A Longitudinal Analysis of the ads.txt Standard

Muhammad Ahmad Bashir   Sajjad Arshad   Engin Kirda

William Robertson   Christo Wilson

Northeastern University
Khoury College of Computer Sciences

What is ads.txt Standard?

Introduced by Interactive Advertising Bureau (IAB)
to combat '**domain spoofing**' fraud
in **Real Time Bidding (RTB)** auctions
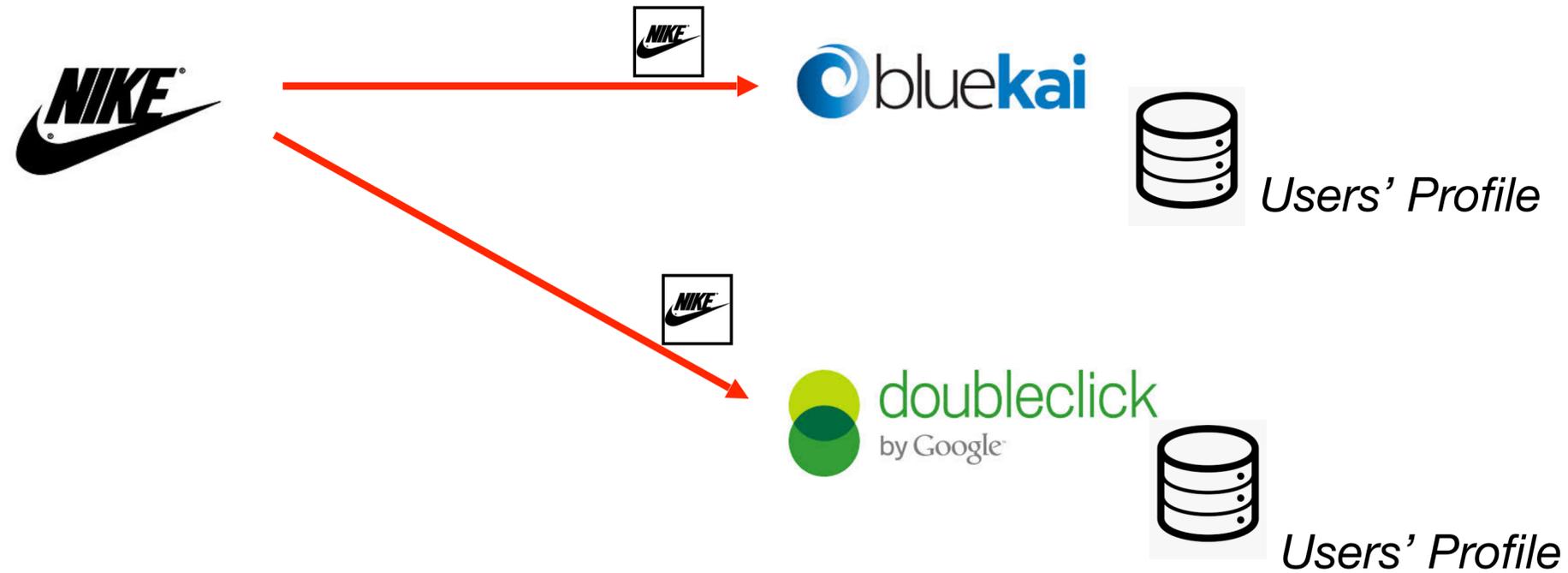
Online Advertising

Website / Publisher   **Advertising & Analytics (A&A) Companies**
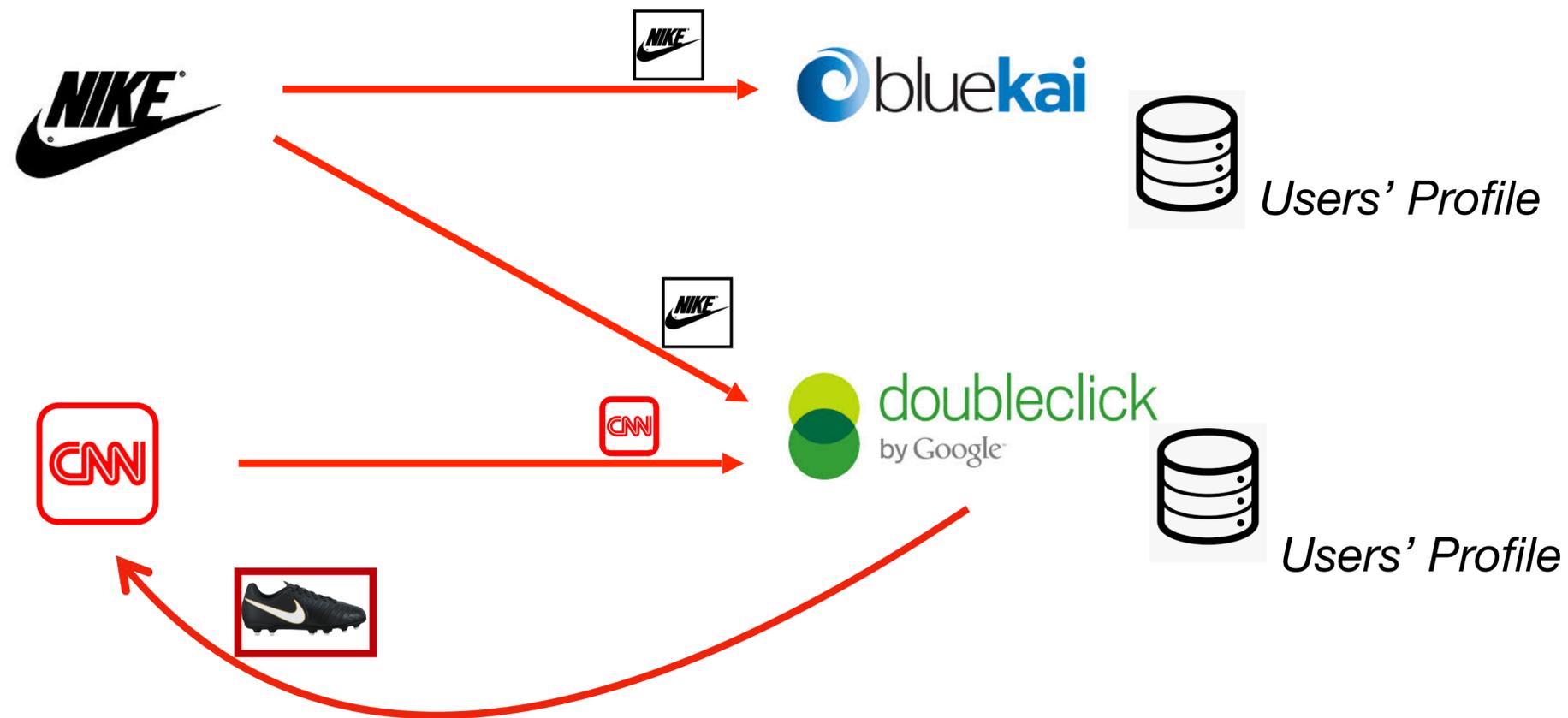
Online Advertising

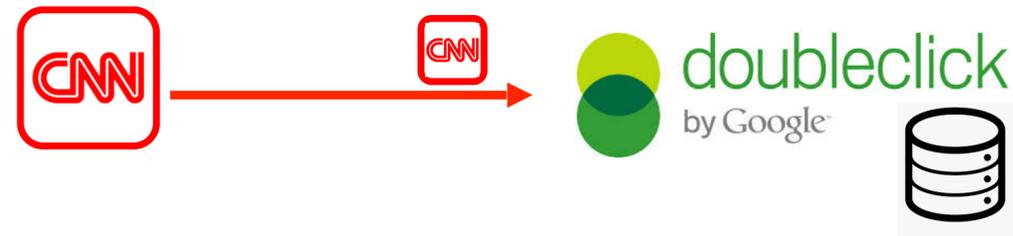**Website / Publisher**   **Advertising & Analytics (A&A) Companies**

Users' Profile

Users' Profile

Online Advertising

Website / Publisher   **Advertising & Analytics (A&A) Companies**

Users' Profile

Users' Profile

Shift Towards Real Time Bidding (RTB)

Website / Publisher   **Advertising & Analytics (A&A) Companies**

━━━━━▶ **Observable via browser**
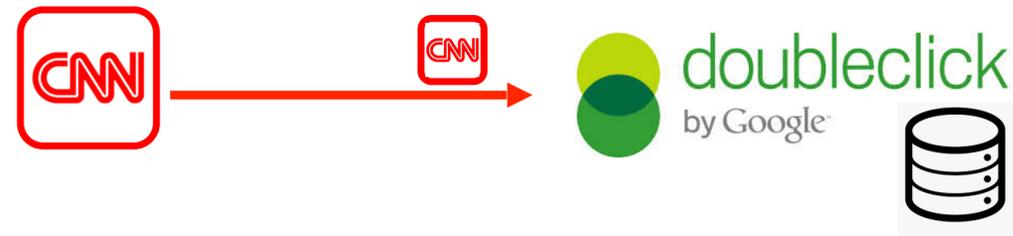
┄┄┄┄▶ **Not observable**

Shift Towards Real Time Bidding (RTB)

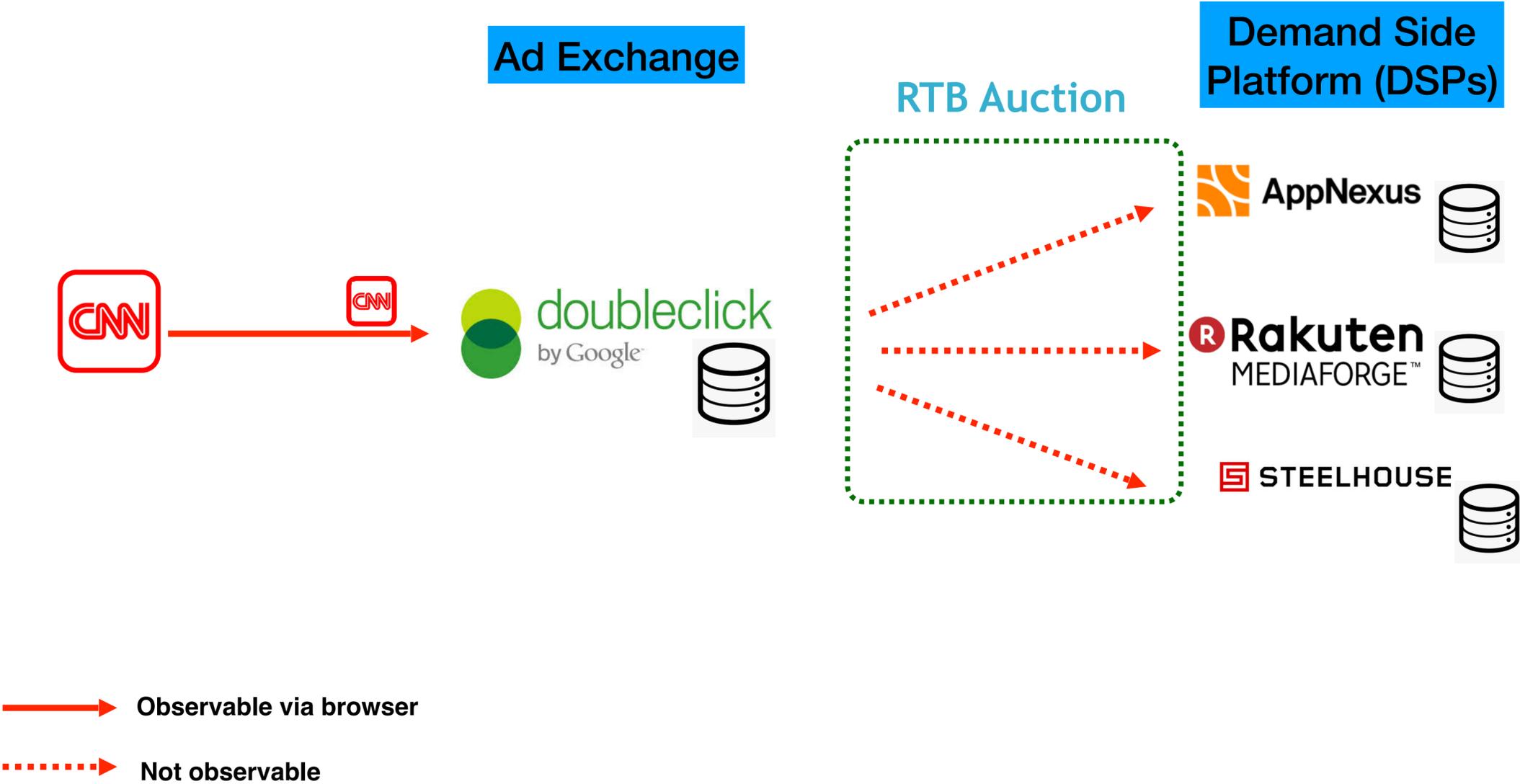**Website / Publisher**　　　**Advertising & Analytics (A&A) Companies**

Ad Exchange

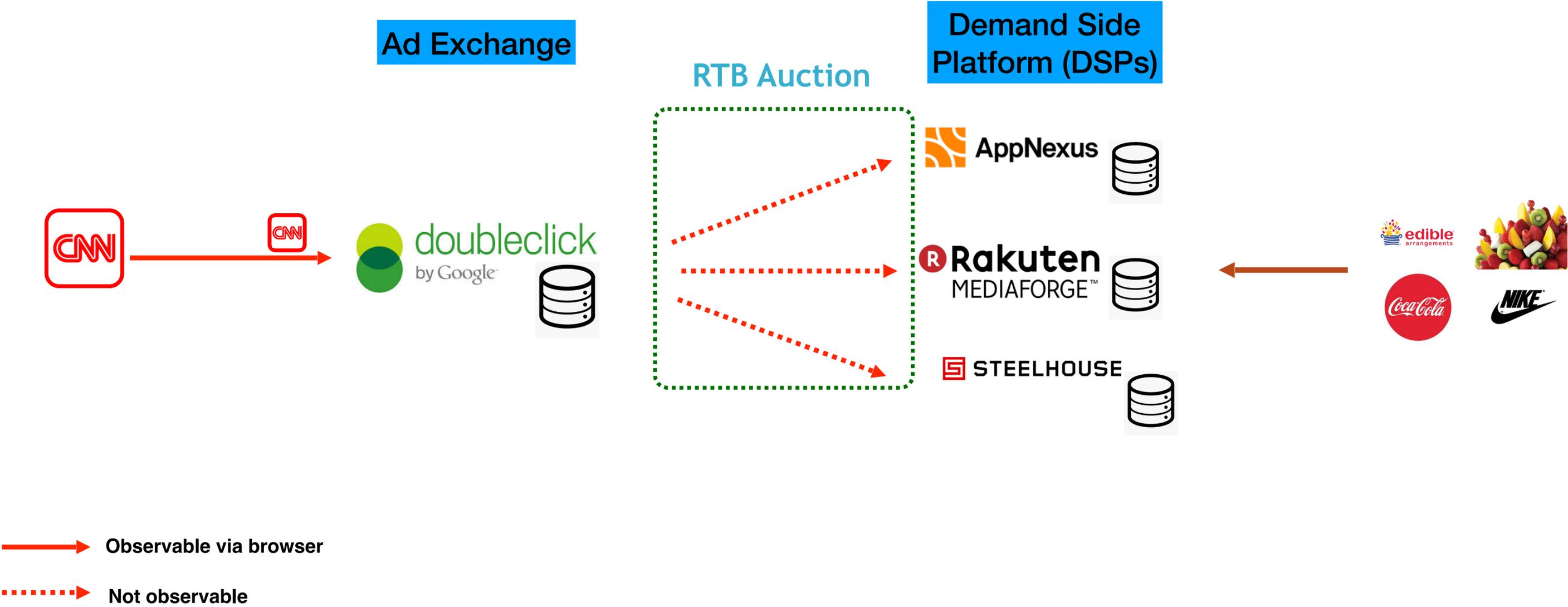

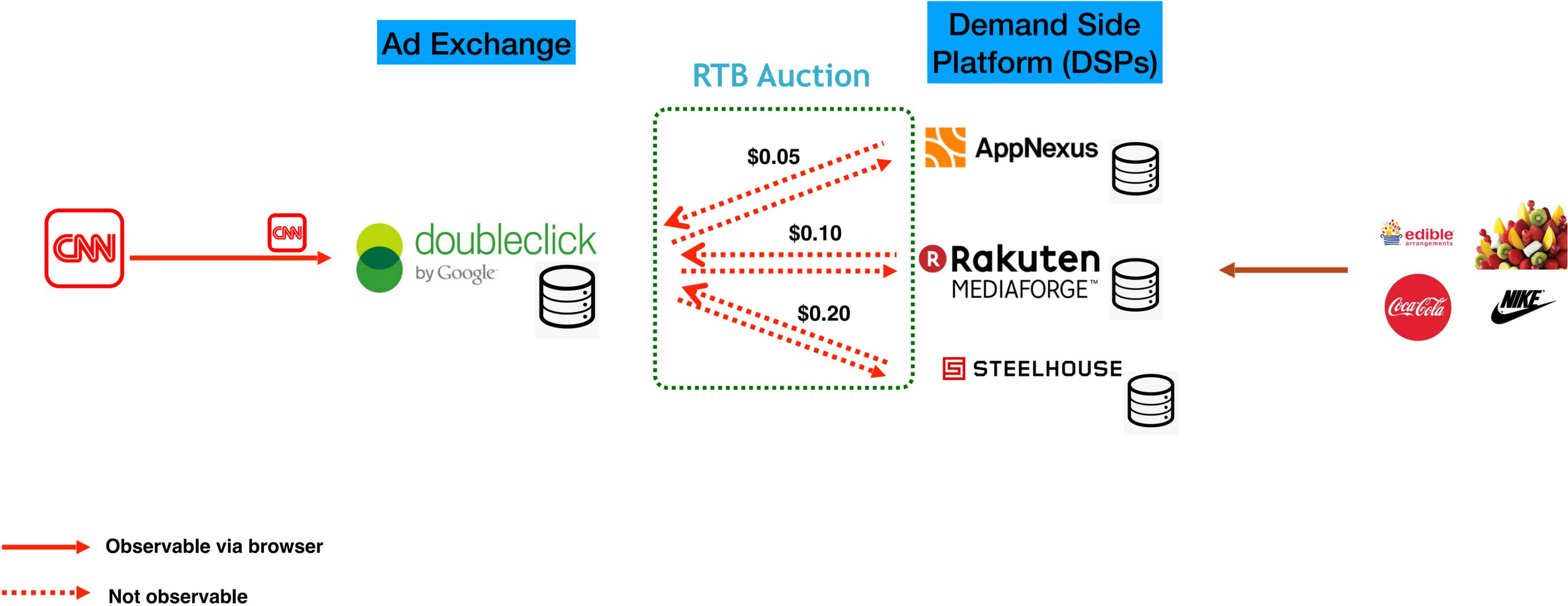→ Observable via browser

┈┈> Not observable

Shift Towards Real Time Bidding (RTB)

