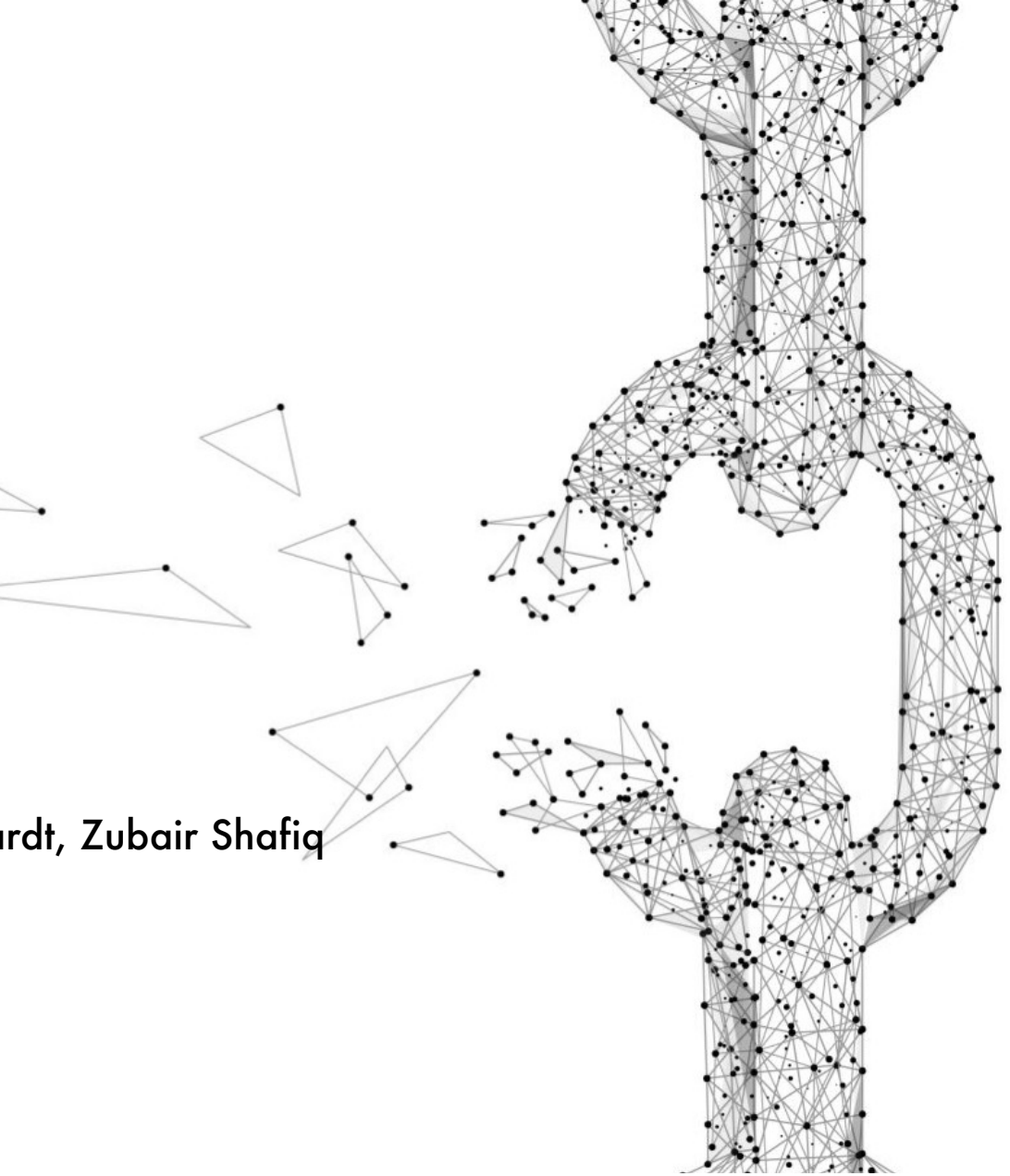
Khaleesi:
Breaker of Advertising & Tracking Request Chains



Umar Iqbal, Charlie Wolfe, Charles Nguyen, Steven Englehardt, Zubair Shafiq

USENIX Security Symposium, 2022

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

IOWA Department of Computer Science

UCDAVIS COMPUTER SCIENCE

DuckDuckGo

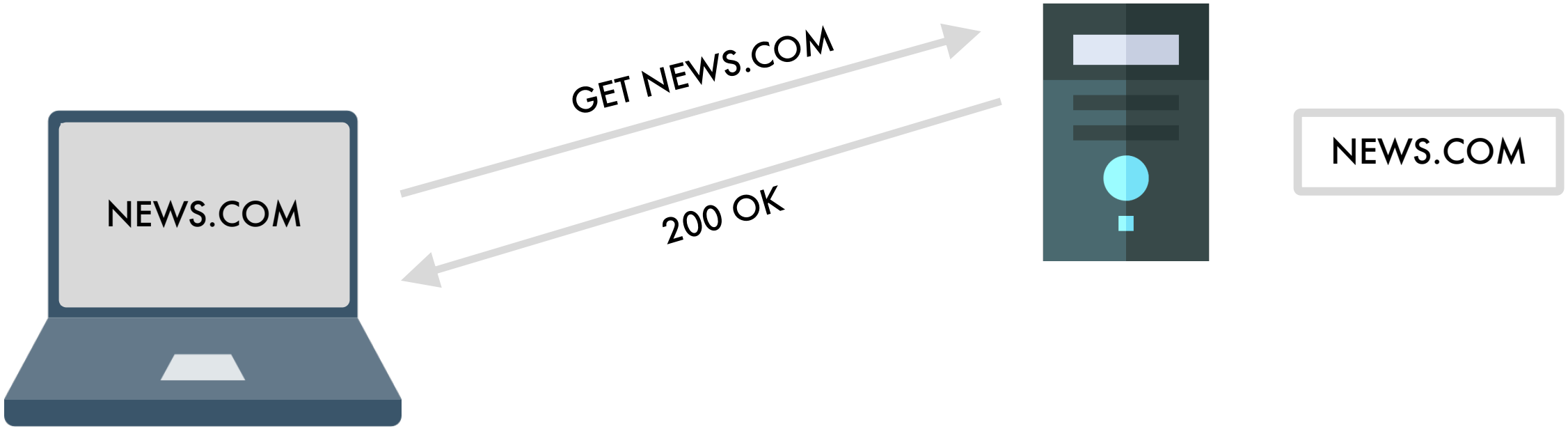
Talk overview

AdTech relies on request chains for tracking & bypassing privacy protections

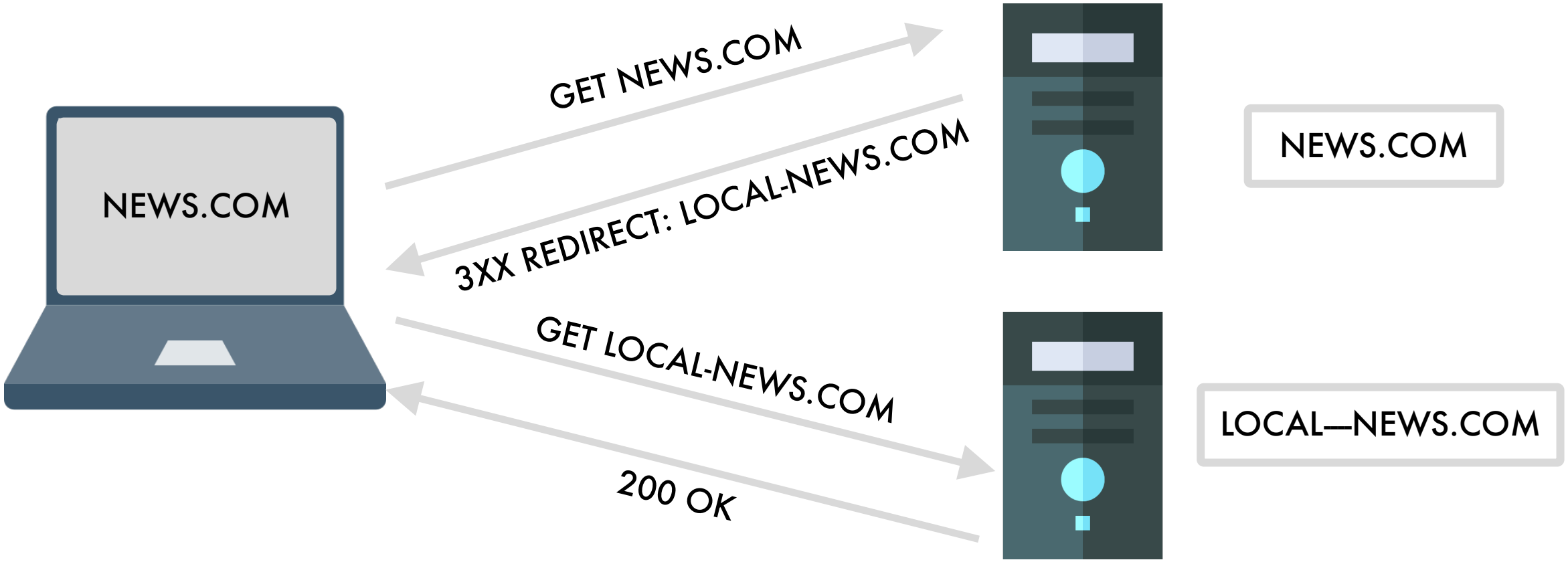
Existing solutions are ineffective against advertising & tracking request chains

Khaleesi:
A purpose-built approach to protect against advertising & tracking request chains

