 1947 - 2047'. The main menu consists of several blue boxes: 'Publications', 'Special Interest Groups', 'U.S. Public Policy', 'Conferences and Events', 'Serving the community', 'Career Opportunities and Development', and 'Chapters and Local Activities'. At the bottom, there are links for '[Text-Only Home Page]', '[WWW Account]', '[WWW Awards]', '[Other WWW Resources]', and '[Feedback]'. A copyright notice at the very bottom reads 'Copyright (C) 1996, Association for Computing Machinery, Inc. All rights reserved.'

2017

A screenshot of the ACM website in 2017. The header features the ACM logo and the tagline 'Advancing Computing as a Science & Profession'. The navigation bar includes links for 'Digital Library', 'CACM', 'Queue', 'TechNews', 'Learning Center', 'Career Center', 'Join', 'Volunteer', 'myACM', and a search bar. Below the header, there's a large banner with the text 'Association for Computing Machinery' and 'Advancing Computing as a Science & Profession'. A quote below the banner says 'We see a world where computing helps solve tomorrow's problems – where we use our knowledge and skills to advance the profession and make a positive impact.' There are four portrait photos of people in white boxes. In the bottom right corner, there's a small video player showing a person speaking at a podium.

Web Archives in Journalism

INTERNET ARCHIVE
Wayback Machine

<https://2009-2017.state.gov/secretary/remarks/2017/01/266711.htm>

FEB MAR APR
◀ 11 ▶
2016 2017 2018 this

45 captures
21 Jan 2017 – 27 Apr 2017

The screenshot shows the official website of the U.S. Department of State. At the top, there is a navigation bar with links for Home, About, Services, News, Press, and Contact. Below this is a banner featuring the Department of State logo and the text "U.S. DEPARTMENT OF STATE" and "DIPLOMACY IN ACTION". To the right of the banner are social media icons for YouTube, Twitter, Facebook, and More. A search bar with the placeholder "Search Site..." and a magnifying glass icon is located at the top right. On the left side, there is a sidebar titled "In This Section:" with a link to "«Go Back". The main content area features a large title: "Apology for Past Discrimination toward Employees and Applicants based on Sexual Orientation". Above the title, a breadcrumb trail shows the page path: Home > Secretary of State John Kerry > Secretary Kerry's Remarks > 2017 Secretary Kerry's Remarks > Remarks by Secretary Kerry: January 2017.

In This Section:

[«Go Back](#)

Home > Secretary of State John Kerry > Secretary Kerry's Remarks > 2017 Secretary Kerry's Remarks > Remarks by Secretary Kerry: January 2017

Apology for Past Discrimination toward Employees and Applicants based on Sexual Orientation

Web Archives as Citations

- Often recommended and used for durable citations in scholarly work
- Wikipedia: **14%** of all **references in feature articles** are to the **Wayback Machine**



WIKIPEDIA
The Free Encyclopedia

Web Archives in Legal Proceedings

- Used as evidence in many legal situations, e.g.,
 - Criminal
 - Civil
 - Administrative
 - Patent applications



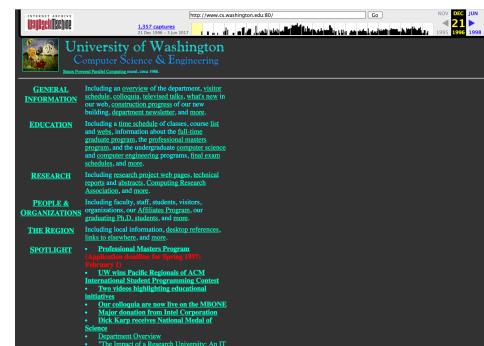
People may have motivations to manipulate web archives



How might they do so?

The Wayback Machine

- Web archive since 1996
- ~15 petabytes of website snapshots
- Accessible to anyone via web browser @ web.archive.org
- Scripts, HTTP headers, CSS, images, etc.
- The model for a large number of other web archives



Rewriting History in Web Archives

Insight: Web archives are used for important things,
so people might want to manipulate them

Vulnerabilities: Multiple vulnerabilities enable manipulation

Attacks: We develop the theory and practice
of exploiting those vulnerabilities

Measurement: These vulnerabilities are very prevalent

Defenses: Exploring the space of defenses that
publishers, clients, and archives could deploy