Website / Publisher

Advertising & Analytics (A&A) Companies

Shift Towards Real Time Bidding (RTB)

Website / Publisher   **Advertising & Analytics (A&A) Companies**   **Advertisers**

Observable via browser

Not observable

Shift Towards Real Time Bidding (RTB)

Website / Publisher

Advertising & Analytics (A&A) Companies

Advertisers

Ad Exchange

RTB Auction

Demand Side Platform (DSPs)
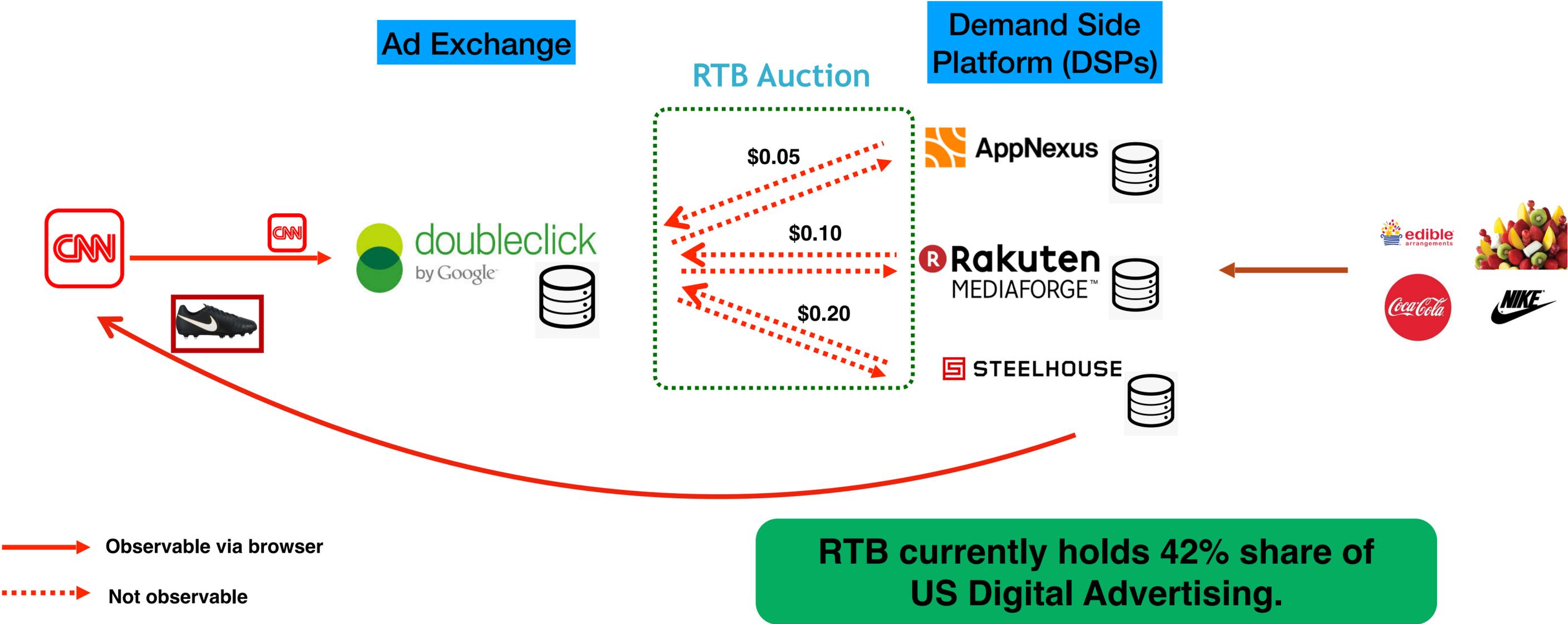
$0.05

$0.10

$0.20

Observable via browser

Not observable

Shift Towards Real Time Bidding (RTB)

Website / Publisher

Advertising & Analytics (A&A) Companies

Advertisers

Observable via browser

Not observable

Shift Towards Real Time Bidding (RTB)

An Example of RTB Auction

Observable via browser

Not observable

UserX

Publisher

Ad Exchange

UserX=123

Demand Side Platforms (DSPs)

UserX=xo$

UserX=ABC

An Example of RTB Auction

An Example of RTB Auction

Observable via browser

Not observable

UserX **Publisher** **Ad Exchange**

doubleclick
by Google

UserX=123
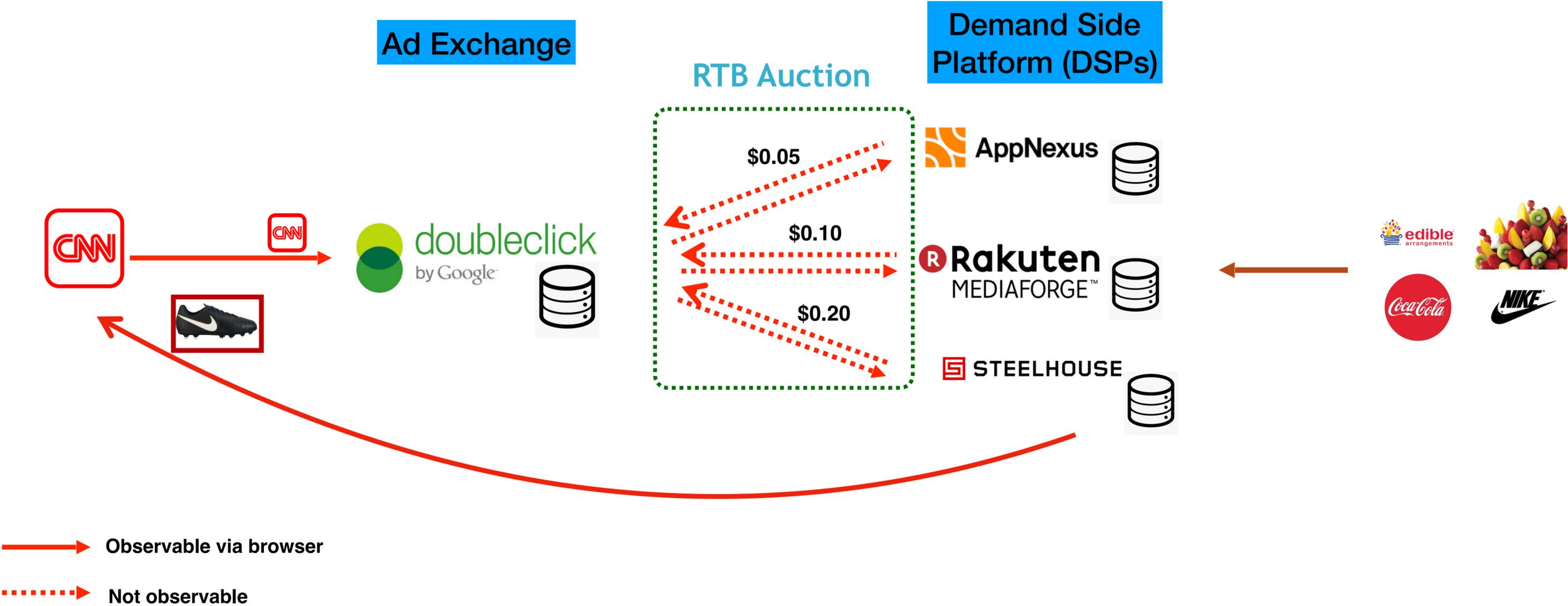
Demand Side Platforms (DSPs)

GET, CNN's Cookie

GET, DoubleClick's Cookie = 123

criteo.

UserX=xo$

rm

UserX=ABC

An Example of RTB Auction

Observable via browser

Not observable

UserX   **Publisher**   **Ad Exchange**   **RTB Auction**   **Demand Side Platforms (DSPs)**

UserX=123

GET, CNN's Cookie

GET, DoubleClick's Cookie = 123

UserX=xo$

UserX=ABC

An Example of RTB Auction

An Example of RTB Auction

Observable via browser

Not observable

UserX — Publisher — Ad Exchange — RTB Auction — Demand Side Platforms (DSPs)

UserX=123

GET, CNN's Cookie

GET, DoubleClick's Cookie = 123

Bid for User 123 ?
$0.15

Bid for User 123 ?
$0.40

criteo
UserX=xo$

rm
UserX=ABC

An Example of RTB Auction

An Example of RTB Auction

Potential for Fraud in RTB Auctions

Publishers   **Ad Exchanges**

Observable via browser
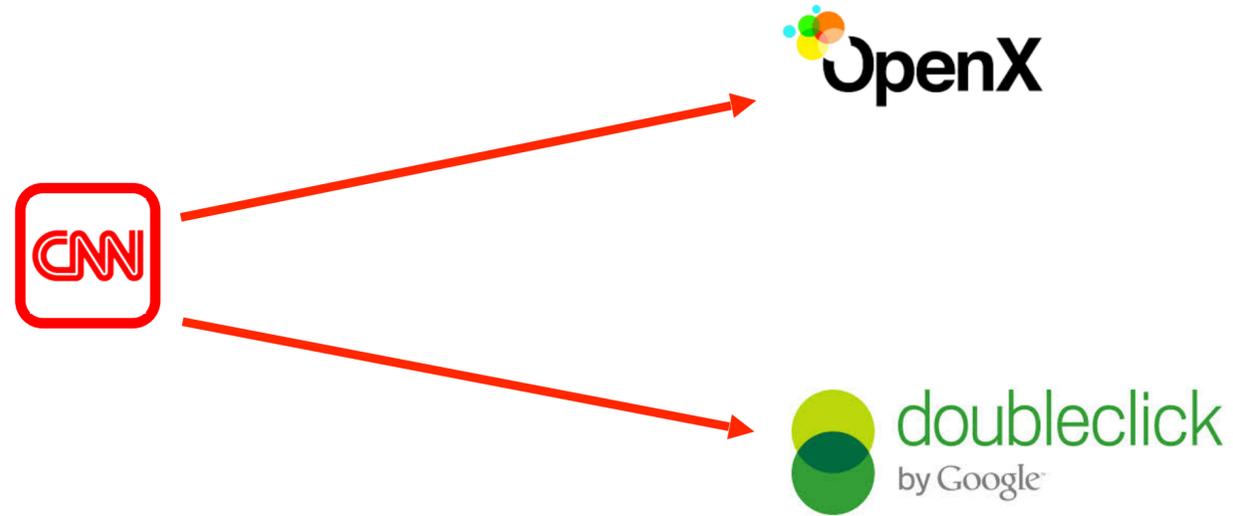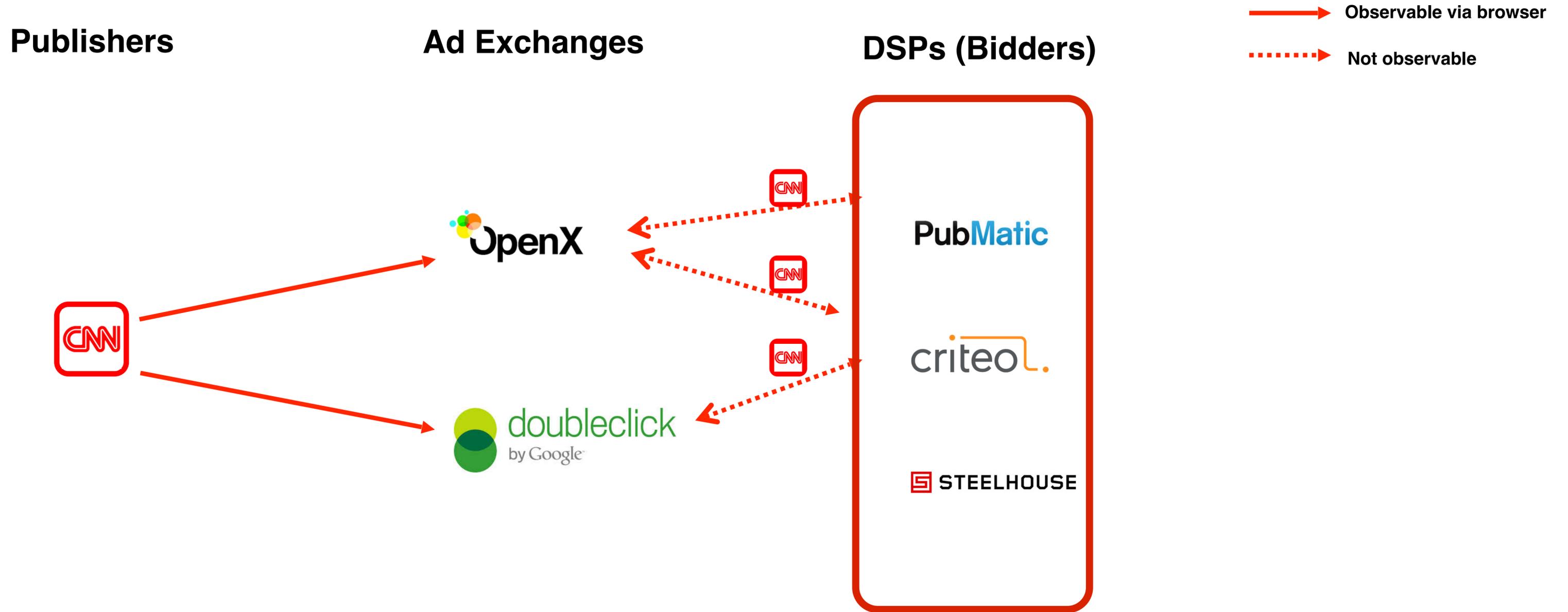
Not observable

Potential for Fraud in RTB Auctions

Publishers

Ad Exchanges

DSPs (Bidders)

Potential for Fraud in RTB Auctions

Publishers

Ad Exchanges

DSPs (Bidders)

