Cached and Confused

Web Cache Deception in the Wild

Seyed Ali Mirheidari
Sajjad "JJ" Arshad

H@CKTIVITYCON | hackerone

CACHED & CONFUSED:

WEB CACHE DECEPTION IN THE WILD

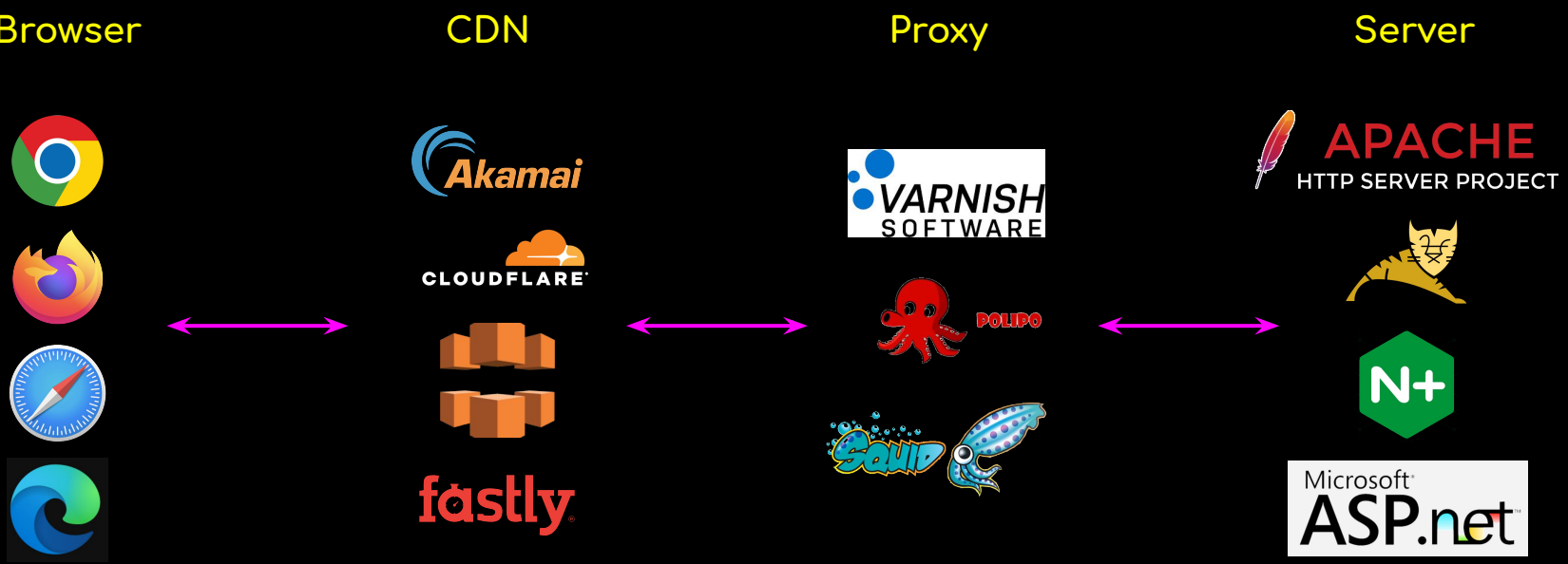
Seyed Ali Mirheidari and Sajjad "JJ" Arshad
Security Researchers

@sajjadium

Agenda

- ➔ Background
- ➔ Path Confusion
- ➔ Web Cache Deception (WCD)
- ➔ Advanced Exploitation Techniques
- ➔ Notable Observations
- ➔ Lesson Learned

Web Cache Technologies



Web Cache Behavior

Cache-Control Response Header

Cache-Control: no-store => Response should not be stored

Cache headers CAN be ignored by CDNs

Caching based on ***resource paths*** and ***extensions*** (e.g., jpg, css, js)

URL 101

scheme://user:password@host:port/path?query#fragment

path => file like structure separated by / (e.g., account/index.php)

query => list of key/value separated by & (e.g., p1=v1&p2=v2)

fragment => arbitrary string (e.g., comment)

Traditional vs Clean URL

Traditional URL

https://example.com/account/index.php?p1=v1

Clean URL

https://example.com/account/v1

Path Confusion 101

https://example.com/account/id

Browser, CDN, and Proxy think account is a directory and id is a file on the server's filesystem

BUT

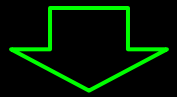
Server **knows** /account is referencing index.php file in the account directory and id is a parameter to index.php

Path Confusion 201

Semantic Disconnect among different framework-independent web technologies (e.g., browser, CDN, proxy, server) which results in different URL *path interpretations*