REUTERS

EDITION: U.S. ▾

Search News & Quotes

[Home](#) [Business](#) [Markets](#) [World](#) [Politics](#) [Technology](#) [Opinion](#) [Money](#) [Life & Culture](#) [Pictures](#) [Video](#)

Recession fears recede as factories show growth

By Lucia Mutikani

WASHINGTON - Manufacturing unexpectedly grew in August and fewer Americans filed new claims for jobless aid last week, despite a slump in confidence that threatened to push the economy back into recession. [Full Article](#) | [Video](#)

White House sharply cuts growth forecast

Retailers' August sales just miss Wall Street view

Online job postings hint at more U.S. hiring ahead

Exclusive: Regulators seek trading secrets

7:24pm EDT

WASHINGTON/NEW YORK (Reuters) - U.S. securities regulators have taken the unprecedented step of asking high-frequency trading firms to hand over the details of their trading strategies, and in some cases, their secret computer codes.

August auto sales defy consumer caution

4:47pm EDT

Fed orders Goldman to review foreclosures

4:56pm EDT

Starz to pull content from Netflix as talks fail

7:20pm EDT

Einhorn strikes out with baseball's New York Mets

4:42pm EDT

Appeals court asked to vacate "Don't Ask" judgment

5:04pm EDT

OPINION



CCS 2017 will be held in Dallas, USA!

By Cliff Young and Chris Jackson

The conventional wisdom now is that the president is beatable in 2012. Taken as a whole, Obama is still a favorite, but the numbers show he is dangerously close to turning from a clear favorite to a struggling contender. [Commentary](#)

Easterbrook: The election's first bogeyman

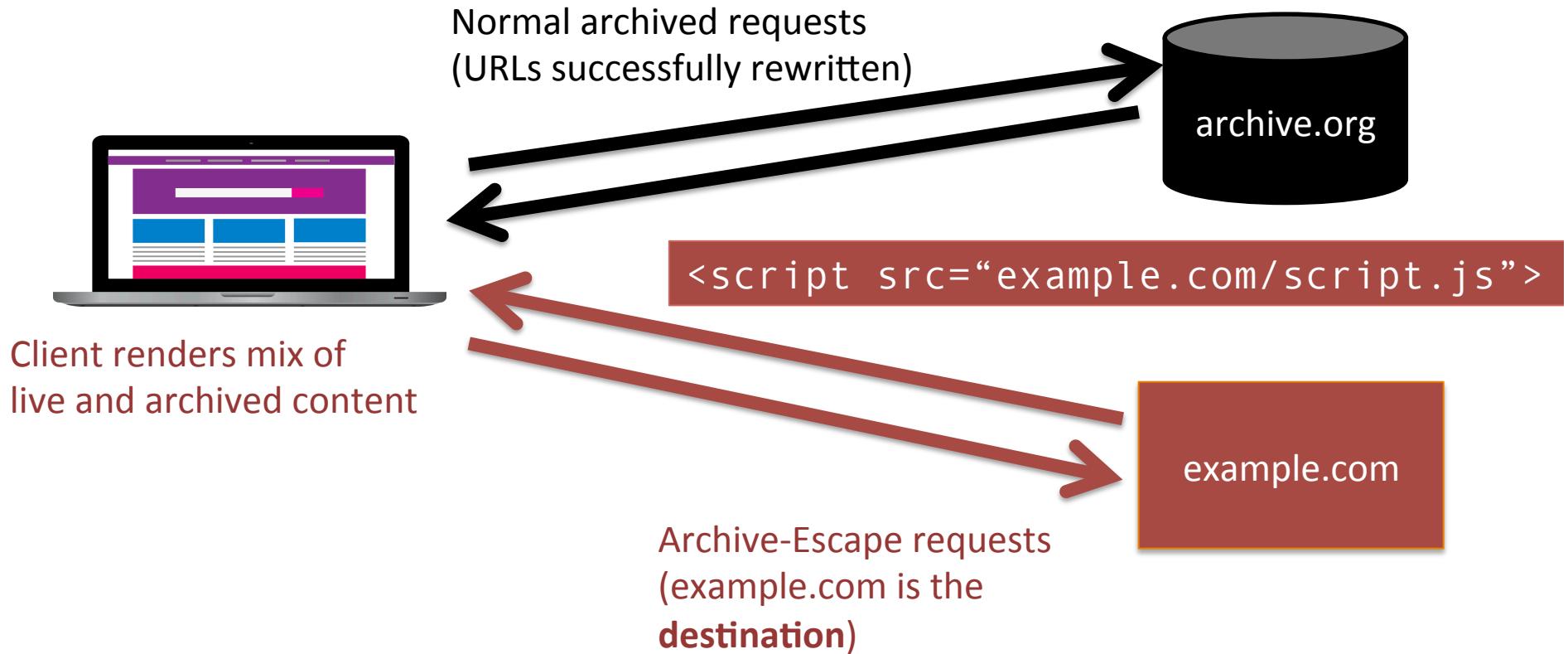
Chen: China is still waiting for inflation to peak