→ Observable via browser

⇢ Not observable

Potential for Fraud in RTB Auctions

Publishers   **Ad Exchanges**   **DSPs (Bidders)**

Observable via browser

Not observable
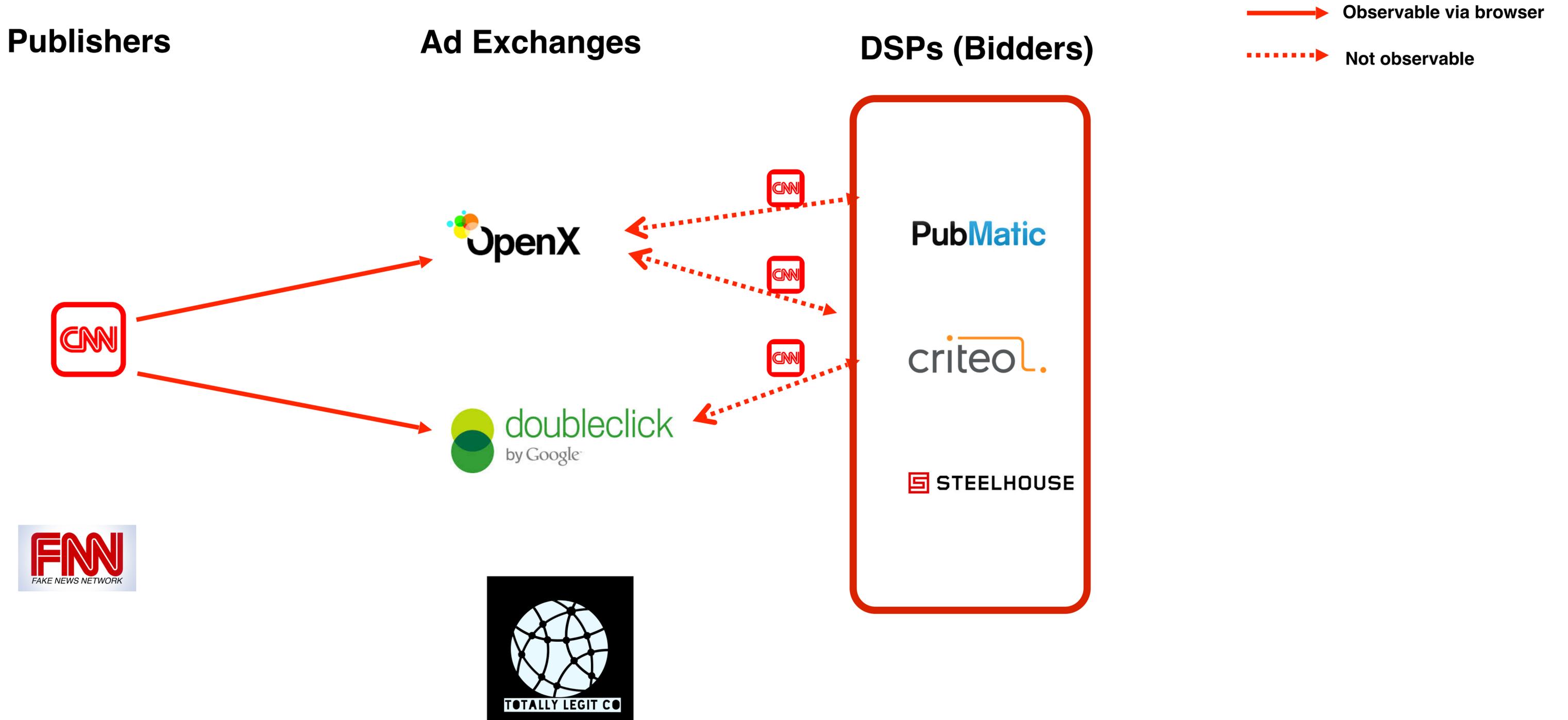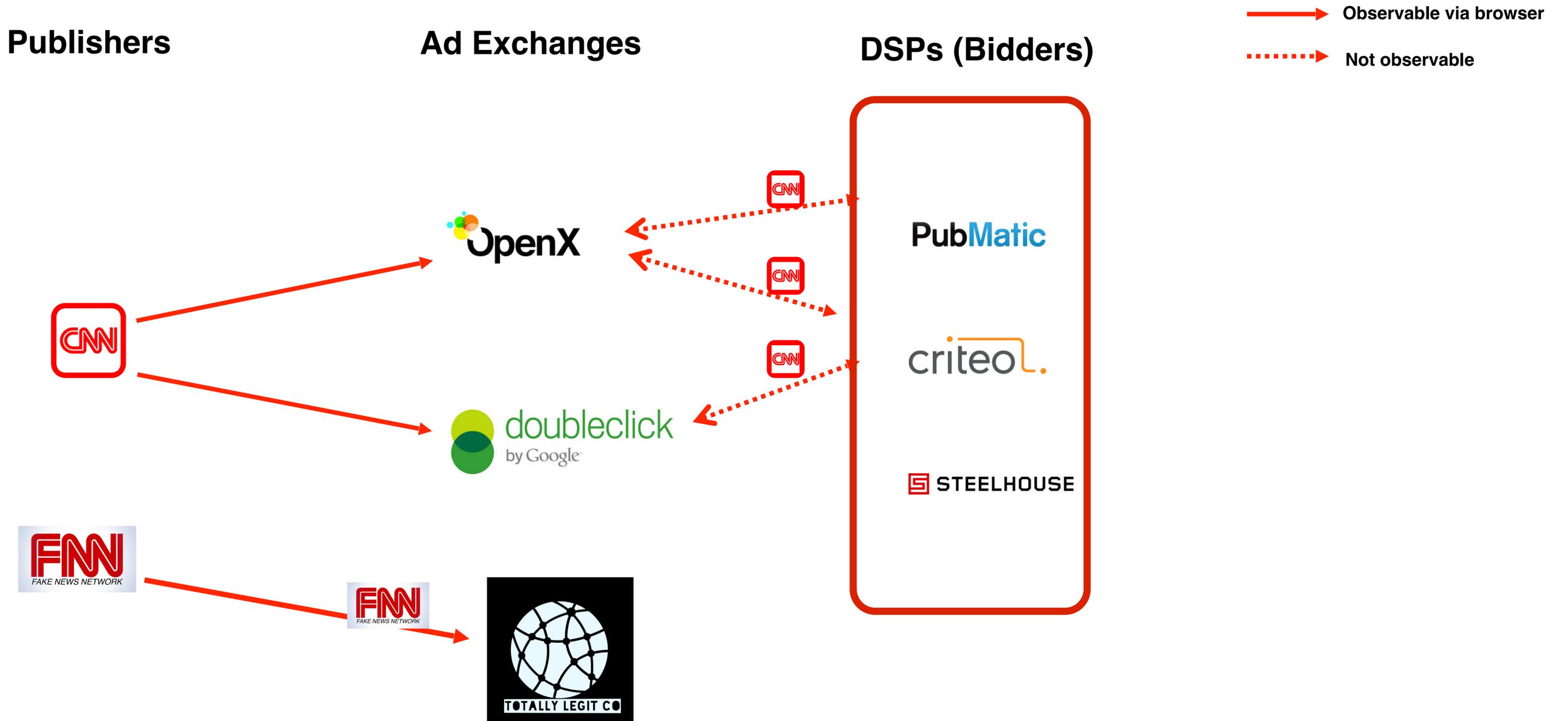
Potential for Fraud in RTB Auctions
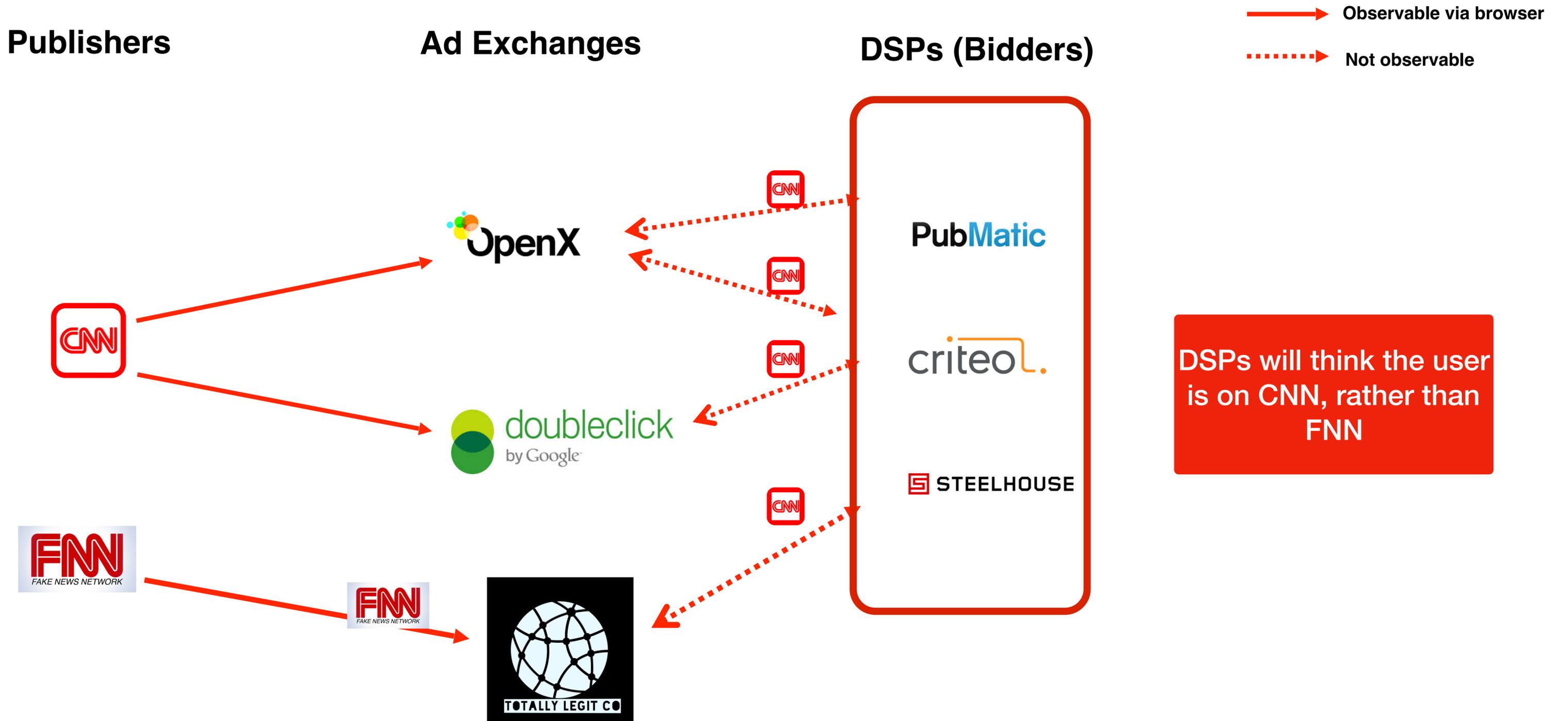
Ad Fraud = Loss in $$$

Ad Fraud = Loss in $$$

Methbot: a $3M-$5M/day video ad-tech fraud

AdAge

| FCB, BBDO moves | Dentsu forecasts | NYT ad sales | McDonald's cups |

Digital →

AD FRAUD WILL COST $7.2 BILLION IN 2016, ANA SAYS, UP NEARLY $1 BILLION

Study Recommends Better Understanding of Programmatic Supply Chain

By George P. Sl...

Talk to an enthusias... and how i...

The $8.2 Billion Adtech Fraud Problem That Everyone Is Ignoring

Samuel Scott   4 years ago

Comment

Login

Startups
Apps
Gadgets
Videos
Audio
Extra Crunch NEW
Newsletters
—
Events
Advertise
Crunchbase
More

Search

Facebook privacy
Google
Transportation
Enterprise

YOU WIN !!!
100 500 millions!
OK

White Ops, a security firm, has published a detailed report on a crime-ring they call "Methbot" that generated $3M-$5M by creating 6,000 fake websites to embed videos in, then generating convincing bots that that appeared to

THE PROGRAMMATIC MARKETER

Domain spoofing remains a huge threat to programmatic

FEBRUARY 28, 2017   *by* Yuyu Chen

ORIGIN from
fraud /frɔːd/
to re

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

https://cnn.com/ads.txt

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT

openx.com, 183753,RESELLER

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

https://cnn.com/ads.txt

```
google.com, pub–7439281311086140, DIRECT
rubiconproject.com, 11078, DIRECT
c.amazon–adsystem.com, 3159, DIRECT
openx.com, 183753,RESELLER
```

File Format:

<FIELD #1>, <FIELD #2>, <FIELD #3>

#1: Exchange domain *
#2: Publisher's account ID
#3: Type of account/relationship

* Not always an exchange

Potential for Fraud in RTB Auctions

Publishers **Ad Exchanges** **DSPs (Bidders)**

→ Observable via browser

⇢ Not observable
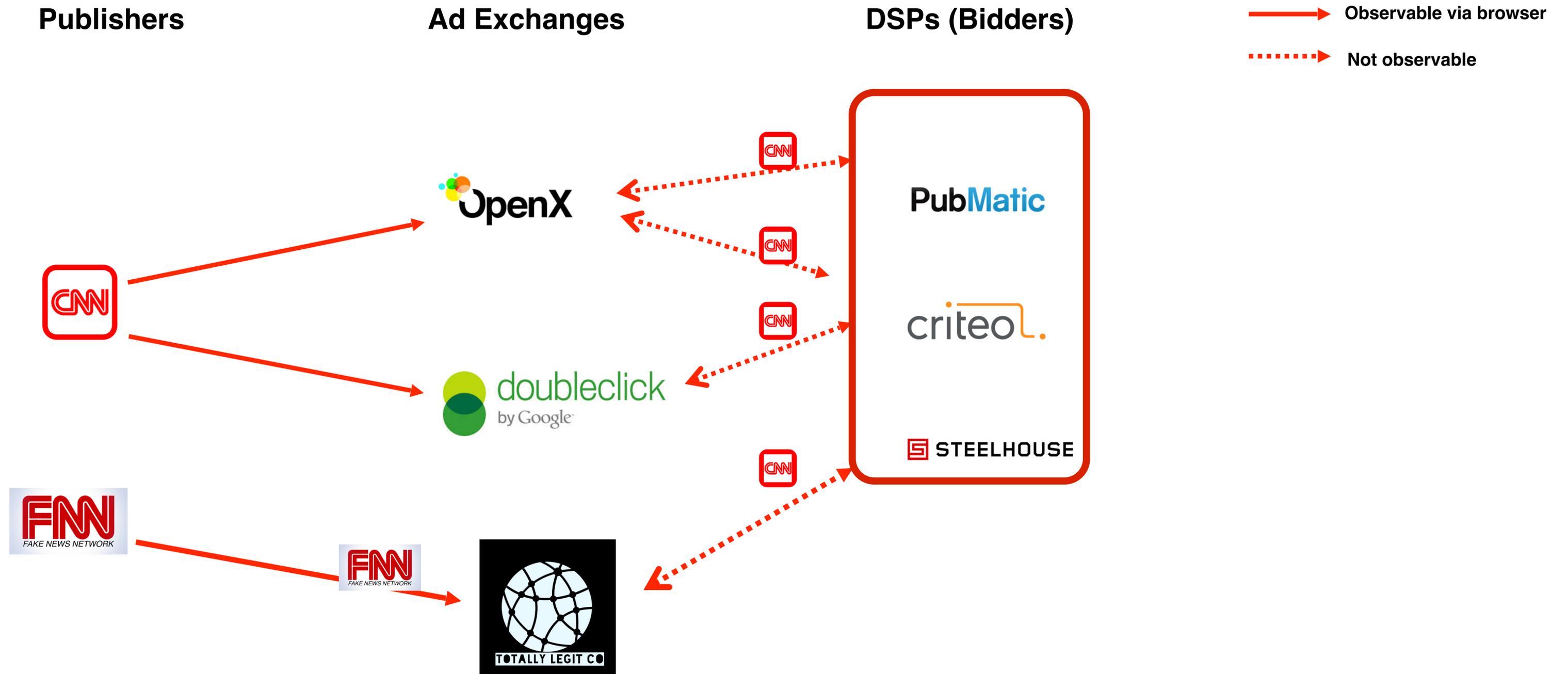
Potential for Fraud in RTB Auctions
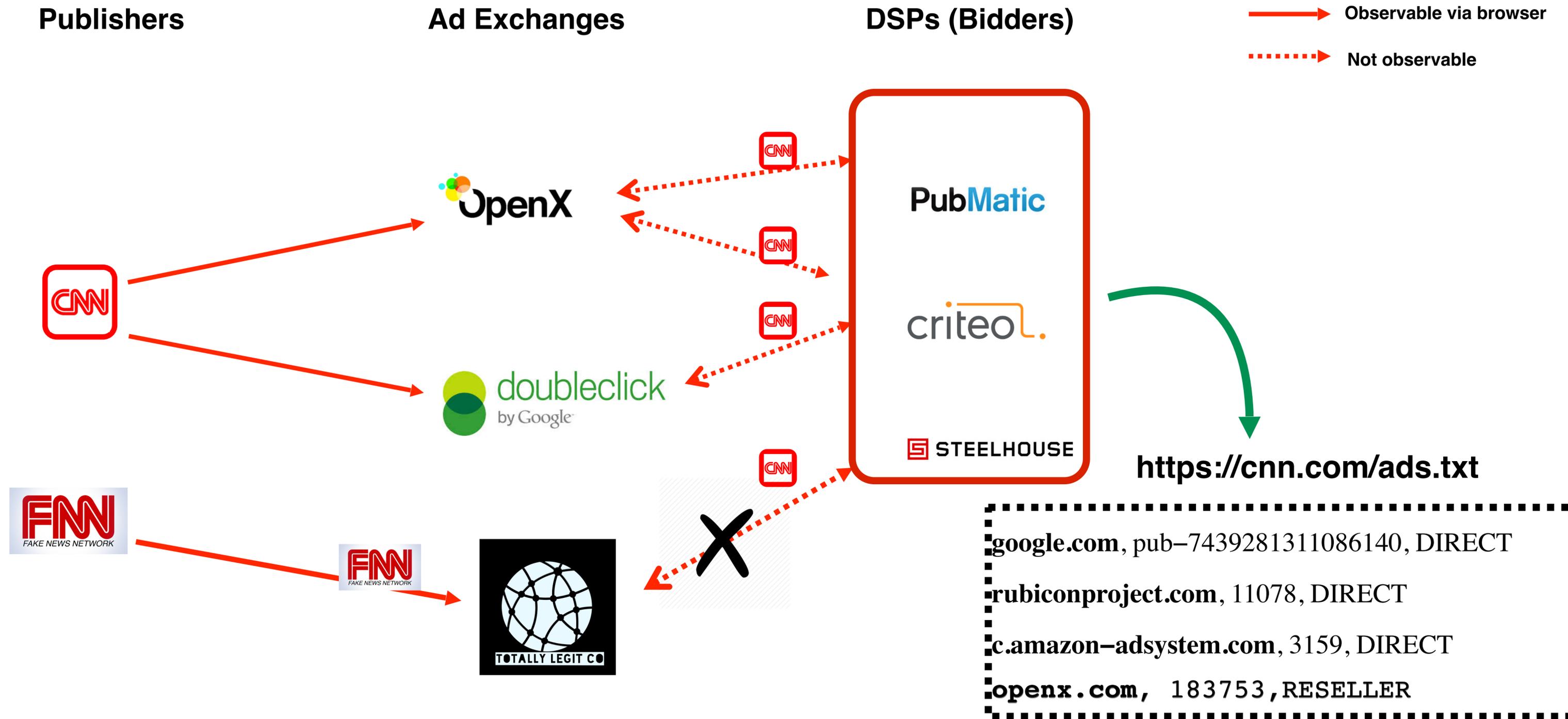
Why Study ads.txt standard?

Why Study ads.txt standard?