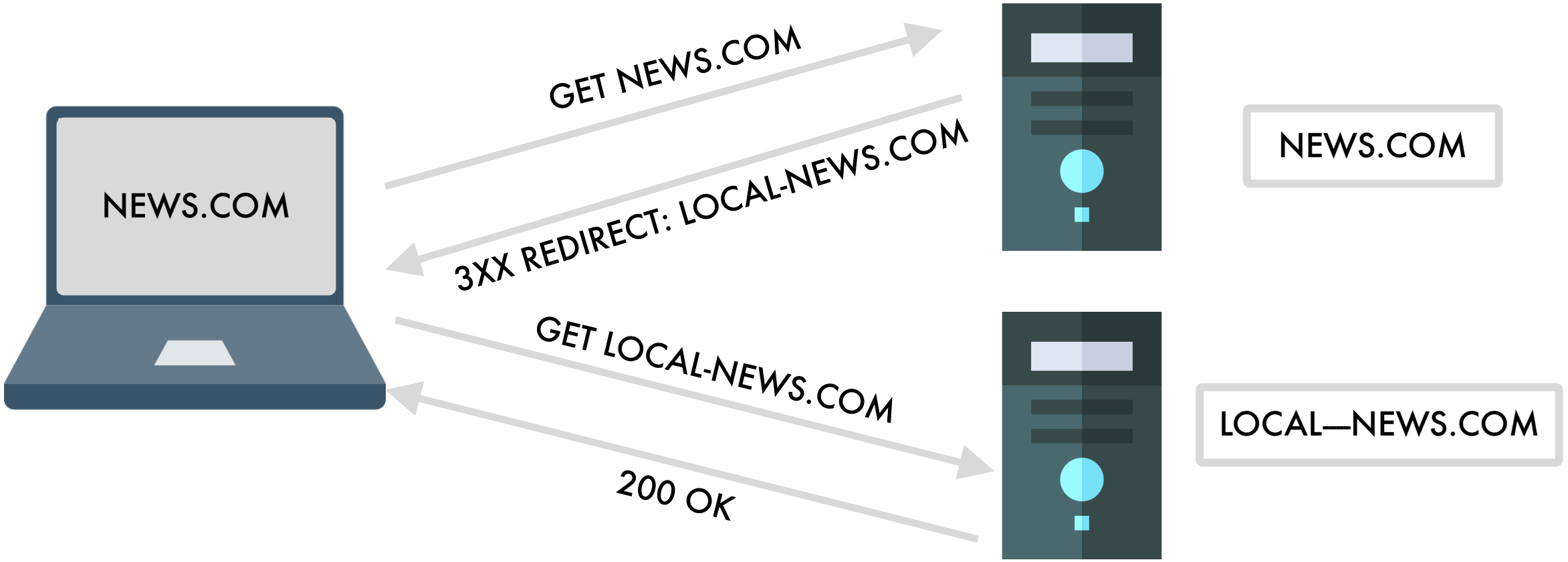
What are request chains?



What are request chains?

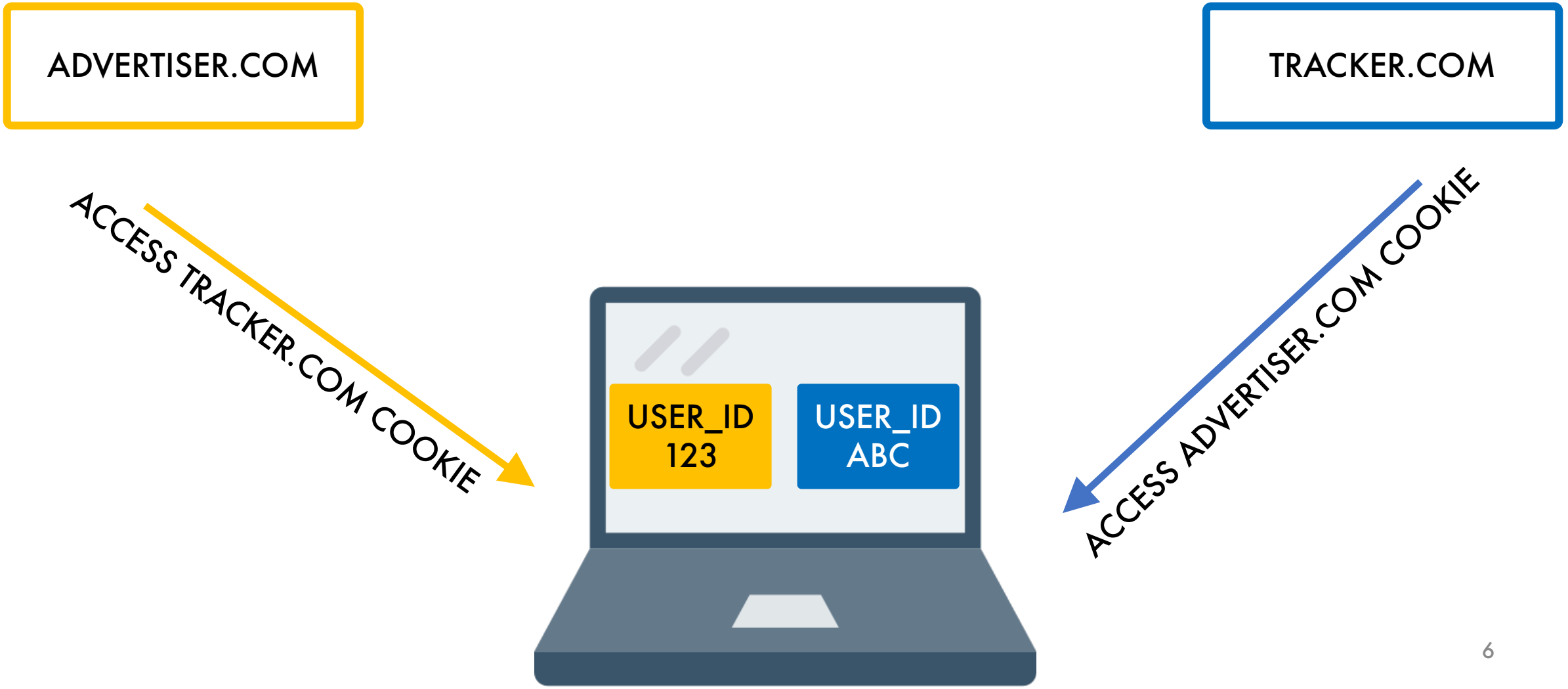


What are request chains?