VIDEOS



Archive Escapes

```
<script src="web.archive.org/2011/example.com/script.js">
```



Rewriting History



What are the technical ways to manipulate client views of the Wayback Machine?

itors seek

RK (Reuters) - have taken the high-hand over the legies, and in computer

lef

OPINION



Obama at the electoral tipping point

itors seek

RK (Reuters) - have taken the high-hand over the legies, and in computer

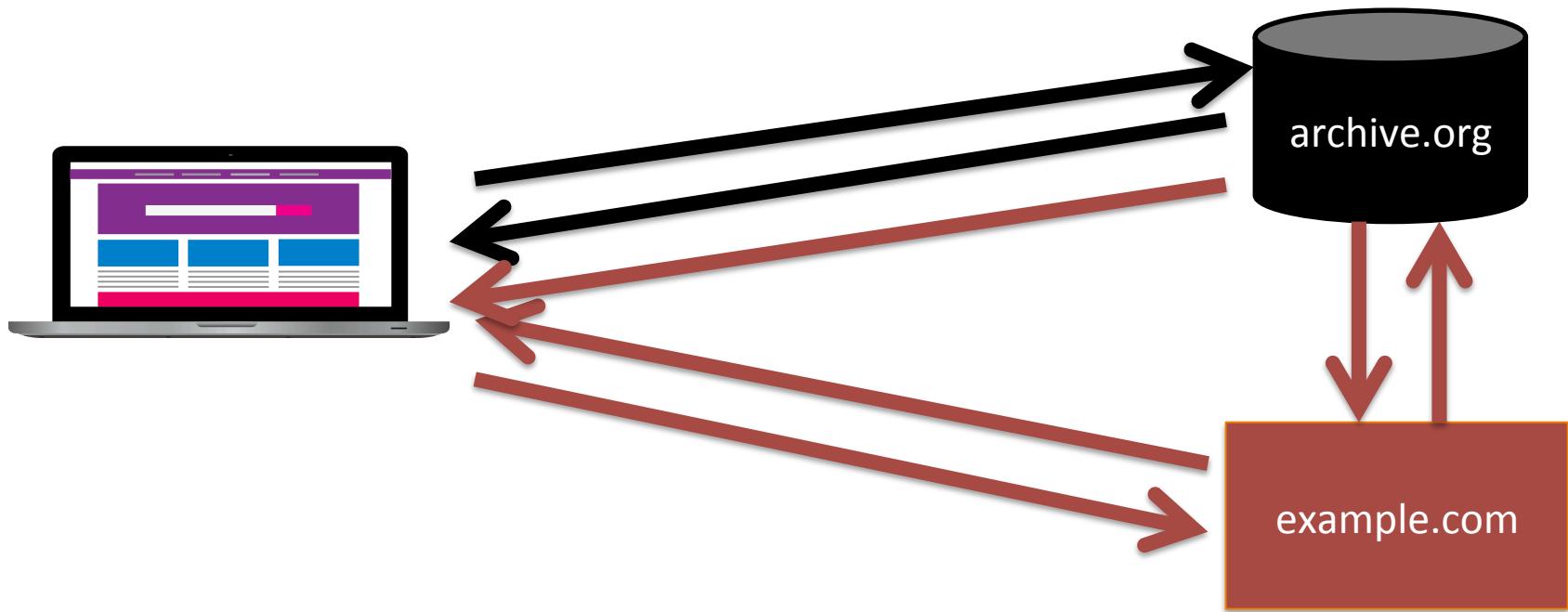
lef

OPINION



CCS 2017 will be held in Dallas, USA!