Potential to combat domain spoofing fraud

- Check for any shortcomings.

- Lessons can be learnt for upcoming standards.

Why Study ads.txt standard?

Potential to combat domain spoofing fraud

- Check for any shortcomings.

- Lessons can be learnt for upcoming standards.

Might be useful for privacy researchers

- Means of bringing transparency in the advertising ecosystem

- Publishers are disclosing their relationships with ad exchanges

Goals / Research Questions

1. Are publishers adopting the ads.txt standard?

2. Are major A&A companies embracing the standard?

   - Which A&A companies have adopted?

   - Do these companies appear to be complying with the standard?

3. Can this data be useful for transparency of the ad ecosystem?
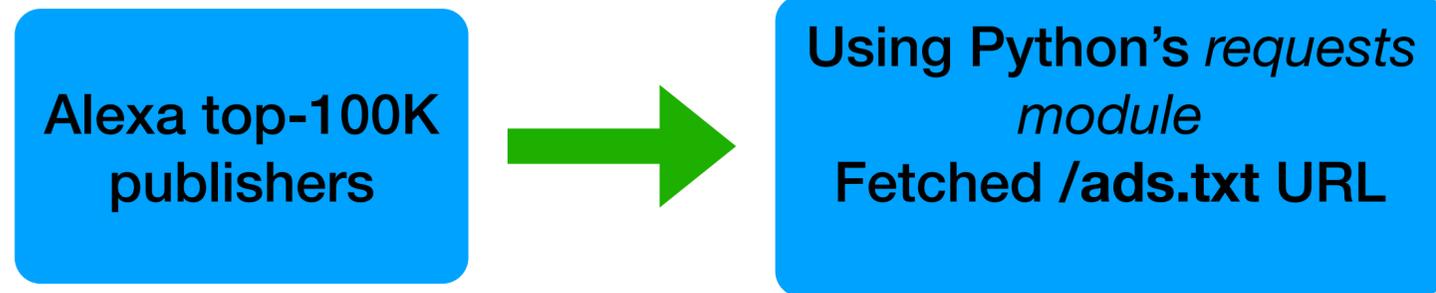
Talk Overview

1. RTB background & motivation for the ads.txt standard

2. Data collection

3. Adoption trends

4. Compliance of the ads.txt standard

5. Means of providing transparency

Data Collection Methodology

Data Collection Methodology

Alexa top-100K publishers

Data Collection Methodology

https://cnn.com/ads.txt

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT

openx.com, 183753,RESELLER

Alexa top-100K publishers

Using Python's *requests module*
Fetched **/ads.txt** URL

Data Collection Methodology

Alexa top-100K publishers

➡

Using Python's *requests module*
Fetched /ads.txt URL

➡

Stored output as base64

https://cnn.com/ads.txt

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT

openx.com, 183753,RESELLER

Data Collection Methodology

Alexa top-100K publishers

Using Python's *requests module*
Fetched **/ads.txt** URL

Stored output as base64

Crawled 15 URLs per publisher using Chrome Debugging

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT
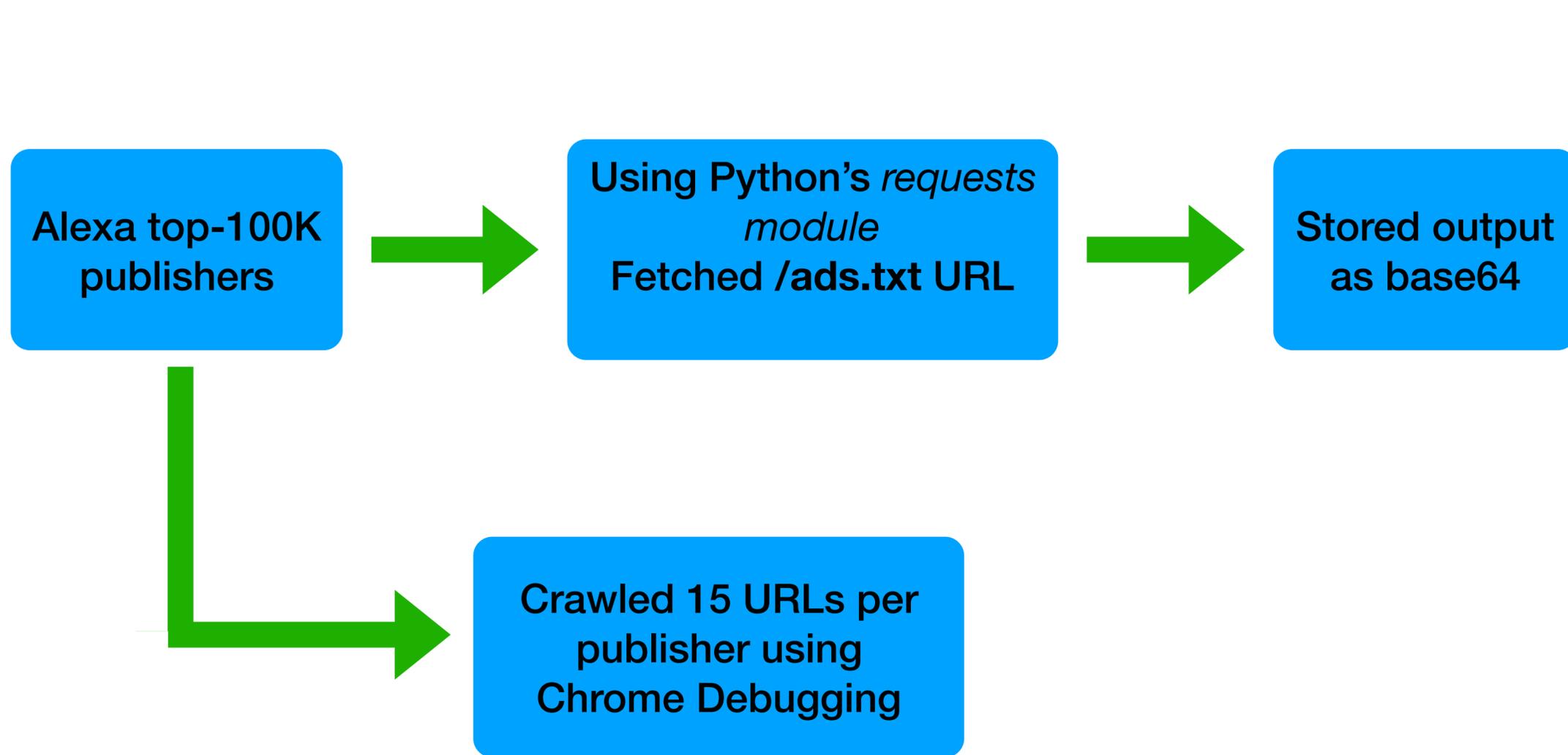
openx.com, 183753,RESELLER

Data Collection Methodology

https://cnn.com/ads.txt

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT
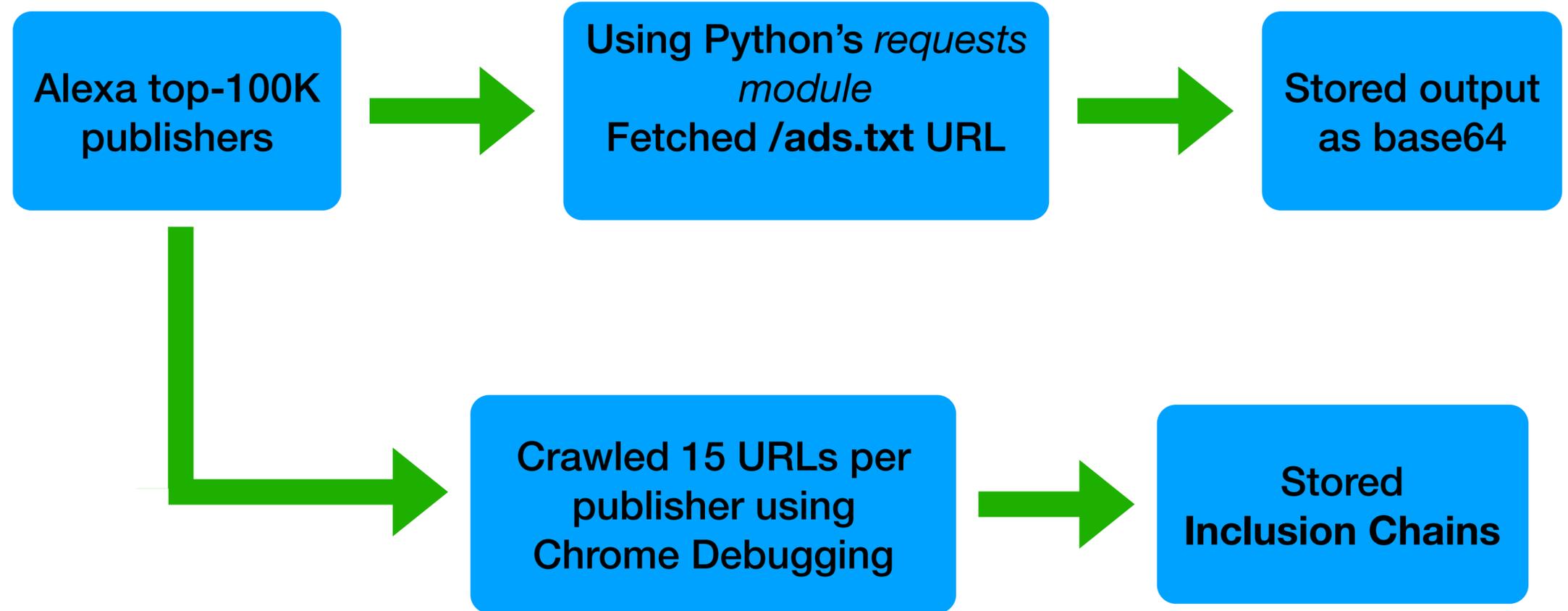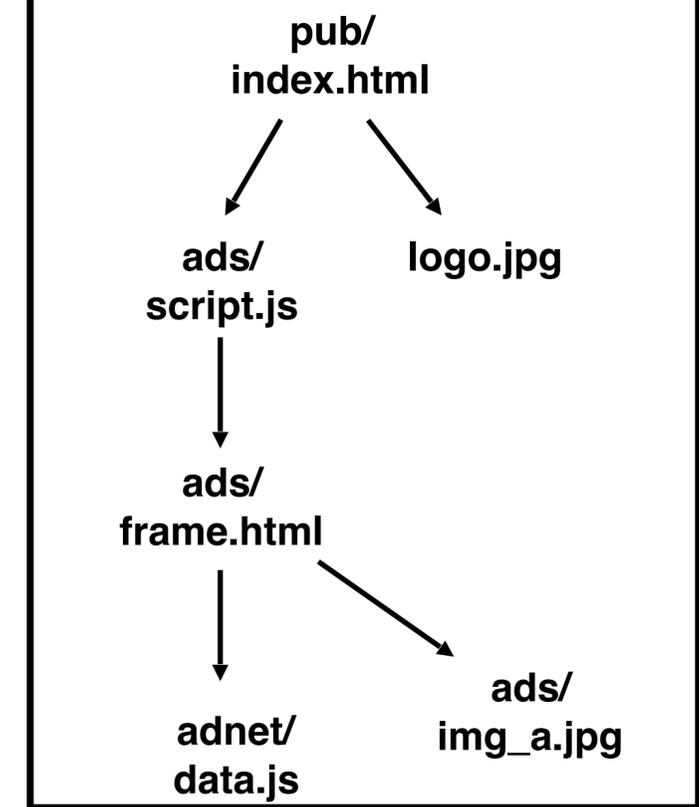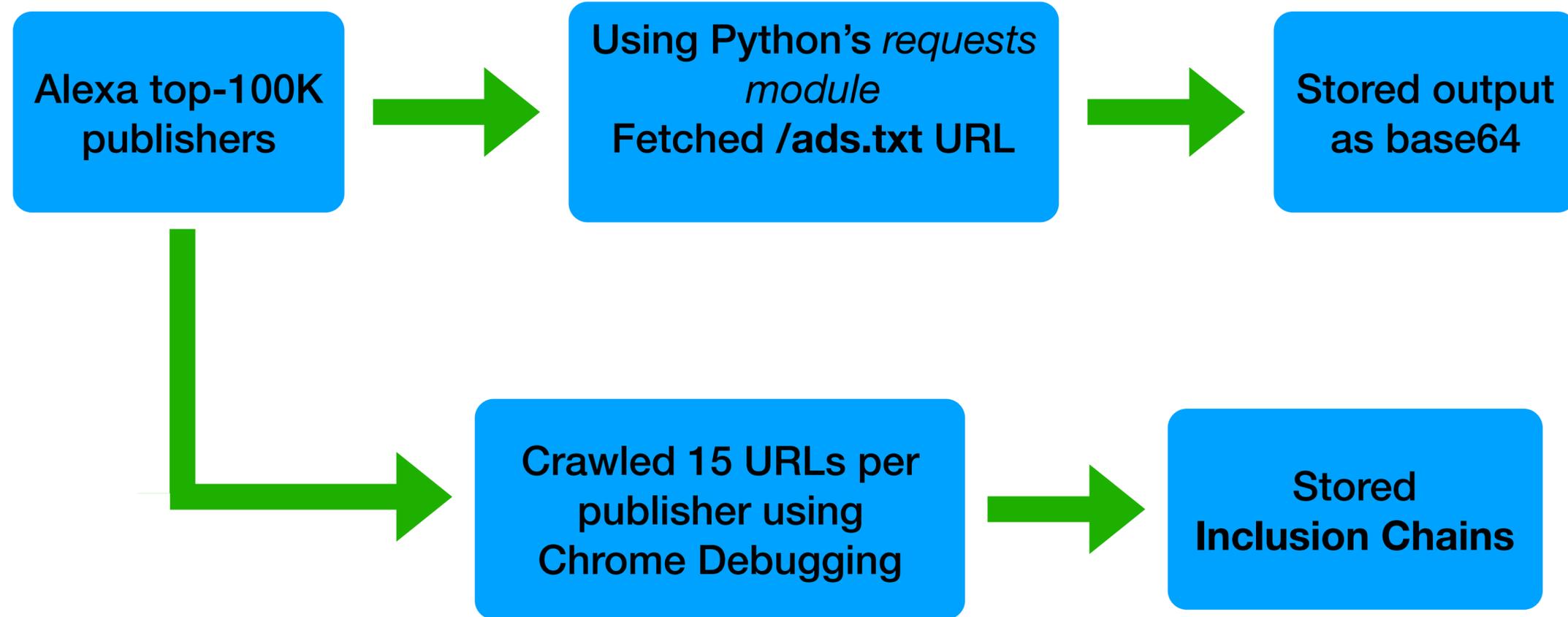
openx.com, 183753,RESELLER

Alexa top-100K publishers → Using Python's *requests module* Fetched **/ads.txt** URL → Stored output as base64

Alexa top-100K publishers → Crawled 15 URLs per publisher using Chrome Debugging → Stored **Inclusion Chains**

Example Inclusions

pub/
index.html

ads/
script.js

logo.jpg

ads/
frame.html

adnet/
data.js

ads/
img_a.jpg

Data Collection Methodology

Alexa top-100K publishers → **Using Python's *requests module* Fetched /ads.txt URL** → **Stored output as base64**

Crawled 15 URLs per publisher using Chrome Debugging → **Stored Inclusion Chains**