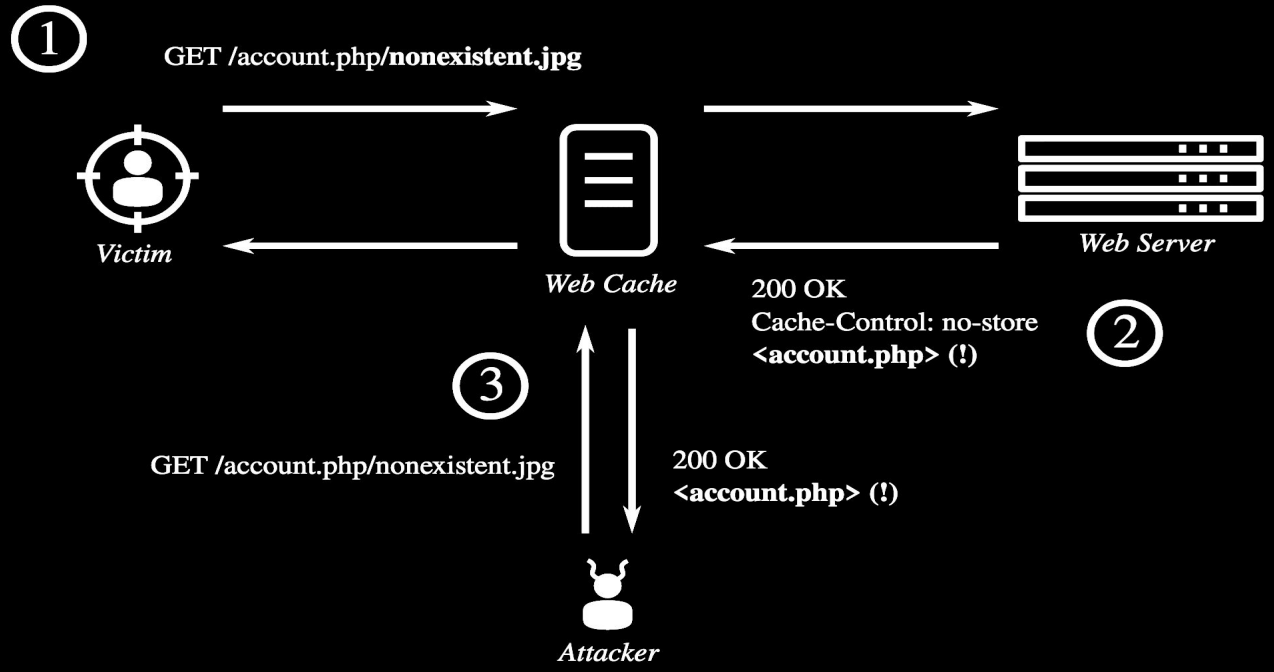
Basic Path Confusion (with Path Parameter)

example.com/account.php



example.com/account.php/nonexistent.jpg

Basic Web Cache Deception



Advanced Path Confusion

The Era of Encoding

URL Encoding

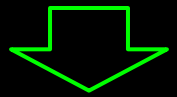
A technique to embed special/non-ascii characters in URL

\n => %0A ; => %3B # => %23 ? => %3F

Browsers & Proxies & CDNs & Servers can get *confused* with customized encoding

Path Confusion with Encoded ?

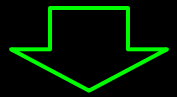
example.com/account.php?name=val



example.com/account.php%3Fname=valnonexistent.css

Path Confusion with Encoded \n

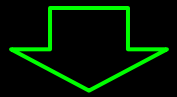
example.com/account.php



example.com/account.php%0Anonexistent.css

Path Confusion with Encoded ;

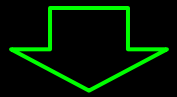
example.com/account.jsp;param



example.com/account.jsp%3Bparamnonexistent.css

Path Confusion with Encoded

example.com/account.php#frag



example.com/account.php%23nonexistent.css

Effectiveness of Encoding

Out of 37 vulnerable sites, 25 were exploited by **basic** WCD

11 sites were exploited ONLY by advanced WCD techniques

Increased number of vulnerable sites by ~ 45%

Interestingly, there were sites ONLY exploited with ONE technique

URL encoding is quite effective to confuse CDNs and Proxies

Path confusion elicit significantly more 200 OK server response !

Notable Observations

Practical Attack Scenarios

- PII (User, Name, Email, Phone, etc.)
- CSRF Tokens
- OAuth State parameter
- Session ID/BREAR Token/API Key
- Dynamic JavaScript file name (XSSI)

CSRF Protection Bypass
OAuth Redirect URI CSRF
Session Hijacking
Cross-Site Script Inclusion