Anachronism Injection



https://web.archive.org/web/20020802031508/http://www.cs.washington.edu/

Internet Archive Wayback Machine

http://www.cs.washington.edu/ Go JUN AUG SEP 2001 2002 2003

1,239 captures 21 Dec 96 – 18 Jul 16

CS&E logo University of Washington Department of Computer Science & Engineering

University of Washington College of Engineering College of Arts and Sciences About Us Search Contact Info

Topics:

- News
- Events and Talks
- Education
- Research
- People
- Diversity
- Organizations
- Computing Facilities
- Outreach
- Faculty/Staff Recruiting
- New CSE Building

Information for:

- Perspective Students
- Current Students
- Alumni
- Faculty/Staff
- Community Colleges
- Industry
- Public/Press
- Visitors

UW: Red Square near dusk

The University of Washington Department of Computer Science & Engineering includes roughly thirty five faculty members, forty staff members, two hundred seventy five graduate majors (150 in the Doctoral program and 125 in the Professional Masters Program), and four hundred twenty five undergraduate majors. Ranked among the top ten programs in the nation, we are active in most of the principal areas of the field, and are engaged in a broad range of interdisciplinary initiatives. We are located in [Beautiful Sieg Hall](#) at the center of the UW campus in Seattle, Washington -- increasingly prominent as a national and international technology center in software, biotech, the health sciences, and other fields.

See information on the [Paul G. Allen Center for Computer Science & Engineering](#) -- a new home for CSE, and the University of Washington's top institutional priority.

Computational Research Associate

Washington

CSE Web Privacy

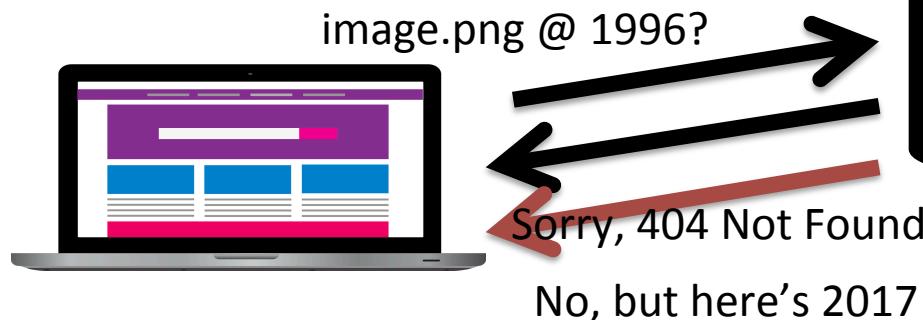
Spotlight News



Department of Computer Science & Engineering
University of Washington
Box 352350
Seattle, WA 98195-2350
(206) 543-1695 voice, (206) 543-2969 FAX
[comments to webmaster@cs.washington.edu]

<https://web.archive.org/web/20171102/http://example.com/image.png>

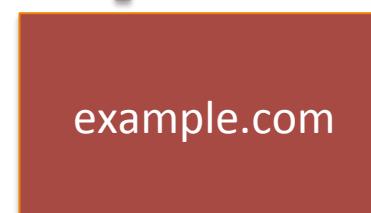
2. Client requests the vulnerable resource. Nearest-neighbor timestamp matching redirects to the anachronistic, malicious content