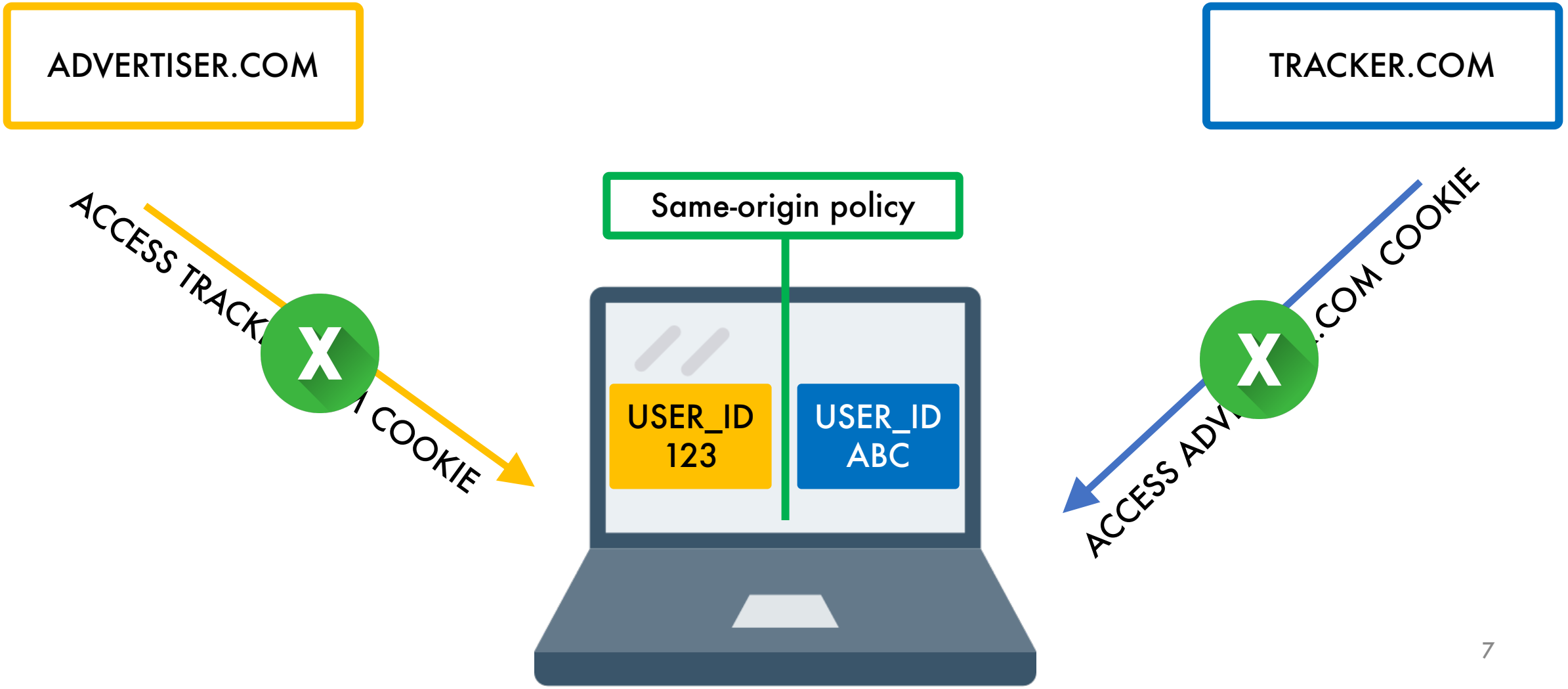


REQUEST → RESPONSE → REQUEST . . . . RESPONSE

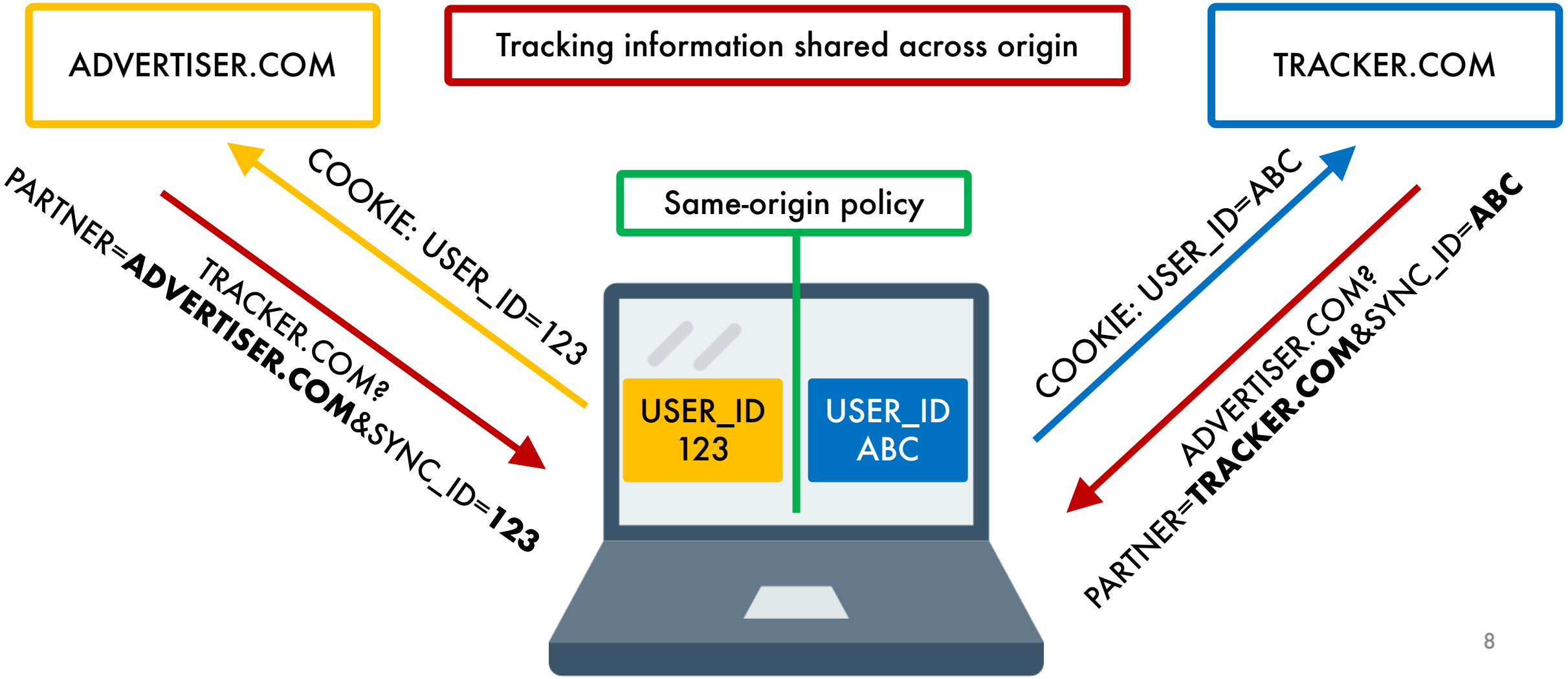
Background: Cookie syncing



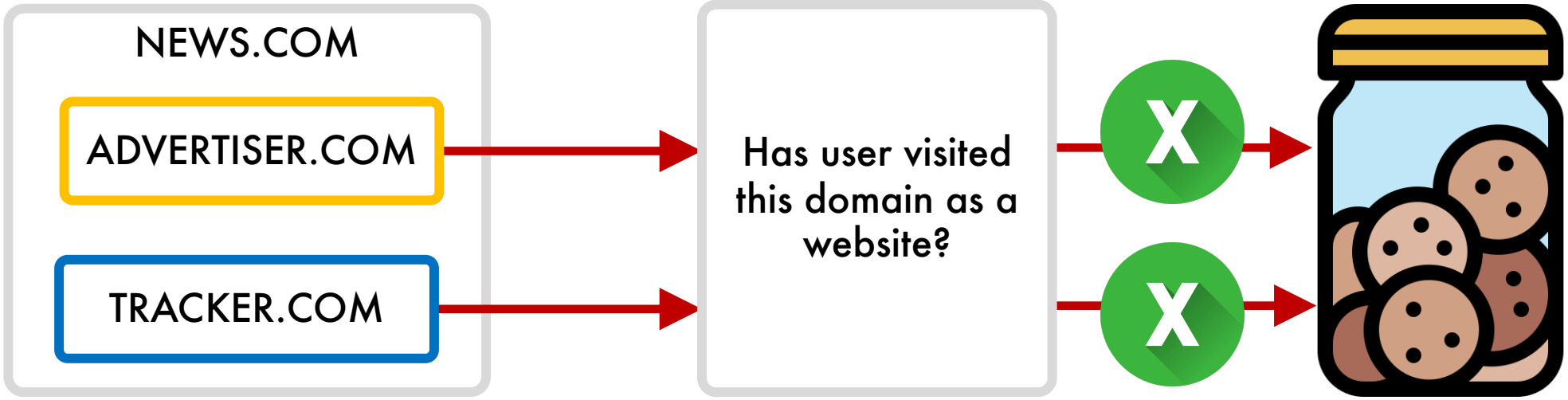
Background: Cookie syncing



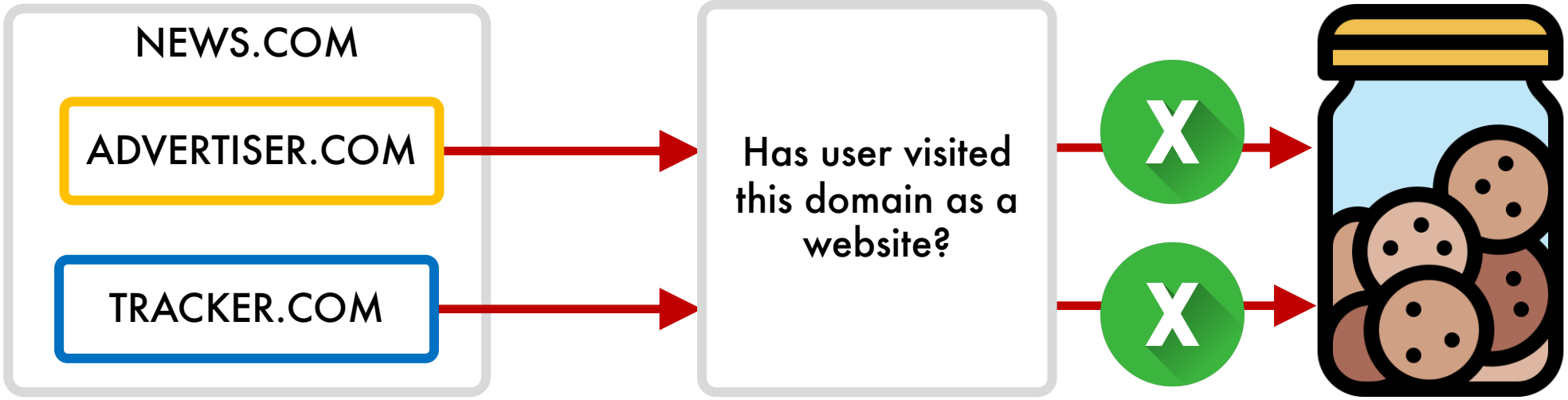
Background: Cookie syncing

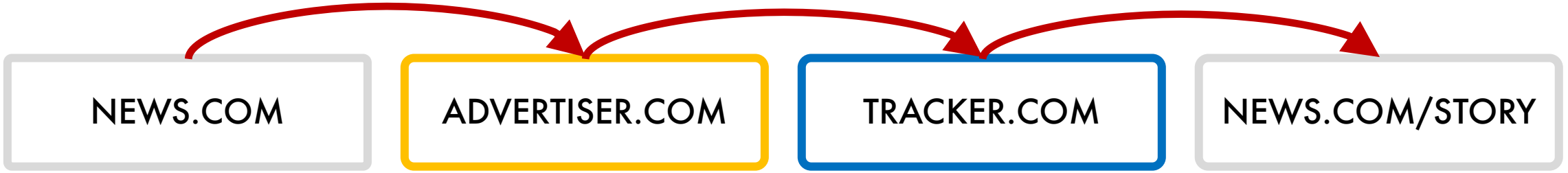


Background: Bounce tracking

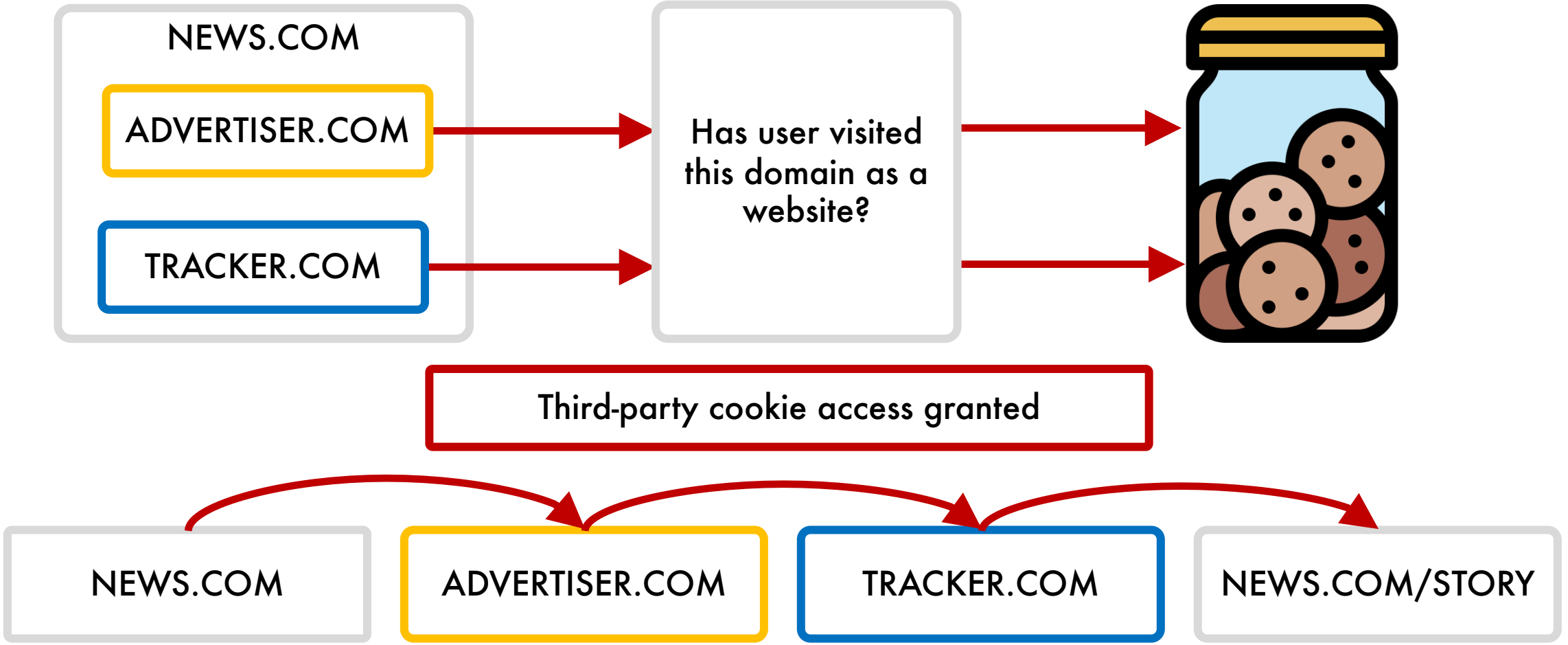


Background: Bounce tracking





Background: Bounce tracking



Talk overview

AdTech relies on request chains for tracking & bypassing privacy protections

Existing solutions are ineffective against advertising & tracking request chains

Khaleesi:
A purpose-built approach to protect against advertising & tracking request chains

Current solutions: Ad/Tracker blocking extensions

Widely used solution

Not equipped to detect request chains : (