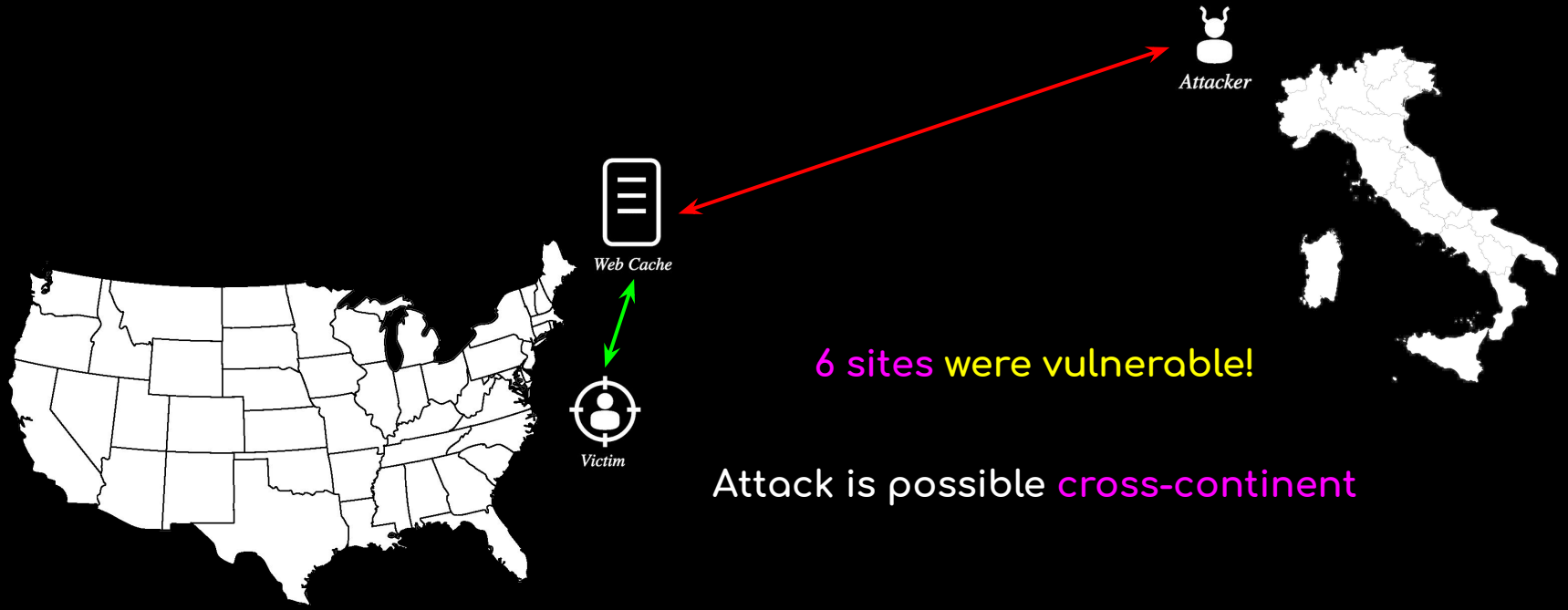
Authenticated vs. Unauthenticated Attacker

WCD DOES NOT require attackers to be authenticated

Increase likelihood of WCD attack

Websites with no public registration can be targeted

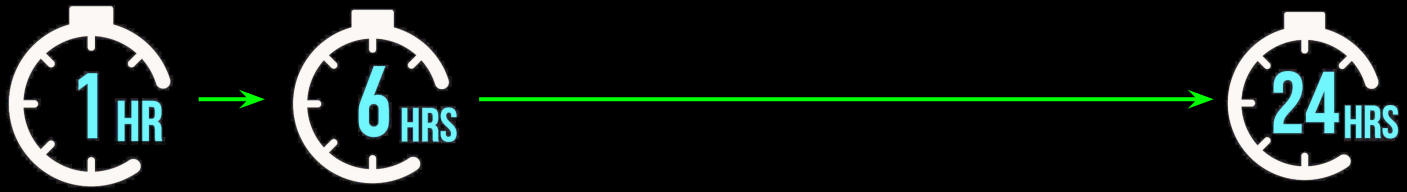
Cache Location



6 sites were vulnerable!

Attack is possible cross-continent

Cache Expiration



16 10 9

No. of vulnerable sites

Attack is still possible after several hours

Cache Configuration

CDN	Default Cached Objects	Honor Headers?		
		no-store	no-cache	private
Akamai	Objects with a predefined list of static file extensions only.	No	No	No
Cloudflare	Objects with a predefined list of static file extensions, AND all objects with cache control headers public or max-age > 0.	Yes	Yes	Yes
CloudFront	All objects.	Yes	Yes	Yes
Fastly	All objects.	No	No	Yes

Lesson Learned

- Configuring web caches correctly is **not trivial**
- CDNs are not plug & play solutions
- There is a widespread lack of user awareness about WCD
- WCD is generally a "system safety" problem
  - No isolated faulty components
  - Complex interactions among web technologies must take into consideration
- Variations of path confusion techniques make it possible to exploit sites that are otherwise not impacted by original attack
  - Some sites were only exploitable with one specific encoding attack
- Path confusion can be used in other attack vectors
  - Relative Path Overwrite (RPO), Cache Poisoning, CPDoS, ...

Thanks! Questions?

Seyed Ali Mirheidari

Sajjad "JJ" Arshad

@sajjadium sajjadium