0. Precondition:
Vulnerable snapshot contains
reference to never-archived resource

Anachronism Injection

1. Attacker uses "Save Page Now" to store malicious content as the **only** capture of the missing resource



<https://web.archive.org/web/19960101/http://example.com/image.png>

Prevalence of "Strong" Vulnerabilities

1. No foresight required
(archive-escape or anachronism-injection)



Prevalence of "Strong" Vulnerabilities

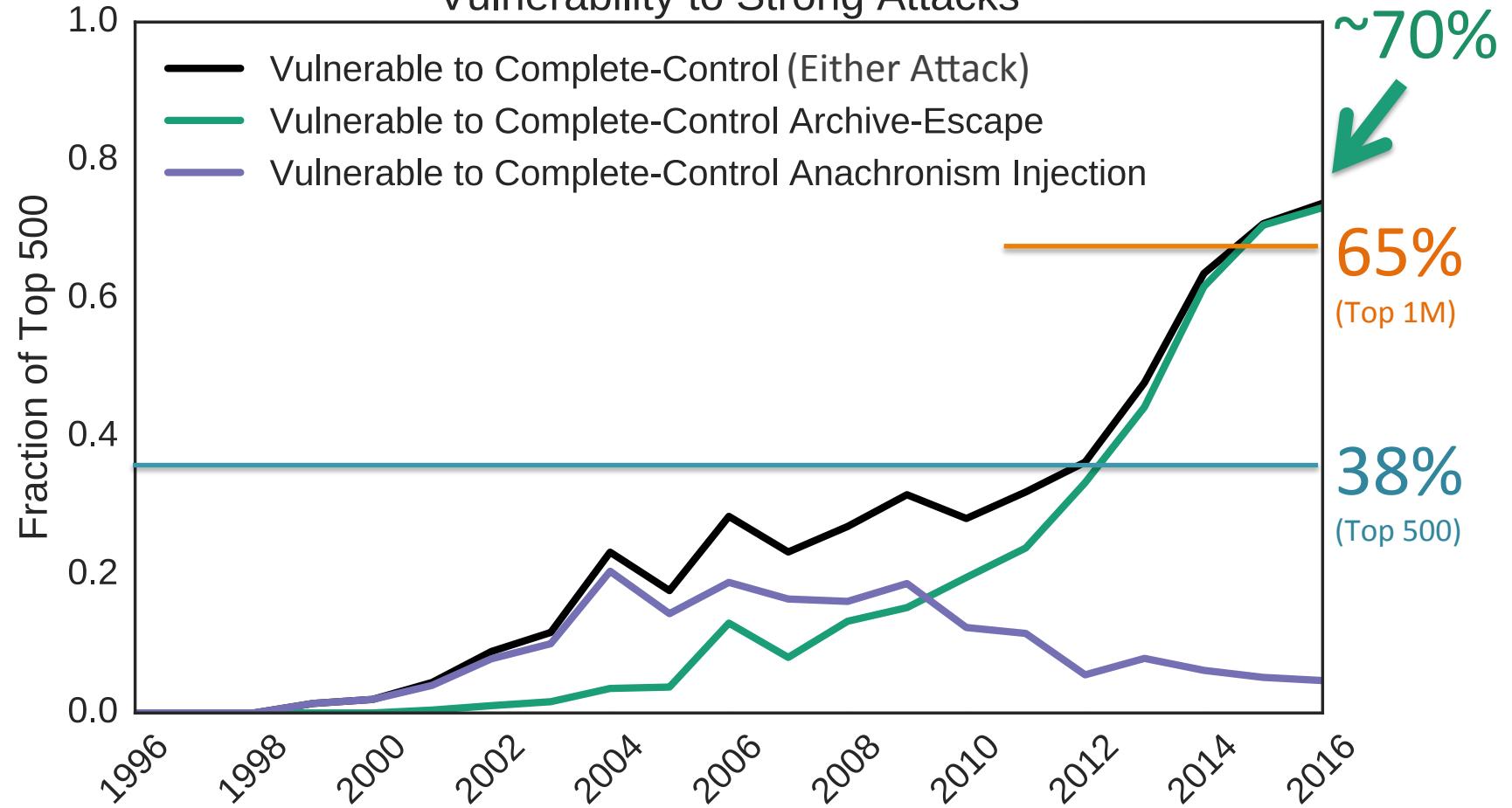
1. No foresight required
(archive-escape or anachronism-injection)
2. Gains **complete control** of snapshot
(vulnerable resource is ***script*** or ***stylesheet***)