Current solutions: Ad/Tracker blocking extensions

Widely used solution

Not equipped to detect request chains : (

Operate at the level of individual requests

Detect known “trackers” through manually curated filter lists

```
||TRACKER.COM
||EXAMPLE.COM
```

Current solutions: Ad/Tracker blocking extensions

Widely used solution

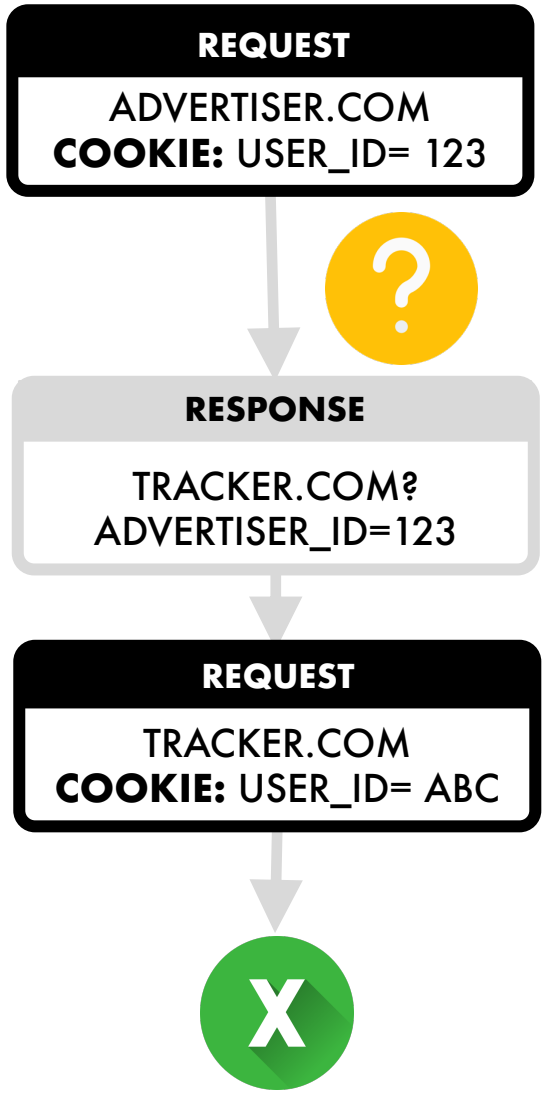
Not equipped to detect request chains : (

Operate at the level of individual requests

Detect known “trackers” through manually curated filter lists

```
||TRACKER.COM
||EXAMPLE.COM
```

Cannot detect “tracking”



Current solutions: Heuristic based detection

Detect “known” tracking

```
[Cookie syncing]
If requests share cookies
[BLOCK]
```

Current solutions: Heuristic based detection

Detect “known” tracking

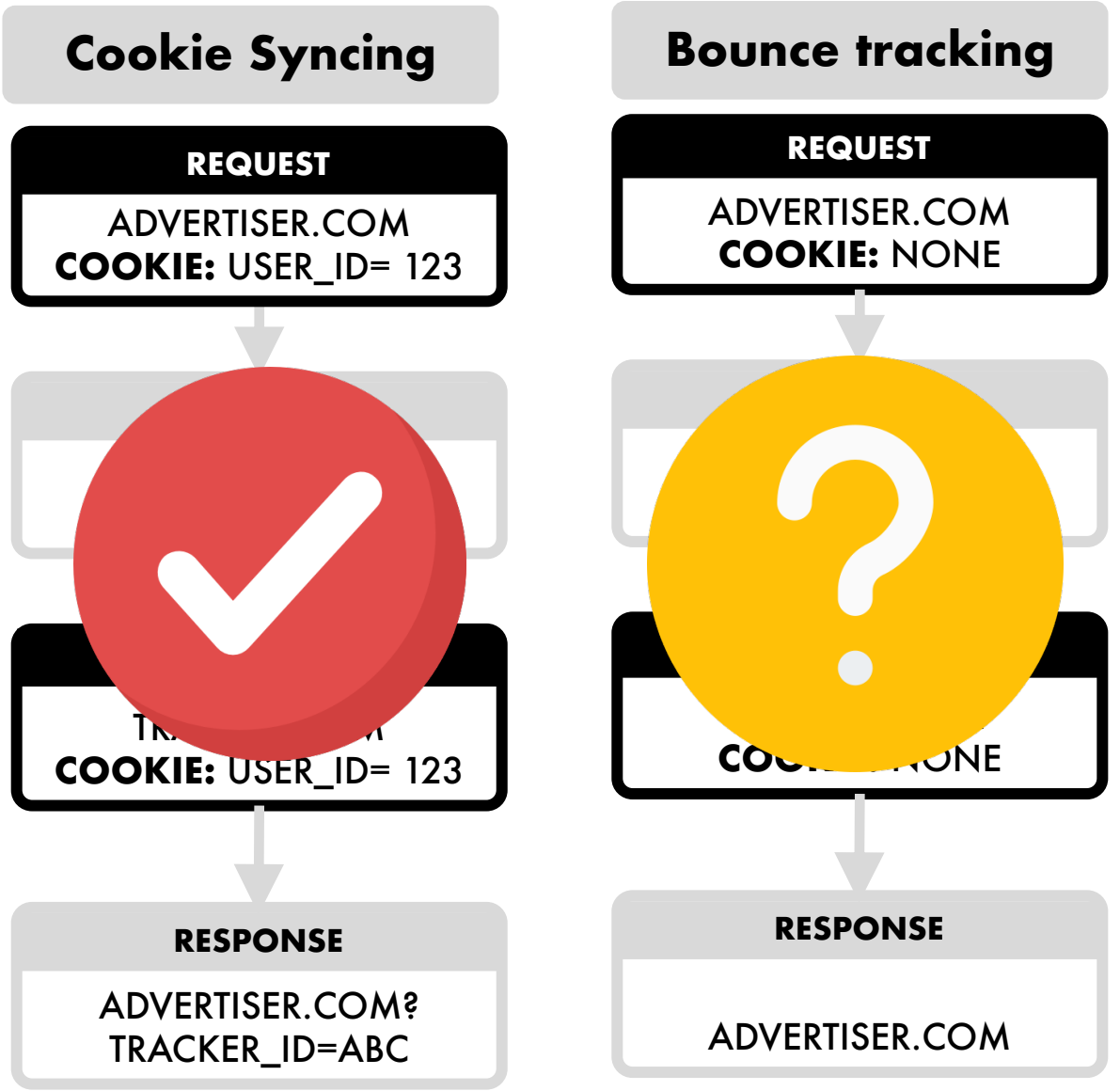
```
[Cookie syncing]
If requests share cookies
[BLOCK]
```