- **Collected 26 snapshots between January 2018 - April 2019**

https://cnn.com/ads.txt

google.com, pub−7439281311086140, DIRECT

rubiconproject.com, 11078, DIRECT

c.amazon−adsystem.com, 3159, DIRECT

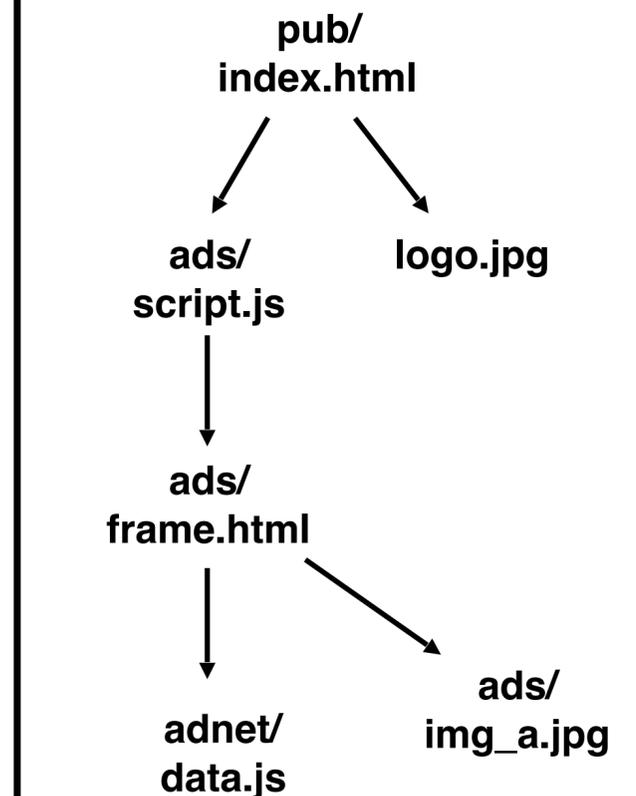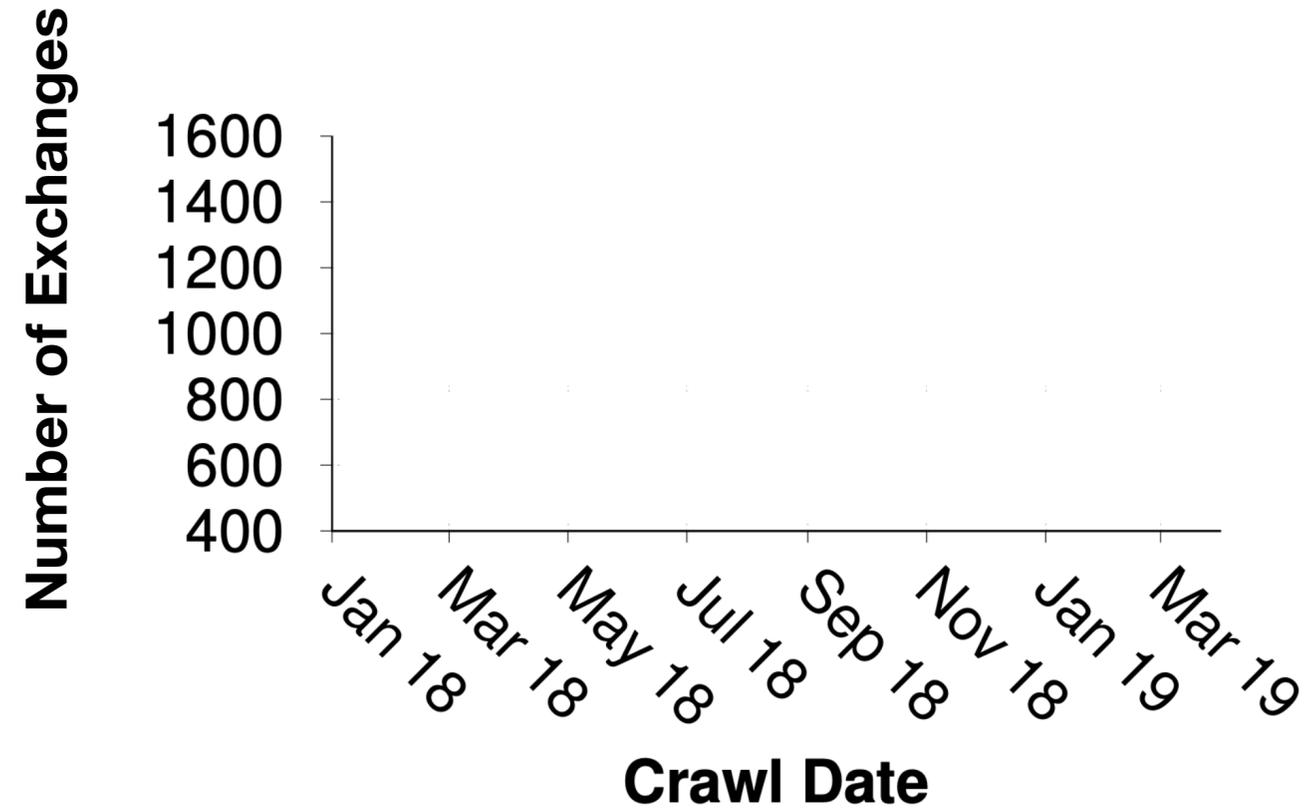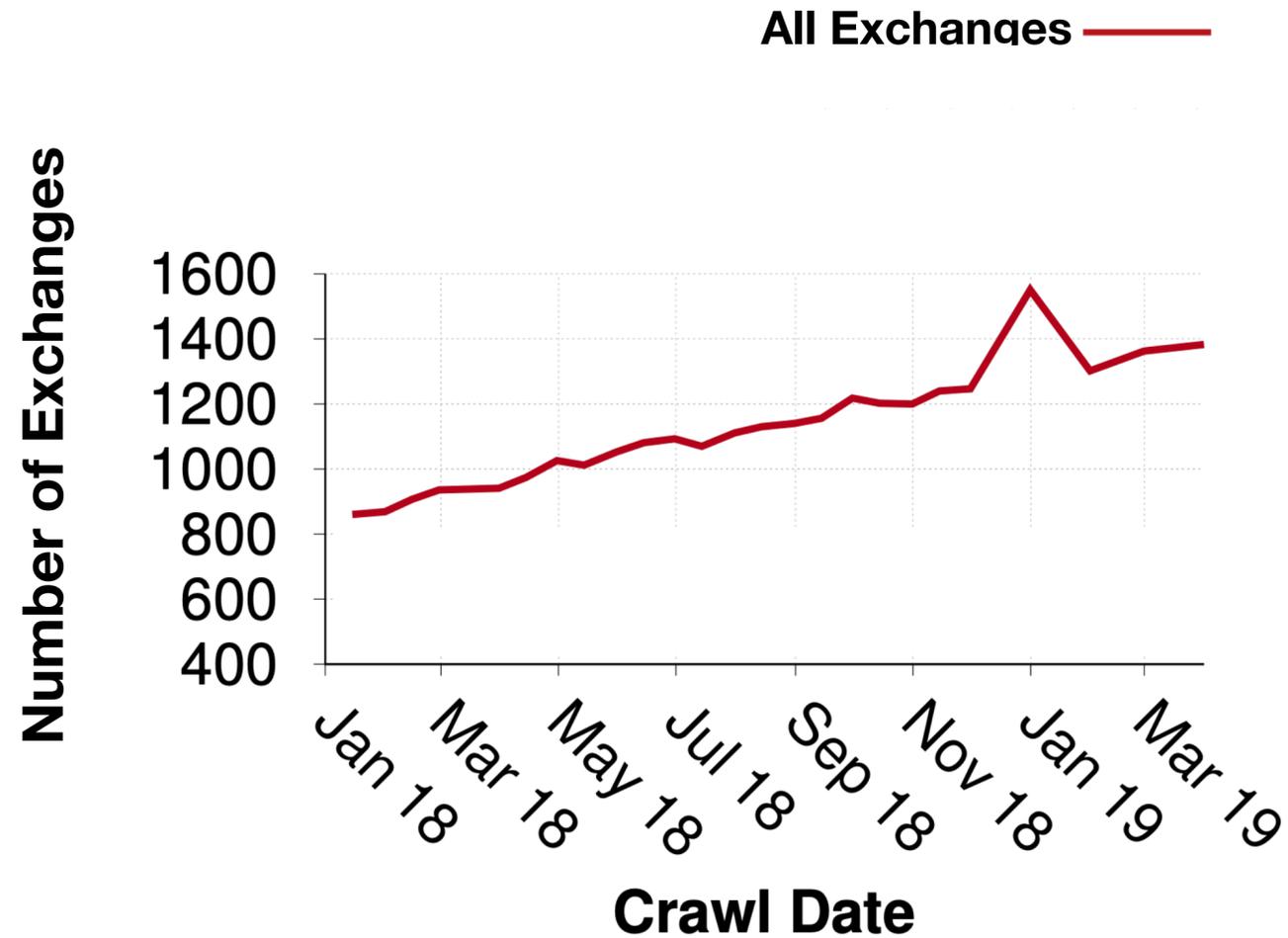openx.com, 183753,RESELLER

Example Inclusions

pub/
index.html

ads/
script.js

logo.jpg

ads/
frame.html

adnet/
data.js

ads/
img_a.jpg

Authorized Exchanges Over Time

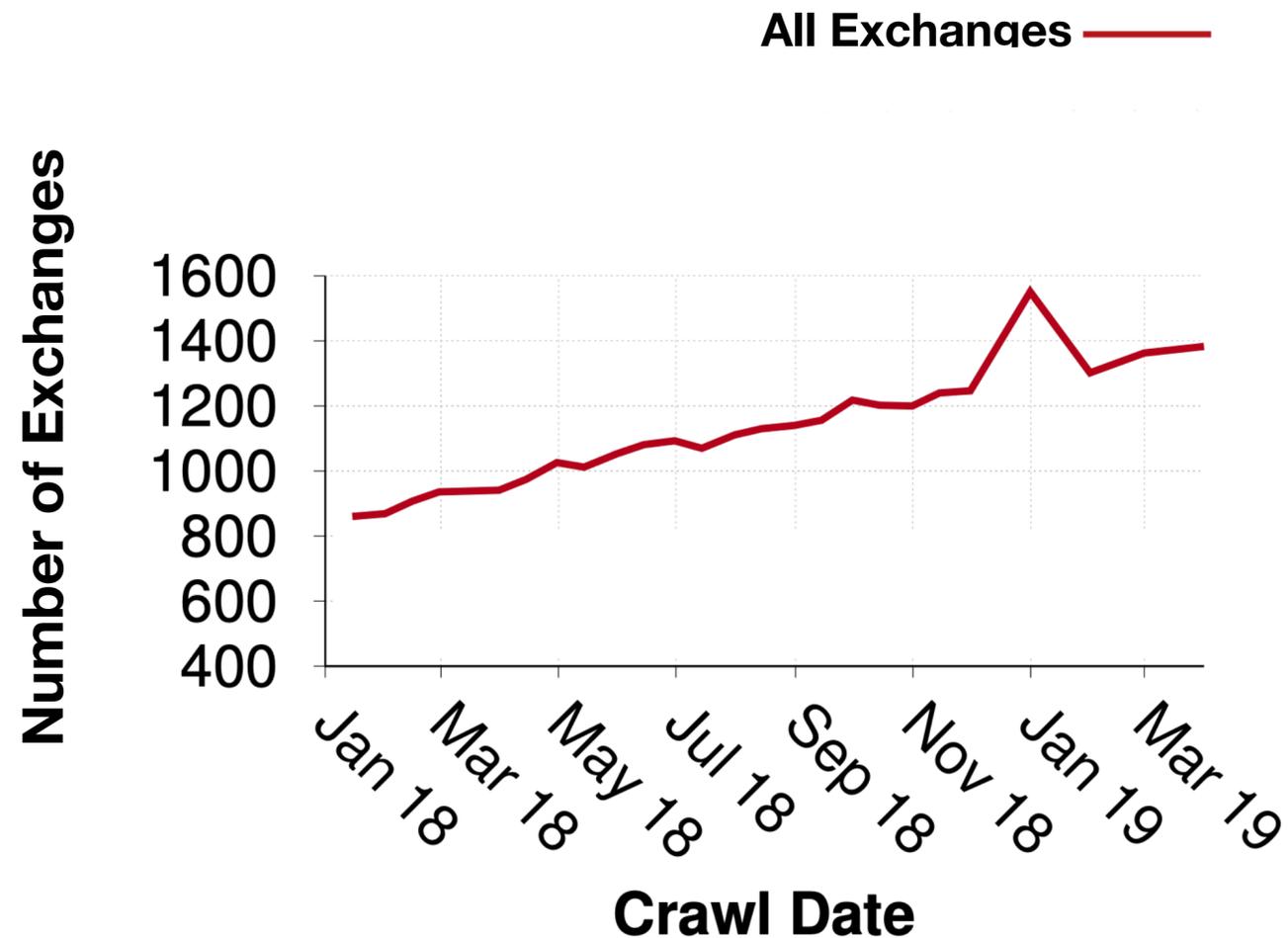Number of authorized exchange domains over time

Authorized Exchanges Over Time

All Exchanges ————

Increase: 860 —> 1400
Total unique exchanges: 2381

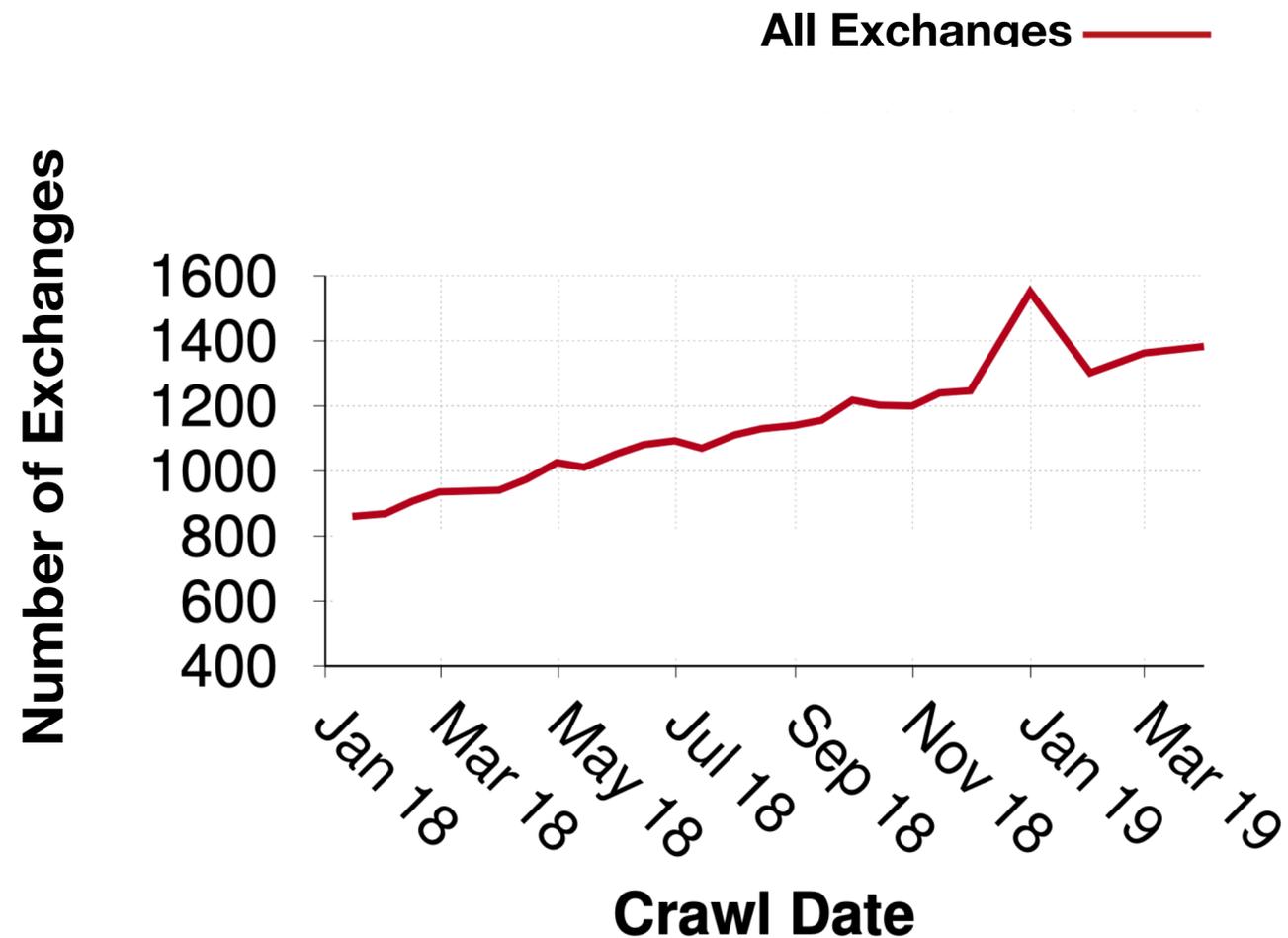Number of authorized exchange domains over time

Authorized Exchanges Over Time

All Exchanges ———

Increase: 860 —> 1400
Total unique exchanges: 2381

Mistakes in ads.txt files

Number of authorized exchange domains over time

Authorized Exchanges Over Time

All Exchanges ———

Number of authorized exchange domains over time

Increase: 860 —> 1400
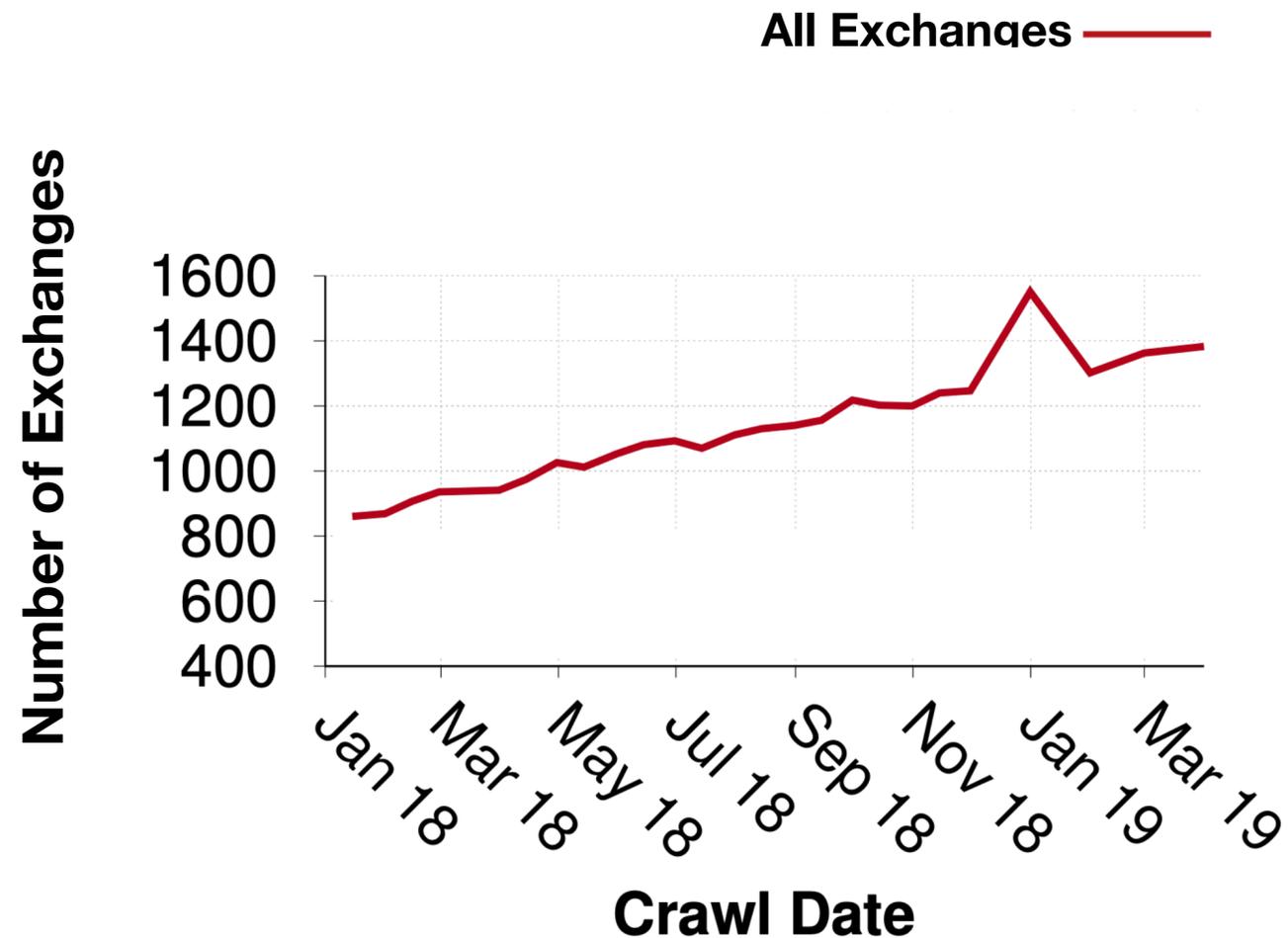**Total unique exchanges: 2381**