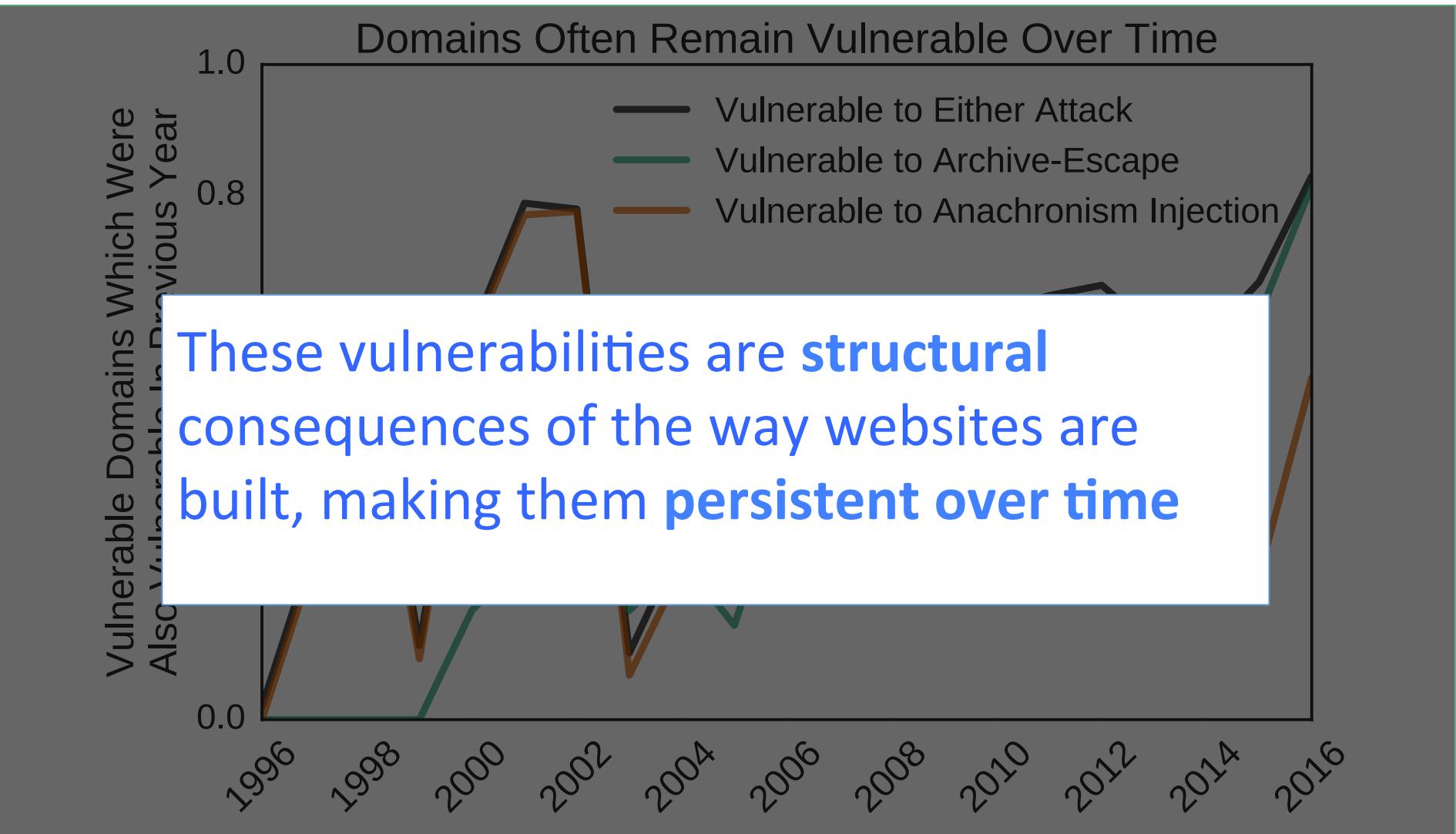




Vulnerability to Strong Attacks

(Recent Top 500 Snapshots)





Same-Origin Escapes

theonion.com



AMERICA'S **WORST** NEWS SOURCE



AMERICA'S FINEST NEWS SOURCE

evil.com

Same-Origin Escapes

Live Web

- Same-Origin Policy prevents modification because 1st- and 3rd-parties are different origins



AMERICA'S FINEST NEWS SOURCE

theonion.com/evil.com

Archived Web

- Same-Origin Policy **doesn't apply** because 1st- and 3rd- parties are the **same, archival origin**



AMERICA'S **WORST** NEWS SOURCE

archive.org/archive.org

Defenses

- Deployed by Archives:
 - Modify/Sandbox JS
 - Serve Archived Domains From Distinct Subdomains
 - Content Security Policy Headers
- Deployed by Publishers
 - Opt-Out of Archives
 - Avoid Dynamic URLs
 - Actively Archive Subresources
- Deployed by Clients:
 - Escape & Anachronism Blocking/Detecting Browser Extension



AMERICA'S WORST NEWS SOURCE

Defending Against Same Origin Escapes

<https://web.archive.org/web/20000101000000/http://first.com>

<https://web.archive.org/web/20000101000000/http://third.com>



Defending Against Same Origin Escapes

`https://web.archive.org/web/20000101000000/http://first.com`

`https://web.archive.org/web/20000101000000/http://third.com`



`https://first.com.web.archive.org/web/20000101000000/http://first.com`

`https://third.com.web.archive.org/web/20000101000000/http://third.com`

Defending Against Archive-Escapes

All archive-escapes are third-party requests!