Current solutions: Heuristic based detection

Detect “known” tracking

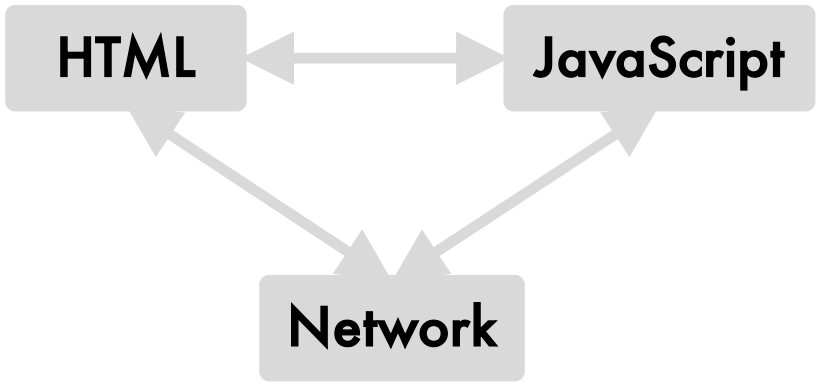
```
[Cookie syncing]
If requests share cookies
[BLOCK]
```

Cannot detect ”unknown” tracking



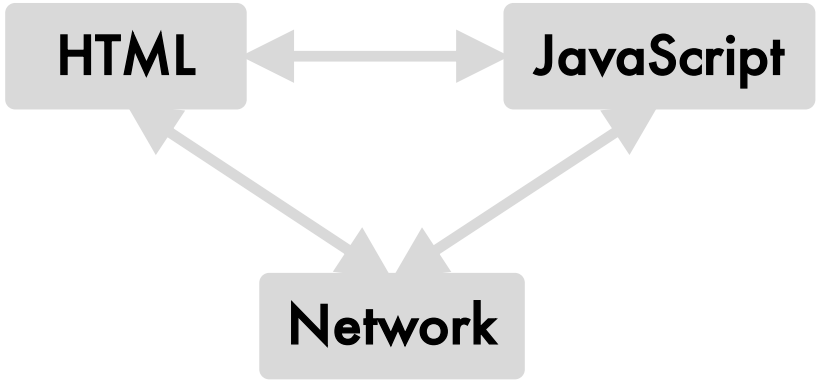
Current solutions: Machine learning based detection

Capture wide range of tracking behaviors



Current solutions: Machine learning based detection

Capture wide range of tracking behaviors

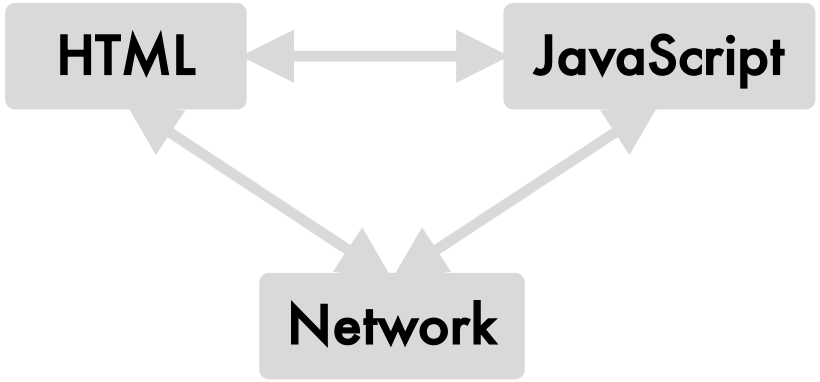


Not **specialized** to detect request chains

Do not **utilize** and **capitalize** on the the sequential nature

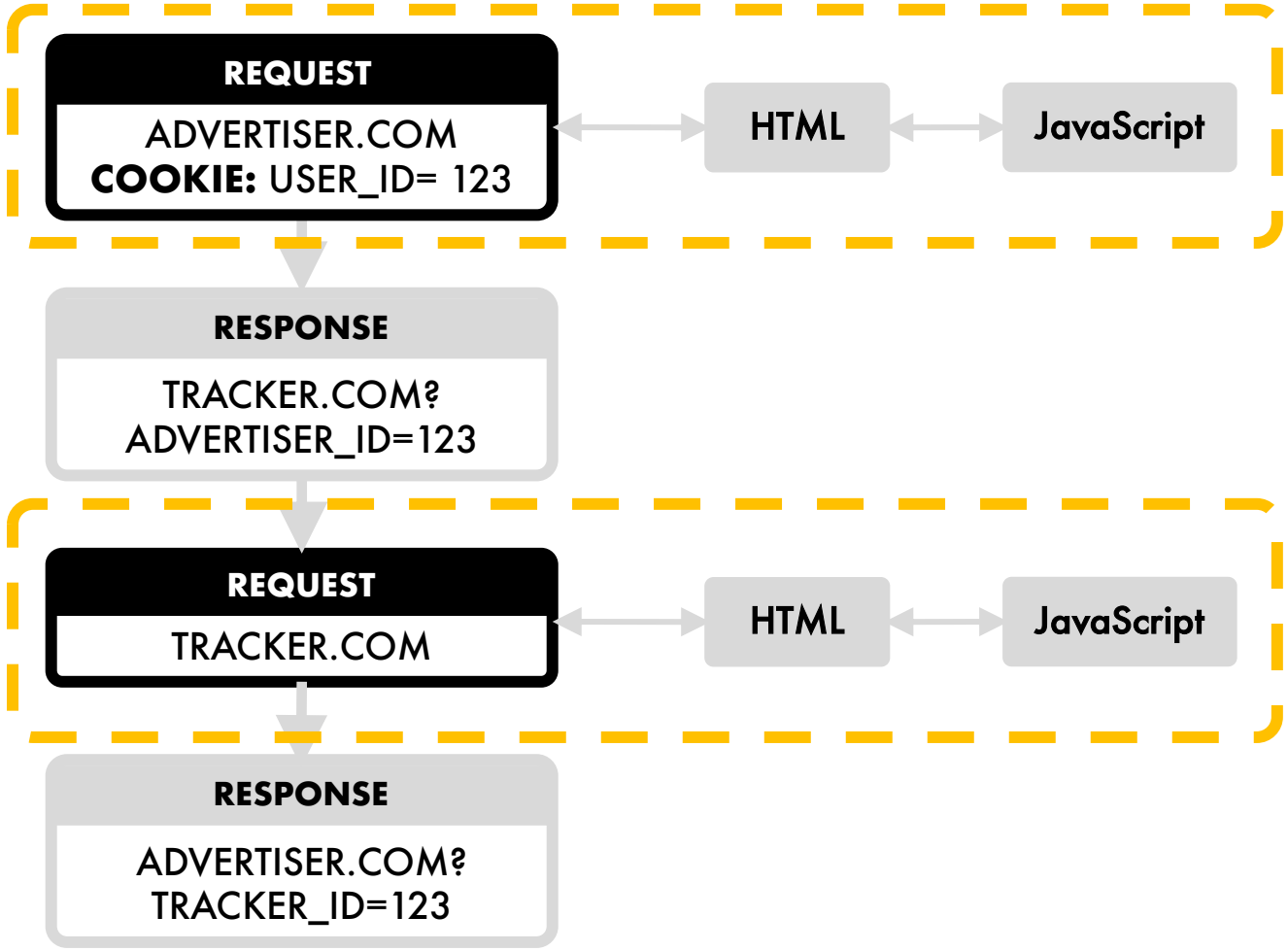
Current solutions: Machine learning based detection