Mistakes in ads.txt files

A. Syntax Errors

google.com pub−7439281311086140, DIRECT

openx.com , pub−7439281311086140

Authorized Exchanges Over Time

All Exchanges ———

Increase: 860 —> 1400
Total unique exchanges: 2381

Mistakes in ads.txt files

Number of Exchanges (y-axis): 400, 600, 800, 1000, 1200, 1400, 1600

Crawl Date (x-axis): Jan 18, Mar 18, May 18, Jul 18, Sep 18, Nov 18, Jan 19, Mar 19

Number of authorized exchange domains over time

A. Syntax Errors

google.com pub−7439281311086140, DIRECT

openx.com , pub−7439281311086140

B. Semantic Errors

Authorized Exchanges Over Time

All Exchanges ——

Number of authorized exchange domains over time

Increase: 860 —> 1400
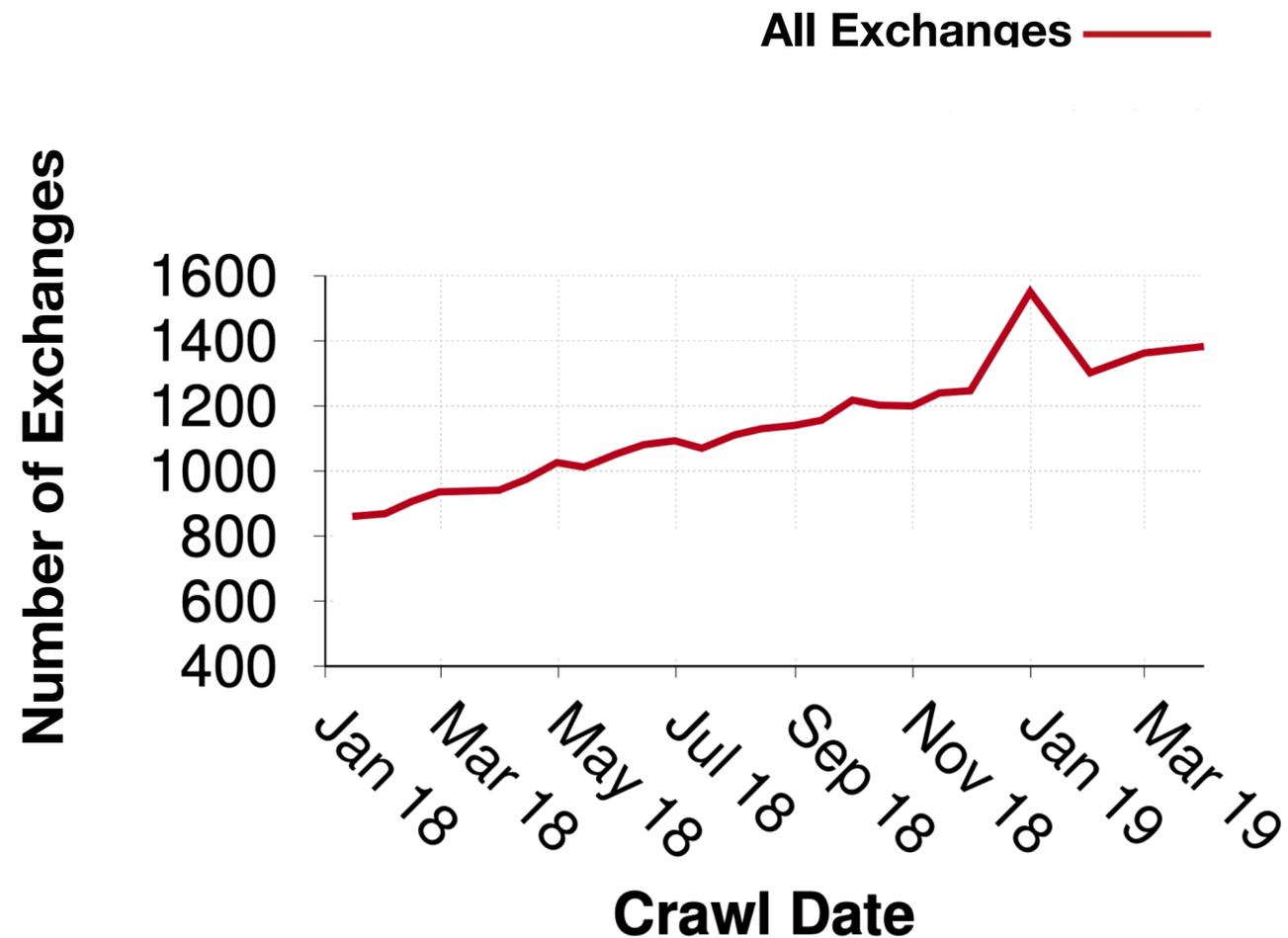Total unique exchanges: 2381

Mistakes in ads.txt files

A. Syntax Errors
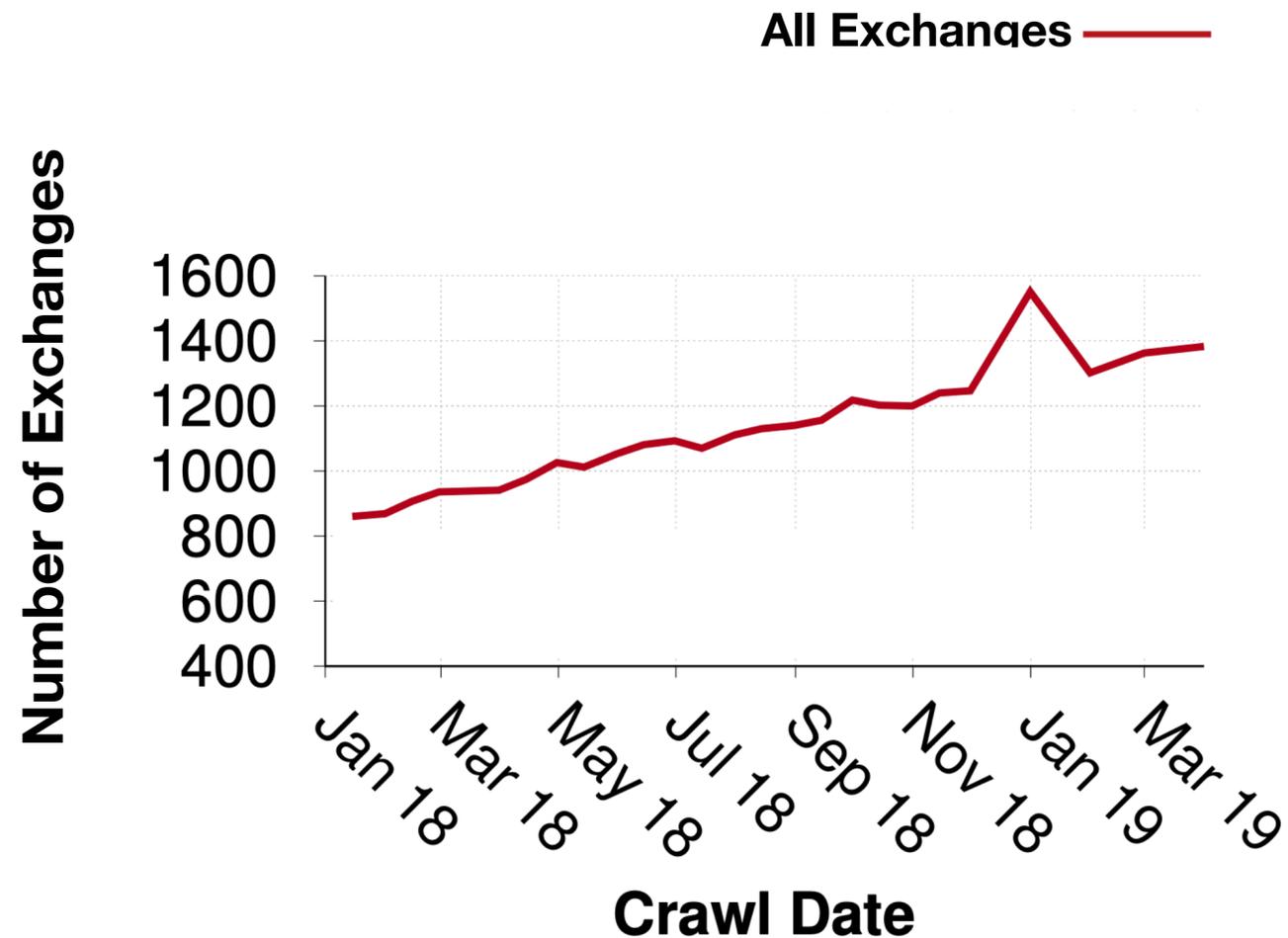
google.com pub−7439281311086140, DIRECT

openx.com , pub−7439281311086140

B. Semantic Errors

rubicnprject.com, 11078, DIRECT ← Misspelled

Authorized Exchanges Over Time

All Exchanges ———

Number of authorized exchange domains over time

Increase: 860 —> 1400
Total unique exchanges: 2381

Mistakes in ads.txt files

A. Syntax Errors

google.com pub−7439281311086140, DIRECT

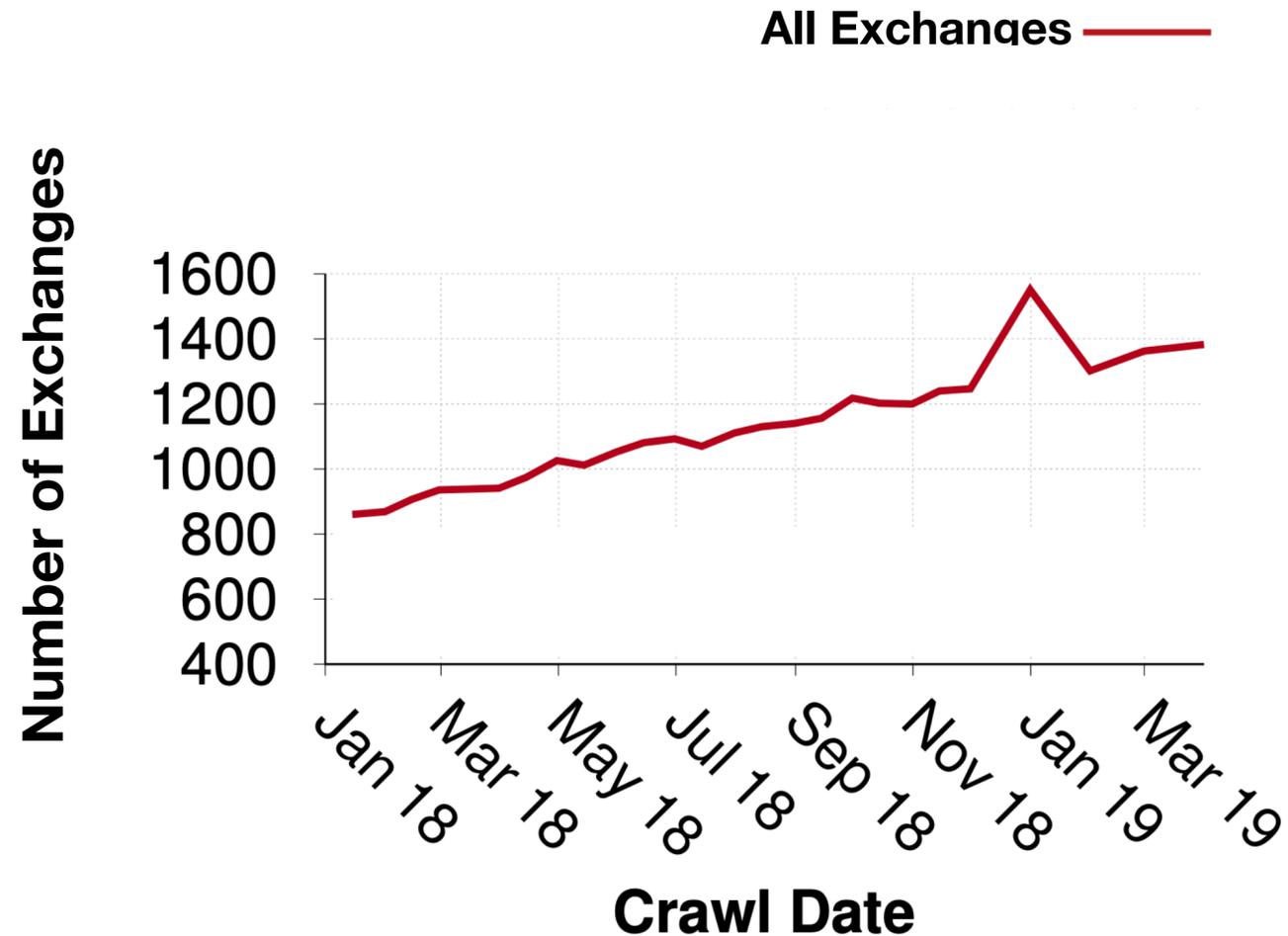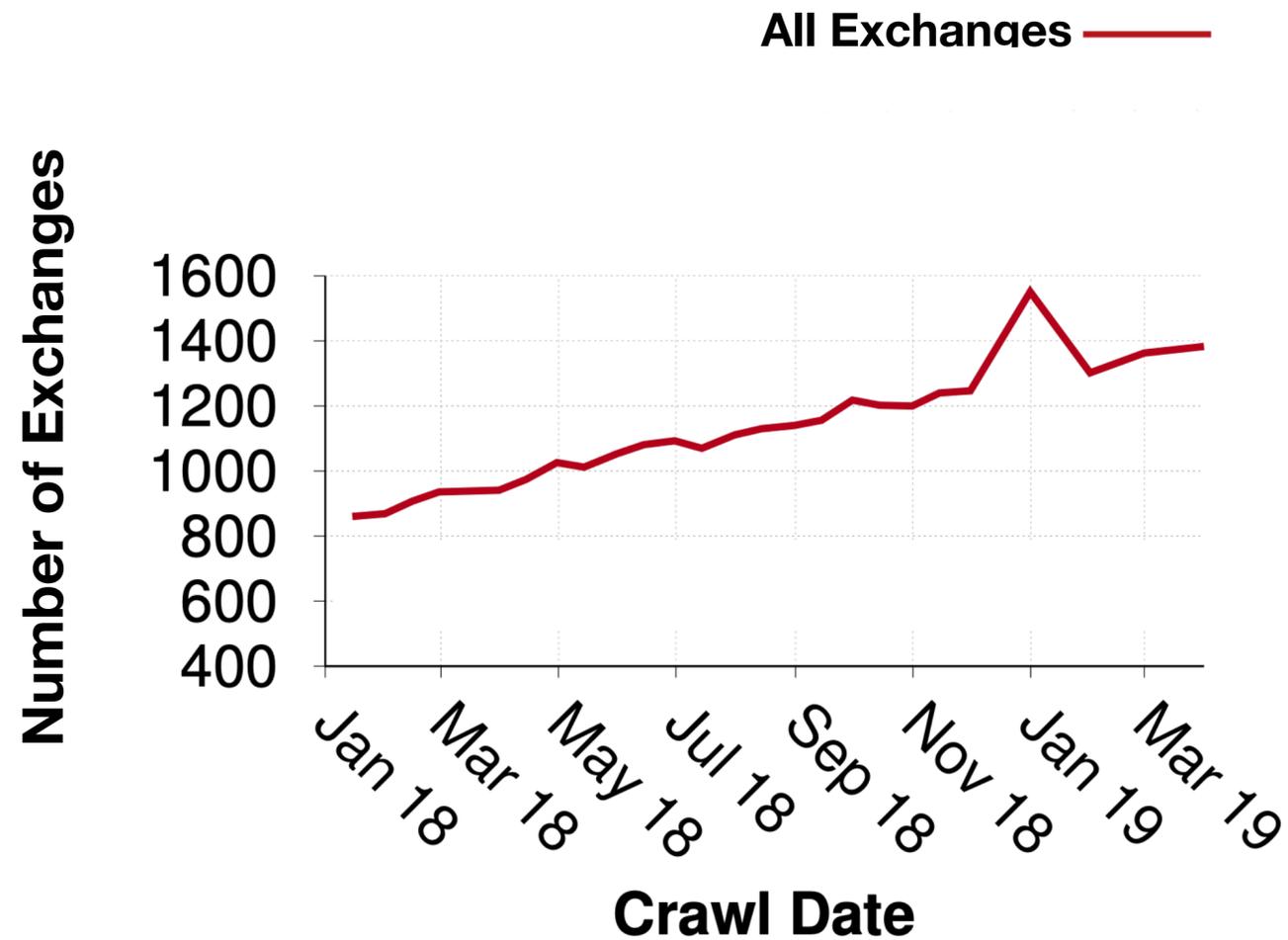**openx.com ,** pub−7439281311086140