Content Security Policy (CSP) Header

```
Content-Security-Policy:  
default-src 'self'  
    'unsafe-eval'  
    'unsafe-inline'  
    data: archive.org web.archive.org analytics.archive.org
```

Defending Against Anachronism Injection

Surface information about subresource timestamps in order to allow experts to detect anachronisms both unintentional and malicious

Defending Against Anachronism Injection

INTERNET ARCHIVE
WayBack Machine

<http://www.cnn.com:80/> JUL AUG SEP
163,388 captures MORE 31 2006 2007
20 Jun 2000 - 25 Oct 2017 
COLLECTED BY

Collection: [Institut national de l'audiovisuel](#)
Crawl data from [Institut national de l'audiovisuel](#) in France. This data is currently not publicly accessible. powered by **YAHOO! SEARCH**

ina.fr INSTITUT NATIONAL DE L'AUDIOVISUEL

Business Sports Analysis Politics Law Tech Science Health Entertainment Offbeat Travel Education Specials Autos Excha
from Wikipedia:
The Institut national de l'audiovisuel (or INA, French for National Audiovisual Institute), is a repository

Make CNN Your Home

TIMESTAMPS

<http://i.cdn.turner.com/cnn/.element/img/2.0/misc/expired/t1land.image.jpg> +5 years 4 months
<http://www.cnn.com/.element/ssi/www/sect/1.3/misc/contextual/MAIN.html?domId=mainCLLinkSpots> +4 years 7 months
<http://www.cnn.com/.element/ssi/sect/1.3/MAIN/staticWeatherBox.html?domId=weatherBox> +4 years 7 months
<http://www.cnn.com/.element/ssi/www/sect/1.3/misc/contextual/MAIN-EMPTY.html?domId=mainCLSponsoredLinks> +4 years 7 months
http://www.cnn.com:80/.element/ssi/www/breaking_news/1.5/banner.exclude.html?domId=cnnBreakingNewsBanner +2 months 28 days
<http://ar.atwola.com/file/adsEnd.js> -6 hours 39 minutes
<http://ar.atwola.com/file/adsWrapper.js> -6 hours 39 minutes
<http://i.a cnn.net/cnn/.element/img/1.5/ceiling/cnnCeildivider.gif> -5 hours 26 minutes with killing
<http://i.a cnn.net/cnn/.element/img/1.3/misc/icon.offsite.gif> -5 hours 26 minutes
<http://i.a cnn.net/cnn/images/1.gif> -5 hours 26 minutes
<http://i.a cnn.net/cnn/.element/img/1.3/misc/icon.offsite.gif> -5 hours 26 minutes TSB: Wrong runway wasn't only error | Video
http://i.a cnn.net/cnn/.element/img/1.5/ceiling/search/hdr_the_web.gif -5 hours 26 minutes for pestling
http://i.a cnn.net/cnn/.element/img/1.5/main/icon_video.gif -5 hours 26 minutes
http://i.a cnn.net/cnn/.element/img/1.5/ceiling/search/hdr_cnn_com.gif -5 hours 26 minutes Video
http://i.a cnn.net/cnn/.element/img/1.5/ceiling/search/hdr_yahoo.gif -5 hours 26 minutes
http://i.a cnn.net/cnn/.element/img/1.5/ceiling/search/hdr_search.gif -5 hours 26 minutes peak for them
<http://i.a cnn.net/cnn/1.gif> -5 hours 25 minutes Study: Average commute a tad shorter | Help
<http://i.a cnn.net/cnn/.element/img/1.5/main/sponsors.rt.corner.gif> -5 hours 25 minutes flight!

Rewriting History in Web Archives

Insight: Socially important uses lead to incentives for manipulation.

Vulnerabilities & Attacks:

Archive-escapes, Anachronism Injection, and Same-Origin Escapes

Measurement: Strong vulnerabilities are very prevalent:

70% of snapshots captured today, 38% since 1996

Defenses: Defenses can help block vulnerabilities, prevent attacks, and allow expert users to trust appropriately in this important resource.

Special thanks: Mark Graham & the Wayback Machine team
at Internet Archive

<https://rewritinghistory.cs.washington.edu>