Capture wide range of tracking behaviors



Not **specialized** to detect request chains

Do not **utilize** and **capitalize** on the the sequential nature



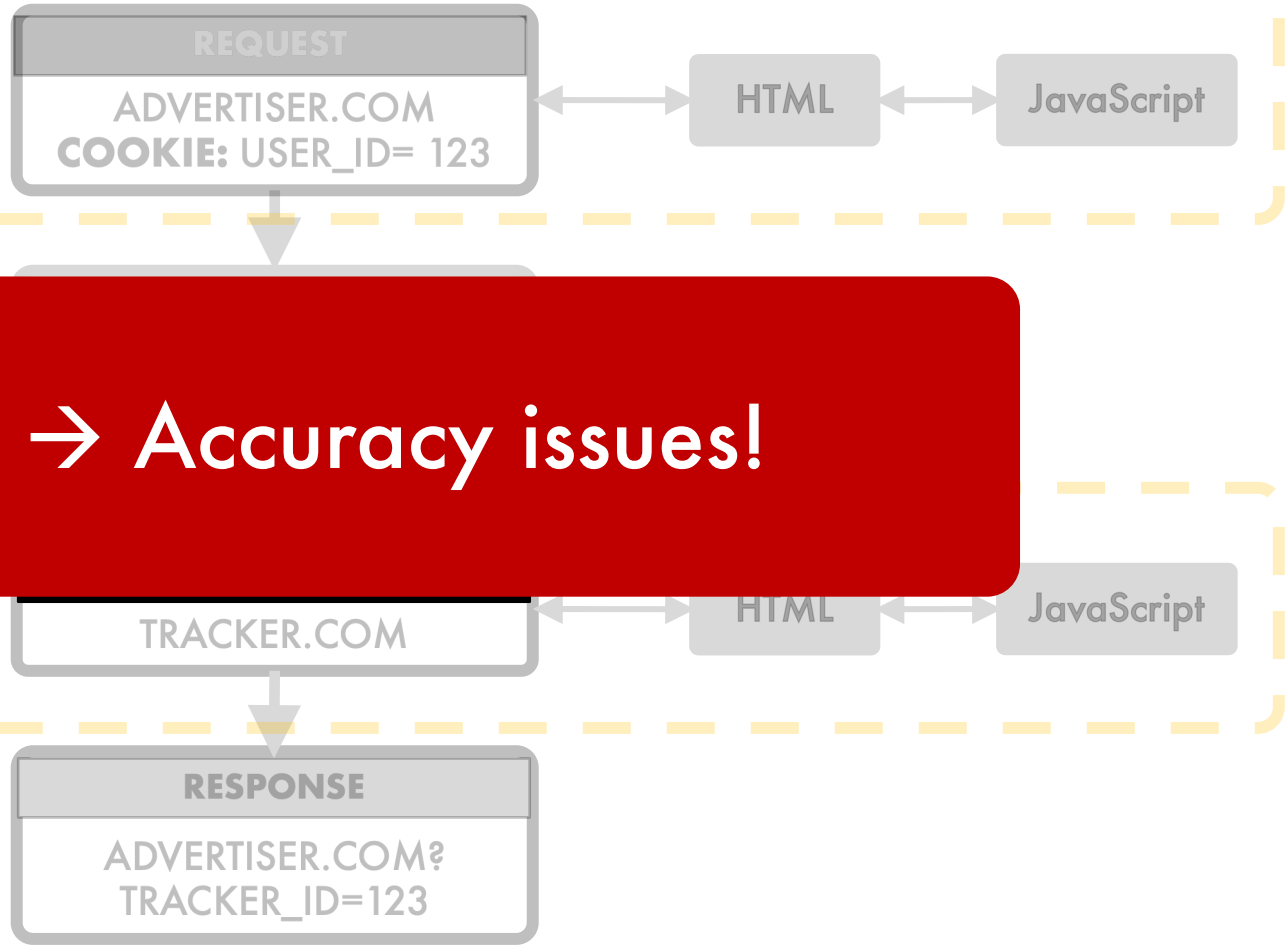
Current solutions: Machine learning based detection

Capture wide range of tracking behaviors

HTML ⟷ JavaScript

Not **specialized** to detect request chains

Do not **utilize** and **capitalize** on the the sequential nature



Not purpose-built → Accuracy issues!

Talk overview

AdTech relies on request chains for tracking & bypassing privacy protections

Existing solutions are ineffective against advertising & tracking request chains

Khaleesi:
A purpose-built approach to protect against advertising & tracking request chains

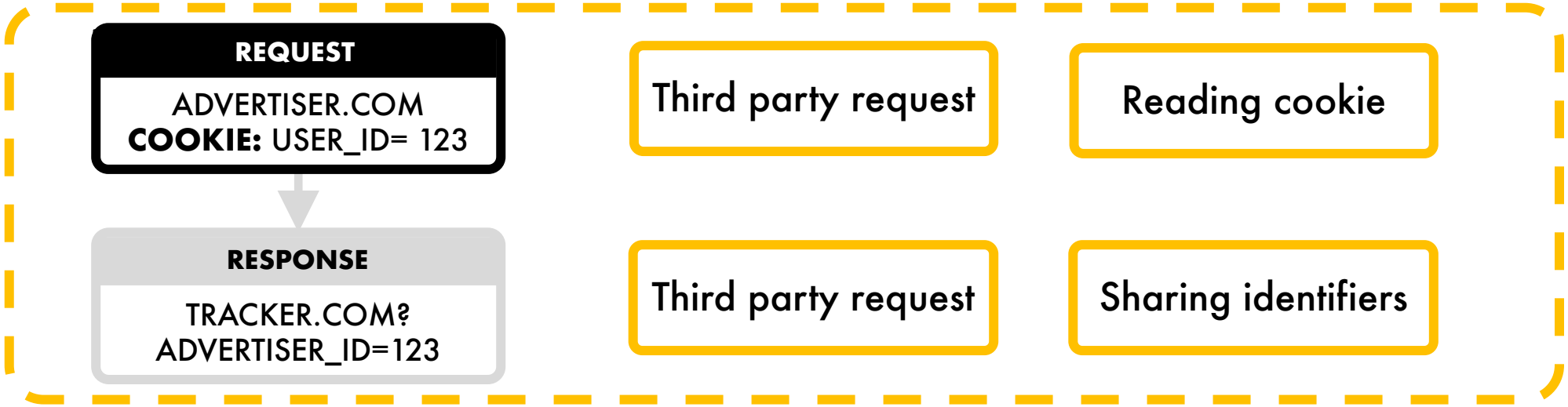
Khaleesi: Motivation & Key idea

Capitalize on the sequential nature of request chains

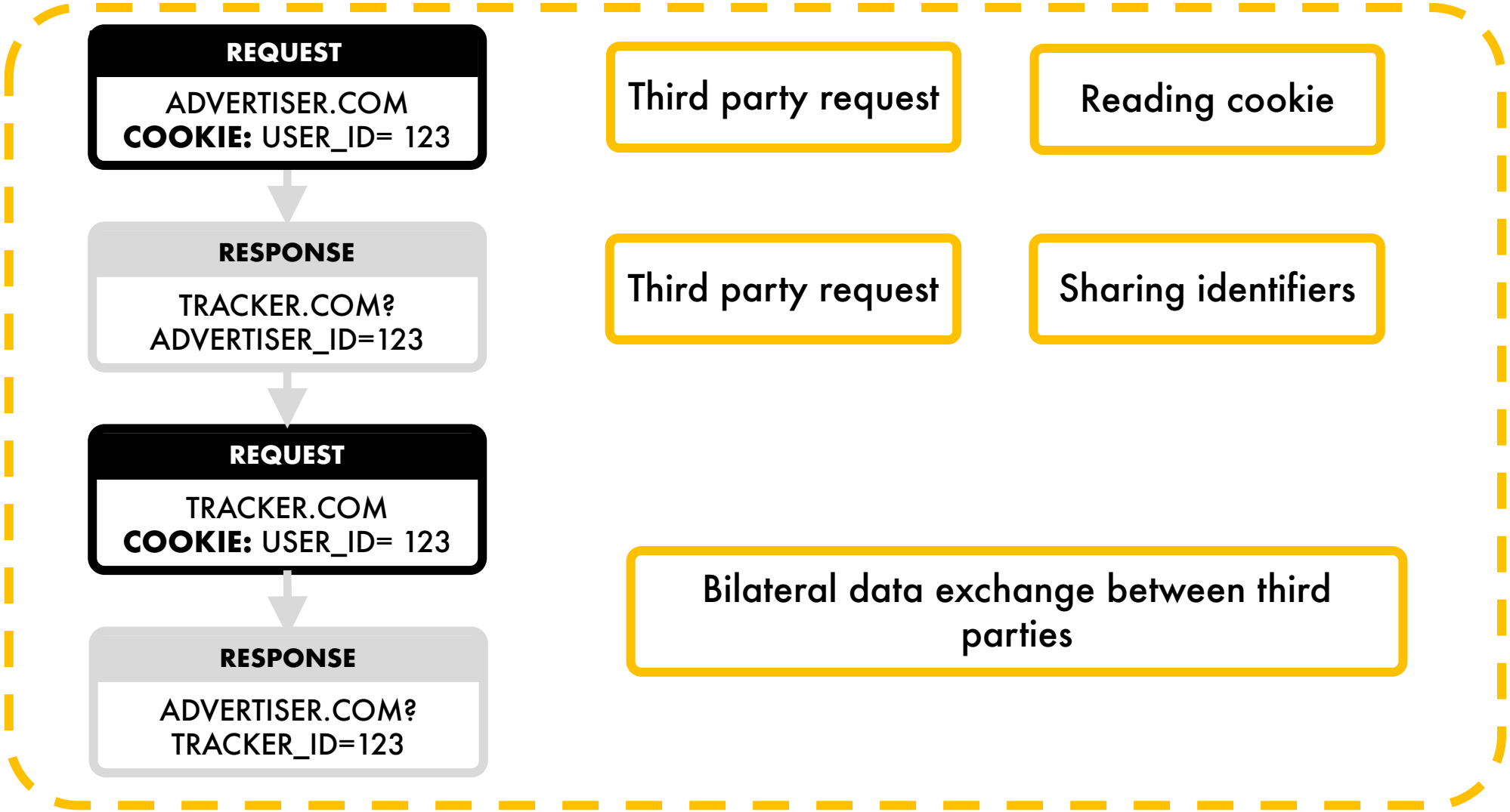
The purpose of request chains becomes clear as they grow