B. Semantic Errors

rubicnprject.com, 11078, DIRECT ← Misspelled

doubleclick.net, 3159, DIRECT

Should be google.com

Authorized Exchanges Over Time

All Exchanges ———

Number of authorized exchange domains over time

Increase: 860 —> 1400
Total unique exchanges: 2381

Mistakes in ads.txt files

A. Syntax Errors

google.com pub−7439281311086140, DIRECT

openx.com **,** pub−7439281311086140

B. Semantic Errors

rubicnprject.com, 11078, DIRECT ◀ **Misspelled**

doubleclick.net, 3159, DIRECT

Should be google.com

Fixing Semantic Errors

All Exchanges ——

Number of authorized exchange domains over time

(Chart: Number of Exchanges vs Crawl Date, showing a rising line from ~860 in Jan 18 to ~1400 by Mar 19, with a peak of ~1560 around Jan 19)
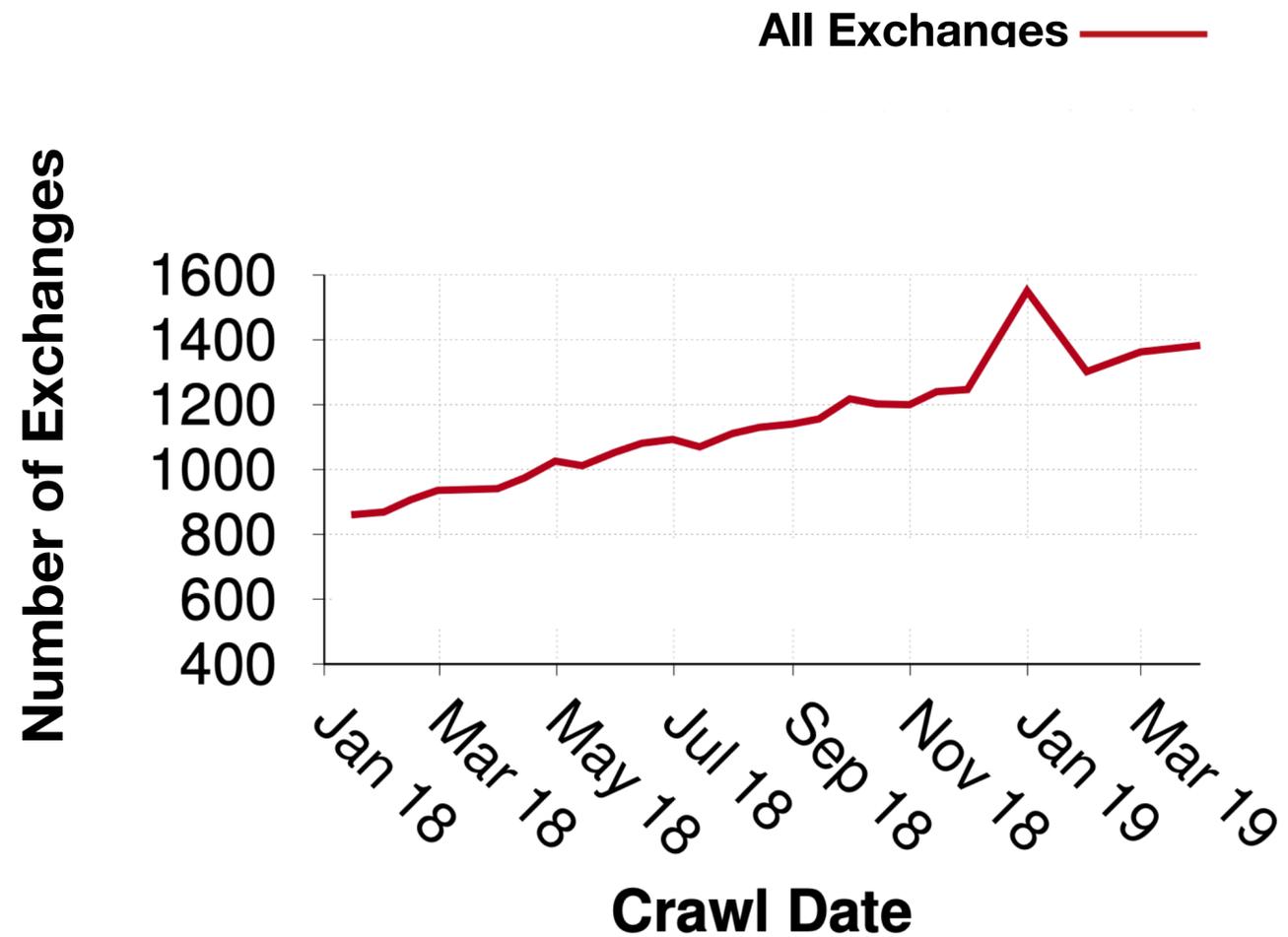
rubicnprject.com, 11078, DIRECT

Misspelled

doubleclick.net, 3159, DIRECT

Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

Fixing Semantic Errors

All Exchanges ⎯⎯⎯

Number of authorized exchange domains over time
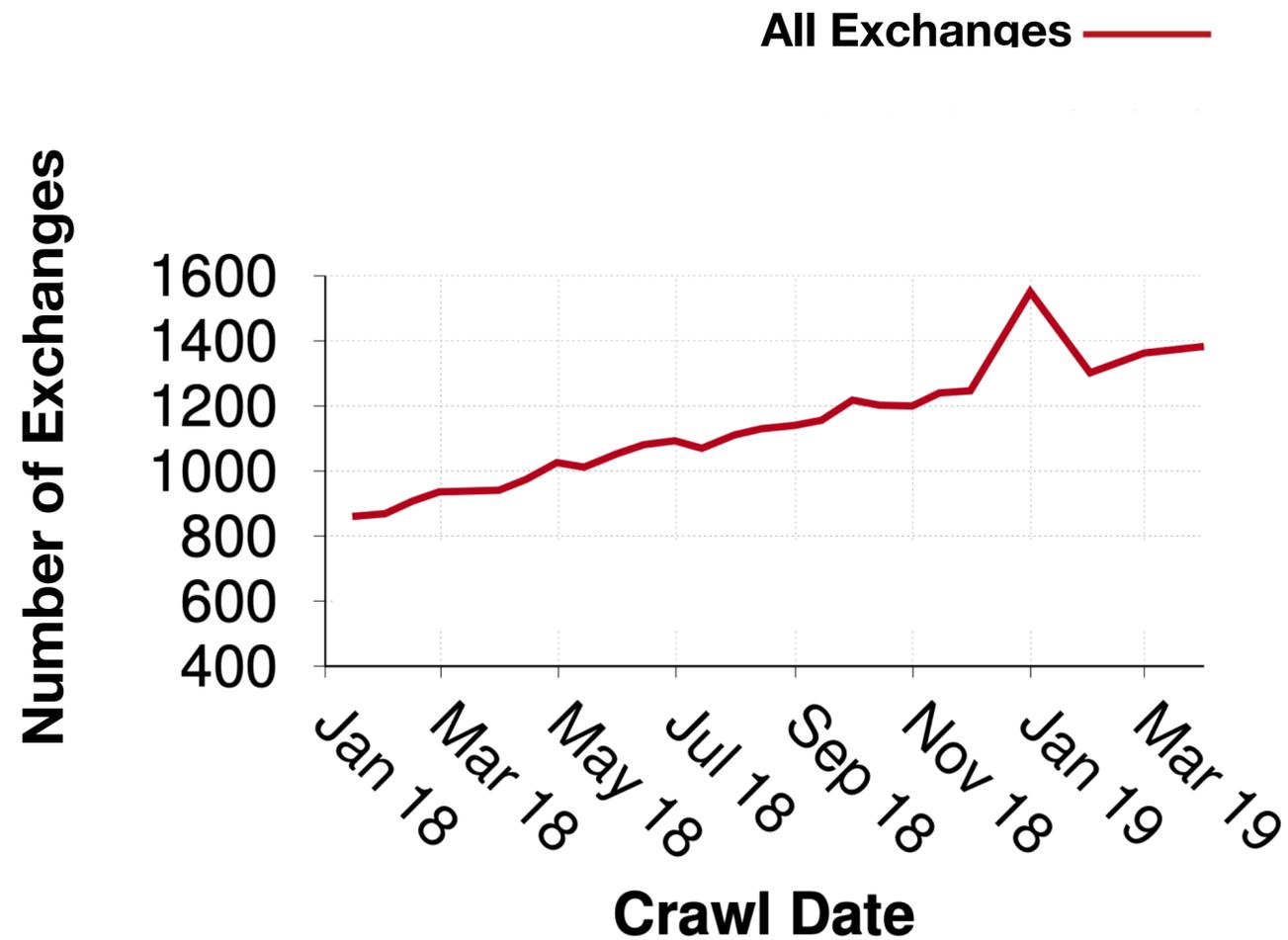
rubicnprject.com, 11078, DIRECT

Misspelled

doubleclick.net, 3159, DIRECT

Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381
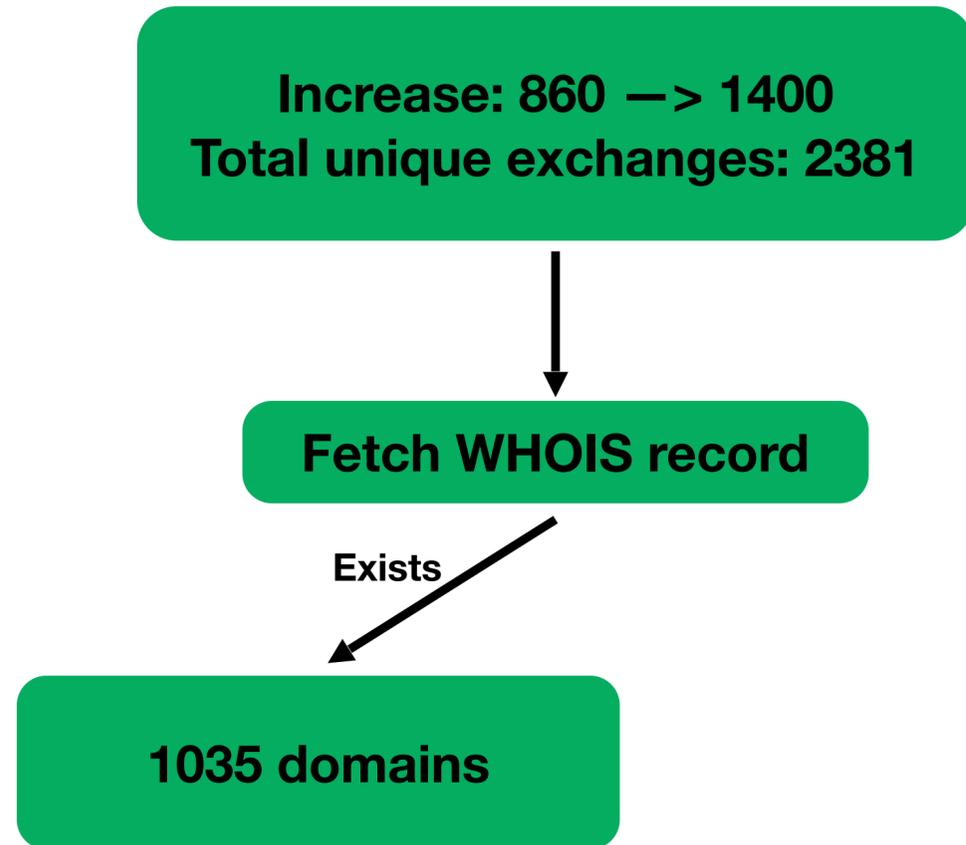
Fetch WHOIS record

Fixing Semantic Errors

Number of authorized exchange domains over time
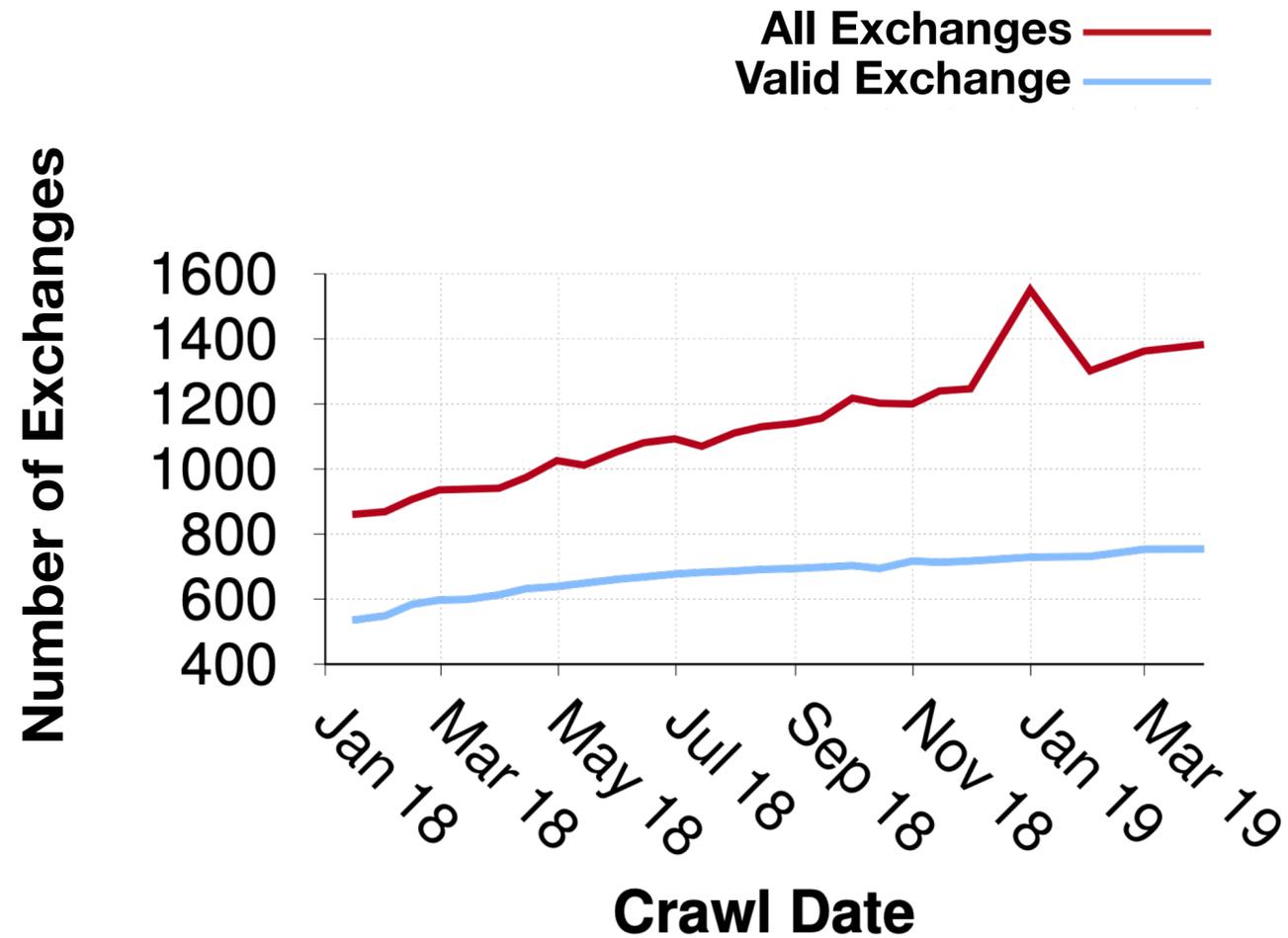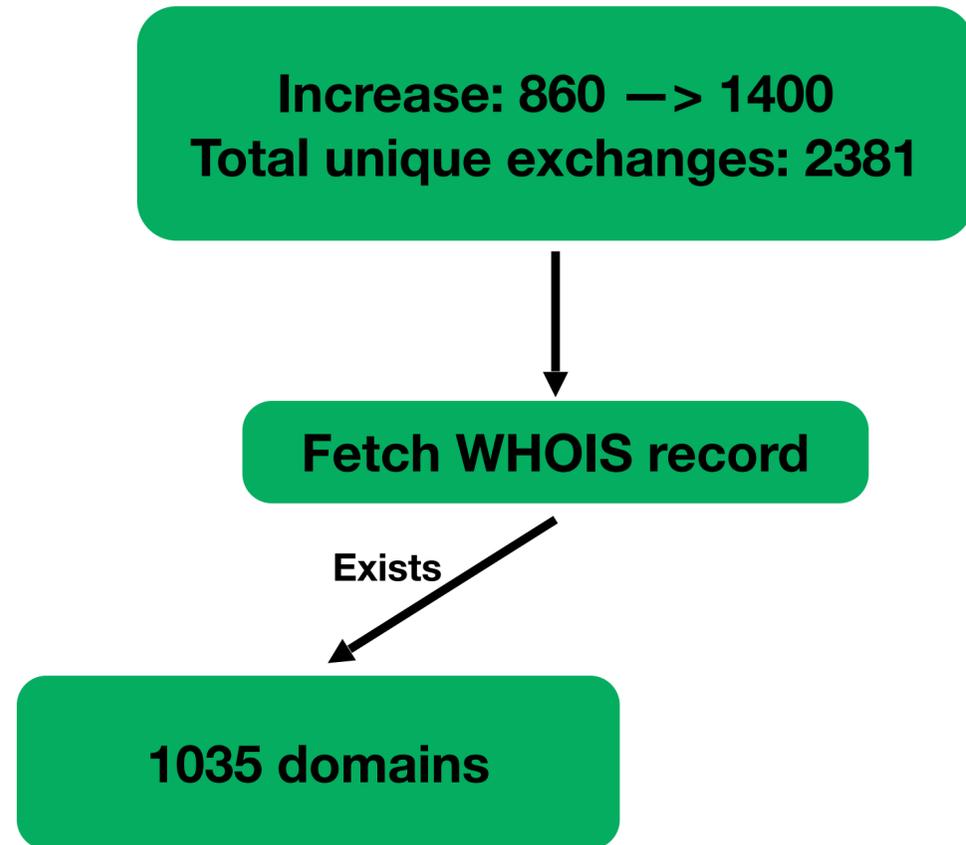
rubicnprject.com, 11078, DIRECT