Khaleesi: Motivation & Key idea

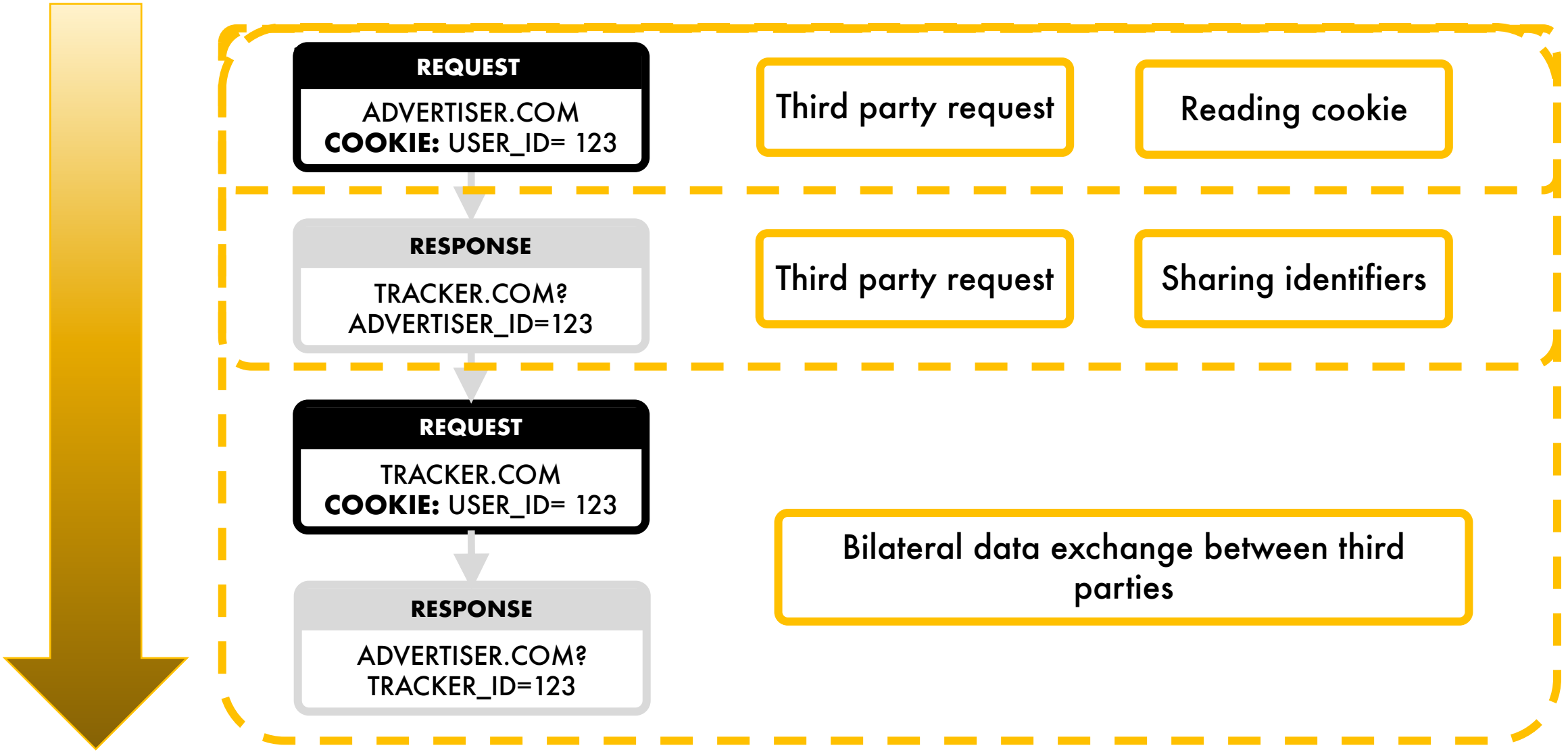
Khaleesi: Motivation & Key idea



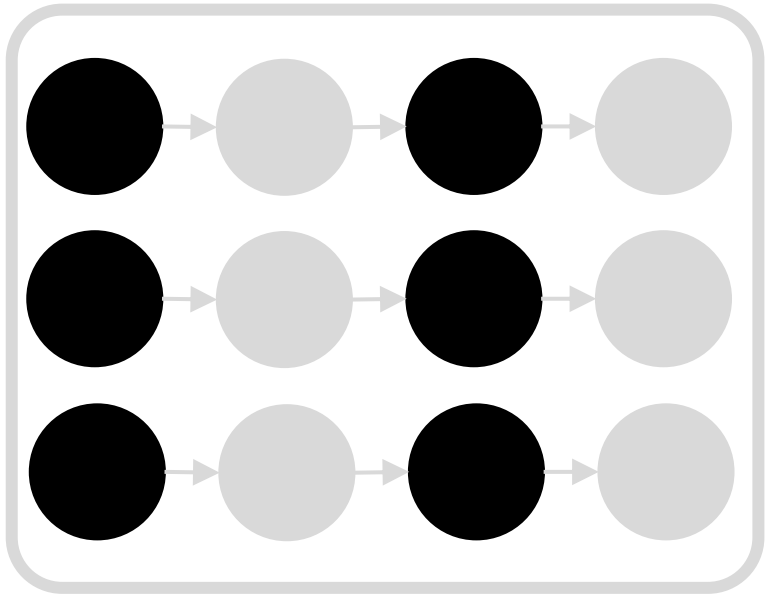
Khaleesi: Motivation & Key idea



Khaleesi: Motivation & Key idea



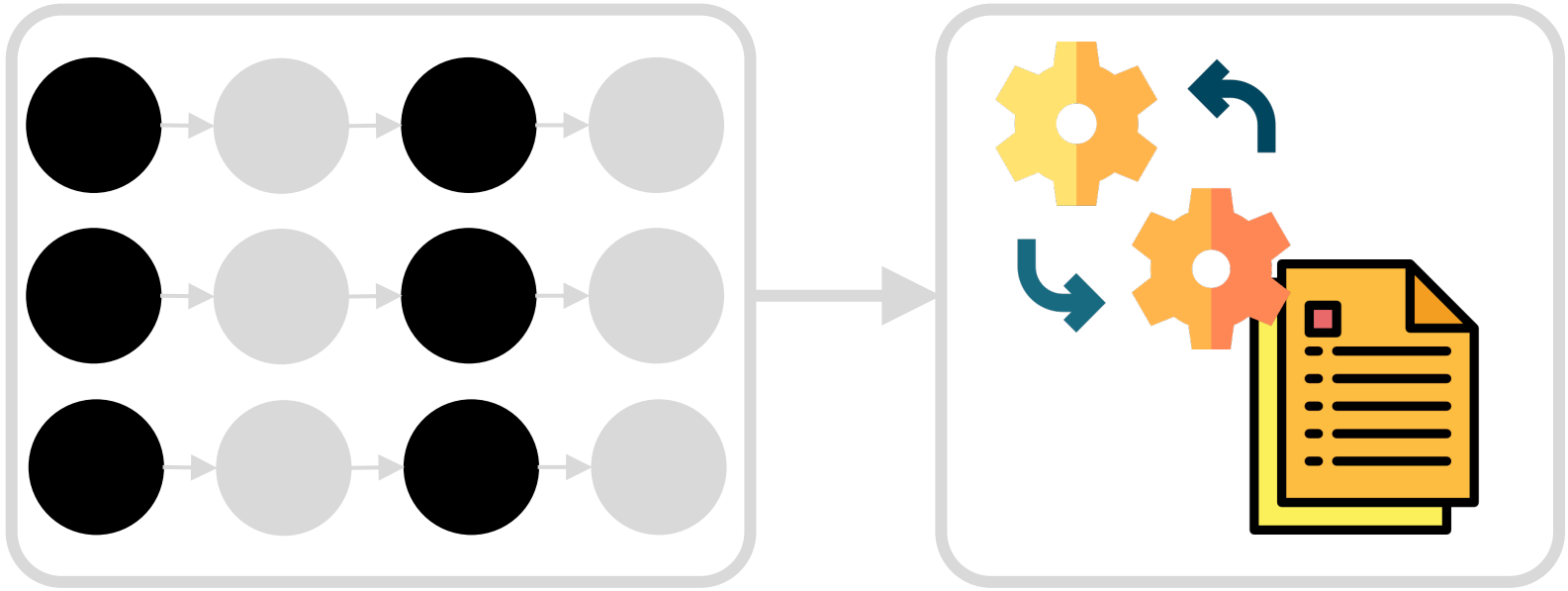
Khaleesi: Overview



Web Extension APIs

HTTP + JavaScript
request chains

Khaleesi: Overview



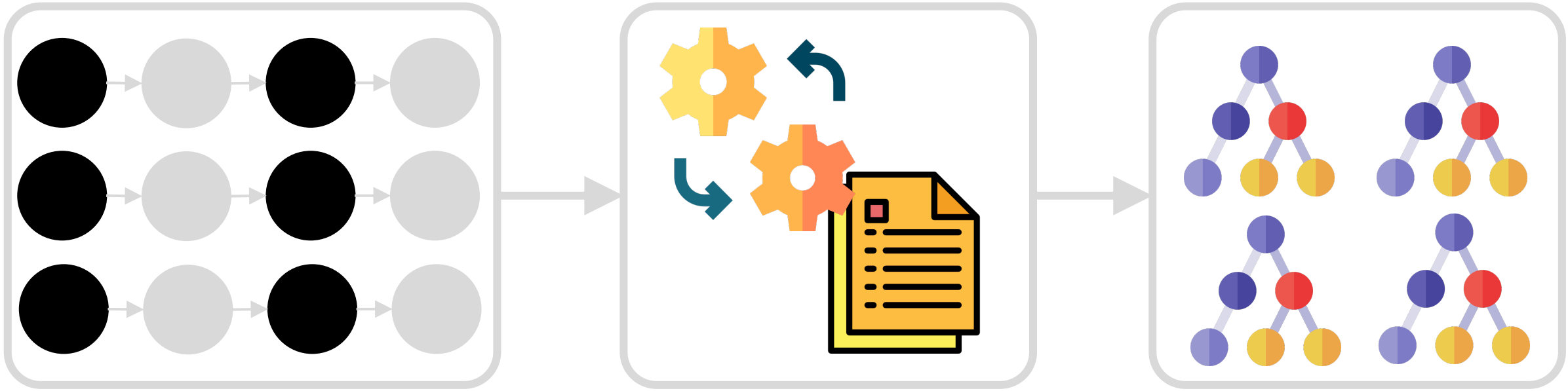
Web Extension APIs

HTTP + JavaScript request chains

Extract request, response, and sequence features

Labeling using Easylist + EasyPrivacy

Khaleesi: Overview



Web Extension APIs

HTTP + JavaScript request chains

Extract request, response, and sequence features

Labeling using Easylist + EasyPrivacy

Machine learning classifier training

Automated detection

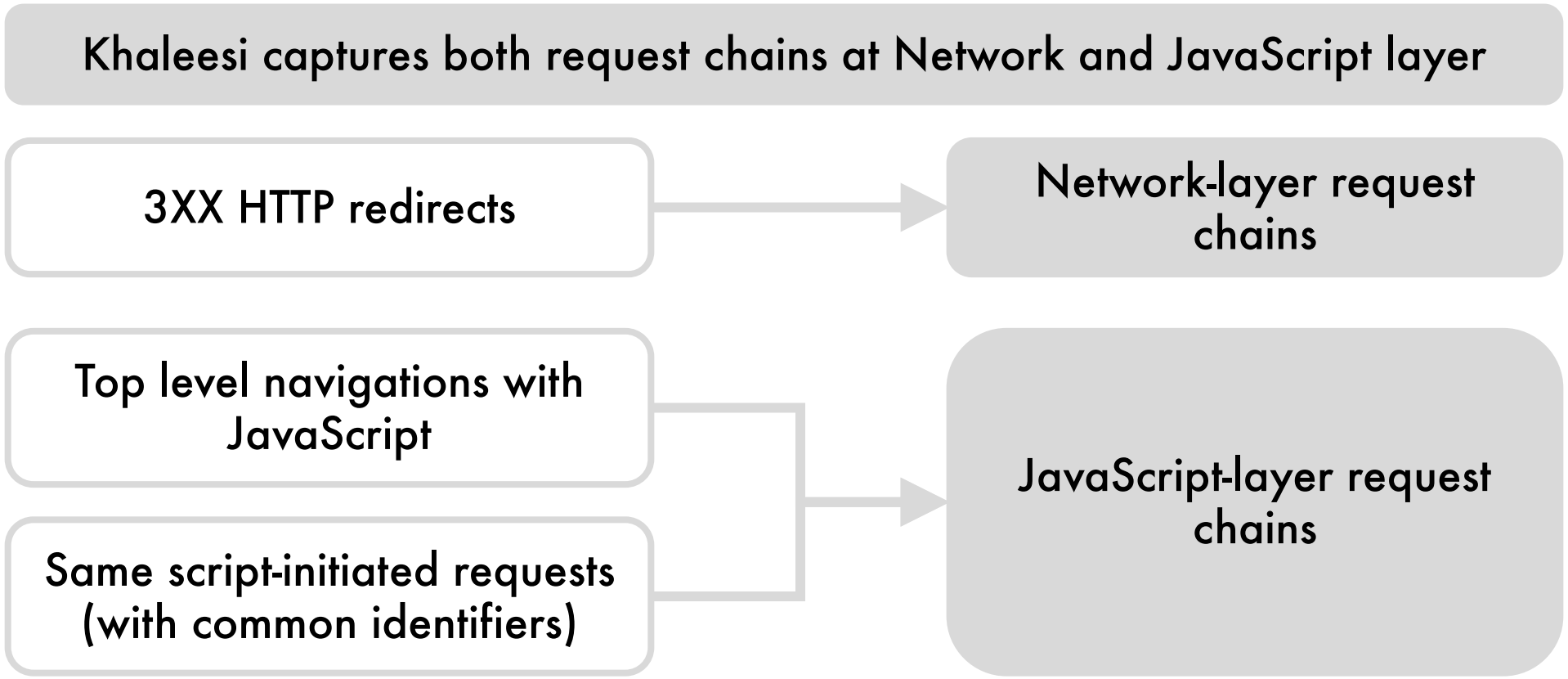
Request chain organization

Khaleesi captures both request chains at Network and JavaScript layer

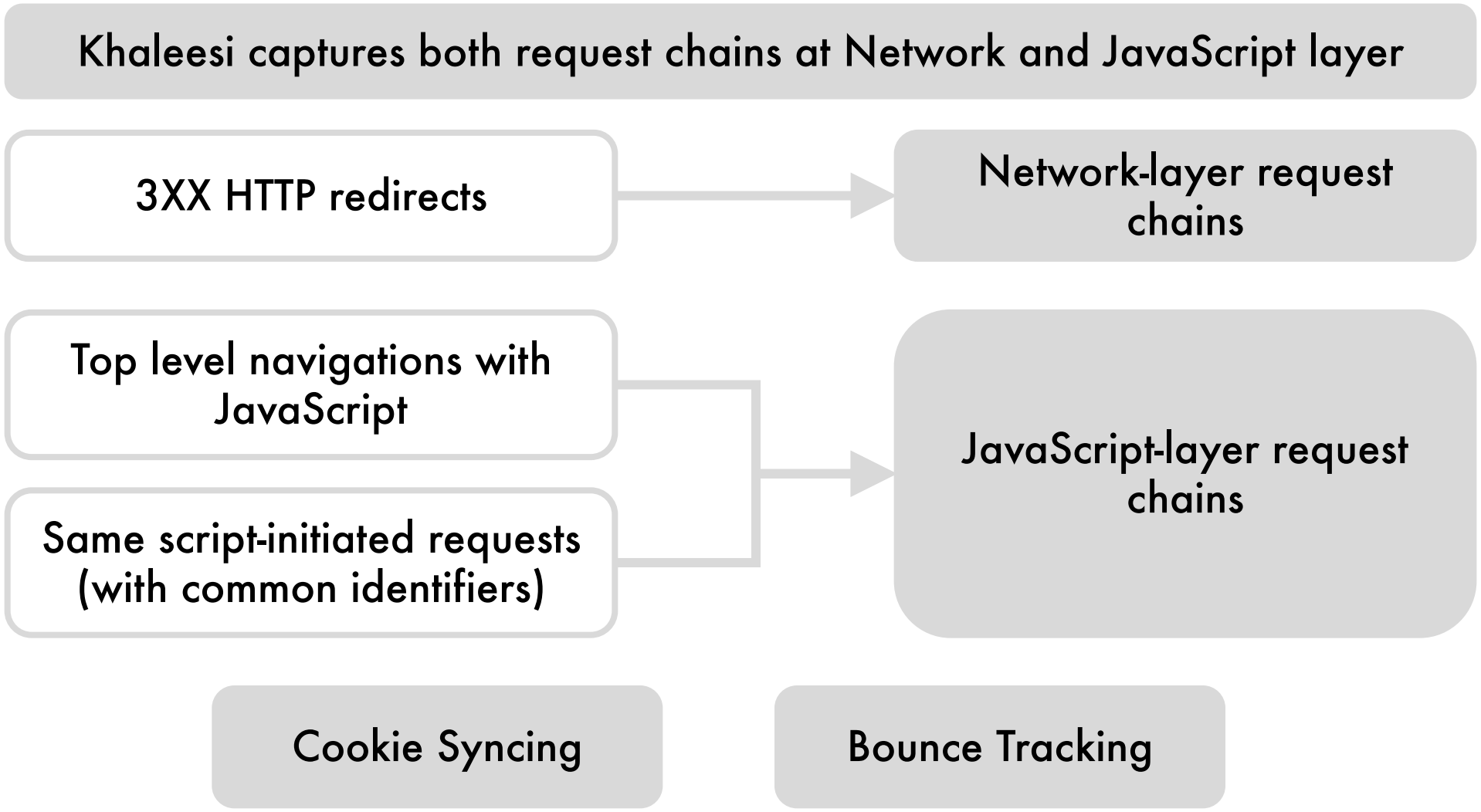
Request chain organization

Khaleesi captures both request chains at Network and JavaScript layer

Request chain organization



Request chain organization



Khaleesi's features

Request features

Response features

Sequence features

Khaleesi's features

Request features

Response features

Sequence features

What the request might do

Example feature

Long URL length → Trackers trying to embed identifiers

Khaleesi's features

Request features	Response features	Sequence features
What the request might do	What the response will do	
Example feature	Example feature	
Long URL length → Trackers trying to embed identifiers	**Length of the response** → loading of a tracking pixel	

Khaleesi's features

Request features	Response features	Sequence features
What the request might do	What the response will do	What the chain has been doing
Example feature	Example feature	Example feature
Long URL length → Trackers trying to embed identifiers	**Length of the response** → loading of a tracking pixel	**Unique domains in a chain** → Trackers sending data to other trackers

Khaleesi: Classification & Accuracy

Random forest ensemble

Khaleesi: Classification & Accuracy

Random forest ensemble

EasyList/EasyPrivacy ground truth

Alexa top-10K websites

Khaleesi: Classification & Accuracy

Random forest ensemble

EasyList/EasyPrivacy ground truth

Alexa top-10K websites

98.63% accuracy

1,259 new ad/tracking domains identified

@UMAARR6

Key Takeaways

Download Khaleesi

AdTech relies on request chains for tracking & bypassing privacy protections

Existing solutions are ineffective against advertising & tracking request chains

Khaleesi:
A purpose-built approach to protect against advertising & tracking request chains

Lots of additional analysis in the paper (e.g., Robustness, Performance, etc.)

More details in the paper!