Misspelled

doubleclick.net, 3159, DIRECT

Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

Fetch WHOIS record

Exists

1035 domains

Fixing Semantic Errors

All Exchanges
Valid Exchange
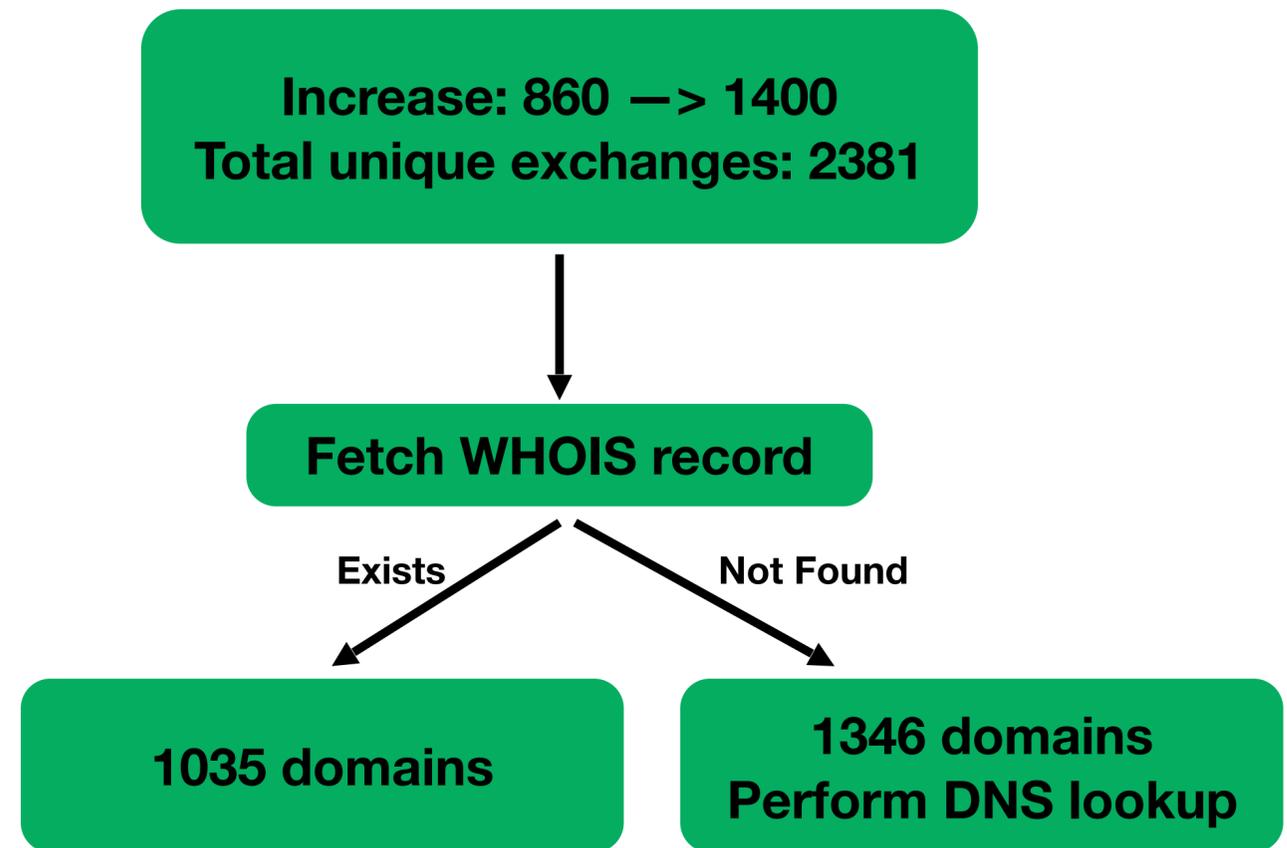
Number of Exchanges vs Crawl Date

Number of authorized exchange domains over time

rubicnprject.com, 11078, DIRECT

Misspelled

doubleclick.net, 3159, DIRECT

Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

Fetch WHOIS record

Exists

1035 domains

Fixing Semantic Errors

Number of authorized exchange domains over time

rubicnprject.com, 11078, DIRECT — Misspelled

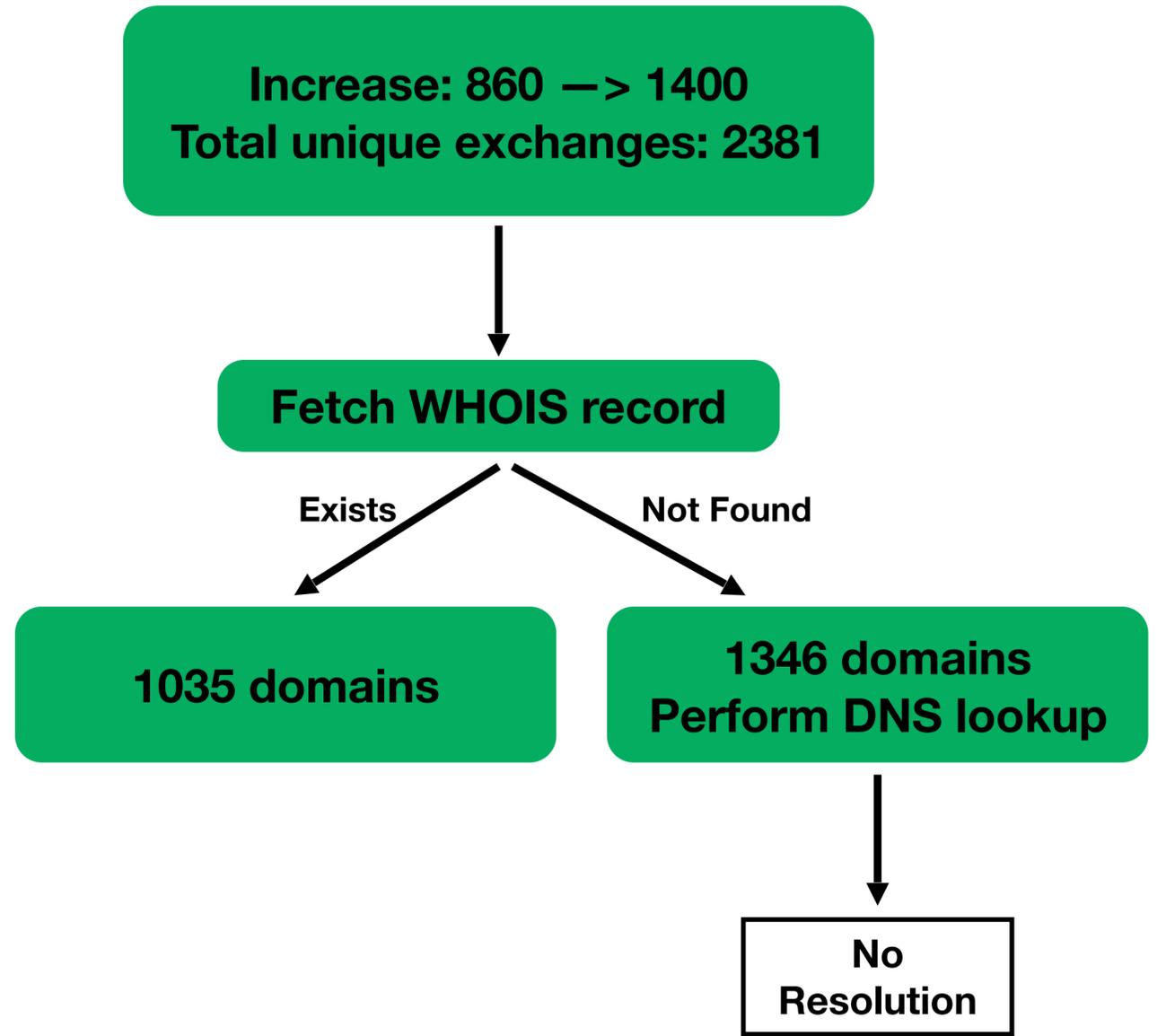**doubleclick.net**, 3159, DIRECT — Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

Fetch WHOIS record

Exists → 1035 domains

Not Found → 1346 domains
Perform DNS lookup

Fixing Semantic Errors

Number of authorized exchange domains over time

rubicnprject.com, 11078, DIRECT → Misspelled

doubleclick.net, 3159, DIRECT → Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

↓

Fetch WHOIS record

Exists ← → Not Found

1035 domains

1346 domains
Perform DNS lookup
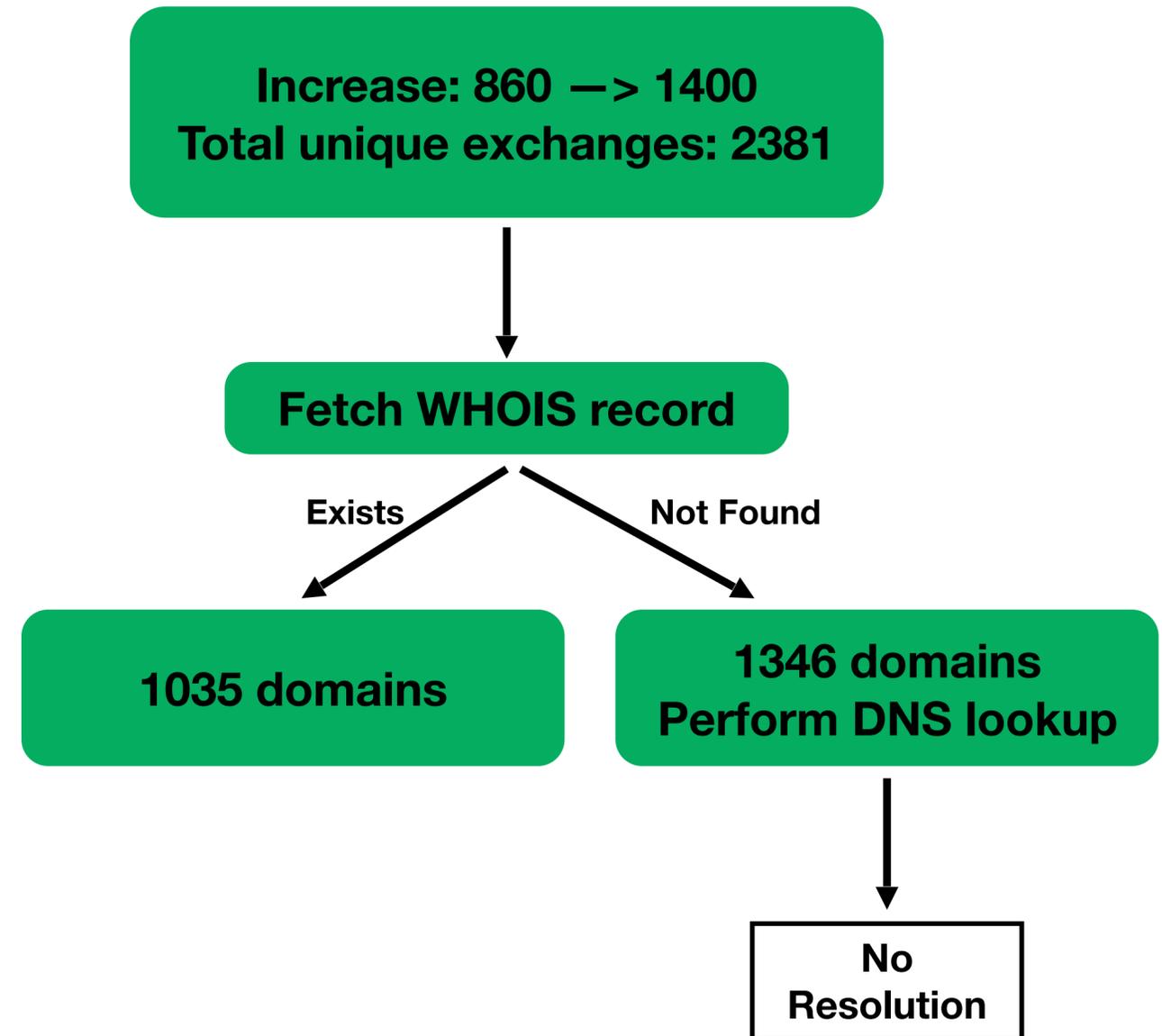
↓

No Resolution

Fixing Semantic Errors

Number of authorized exchange domains over time

rubicnprject.com, 11078, DIRECT — Misspelled

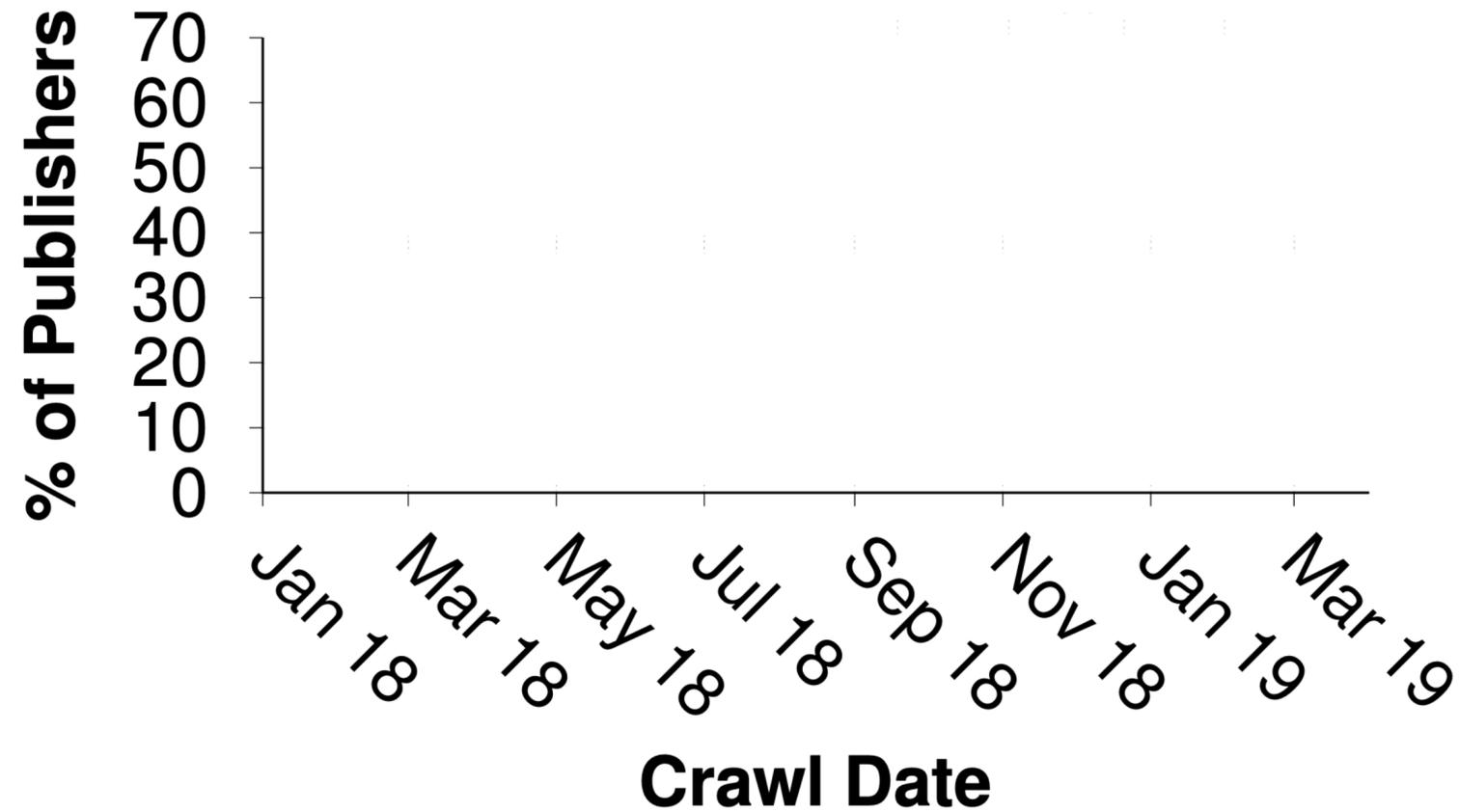**doubleclick.net**, 3159, DIRECT — Should be google.com

Increase: 860 —> 1400
Total unique exchanges: 2381

Fetch WHOIS record

Exists → 1035 domains

Not Found → 1346 domains / Perform DNS lookup → No Resolution
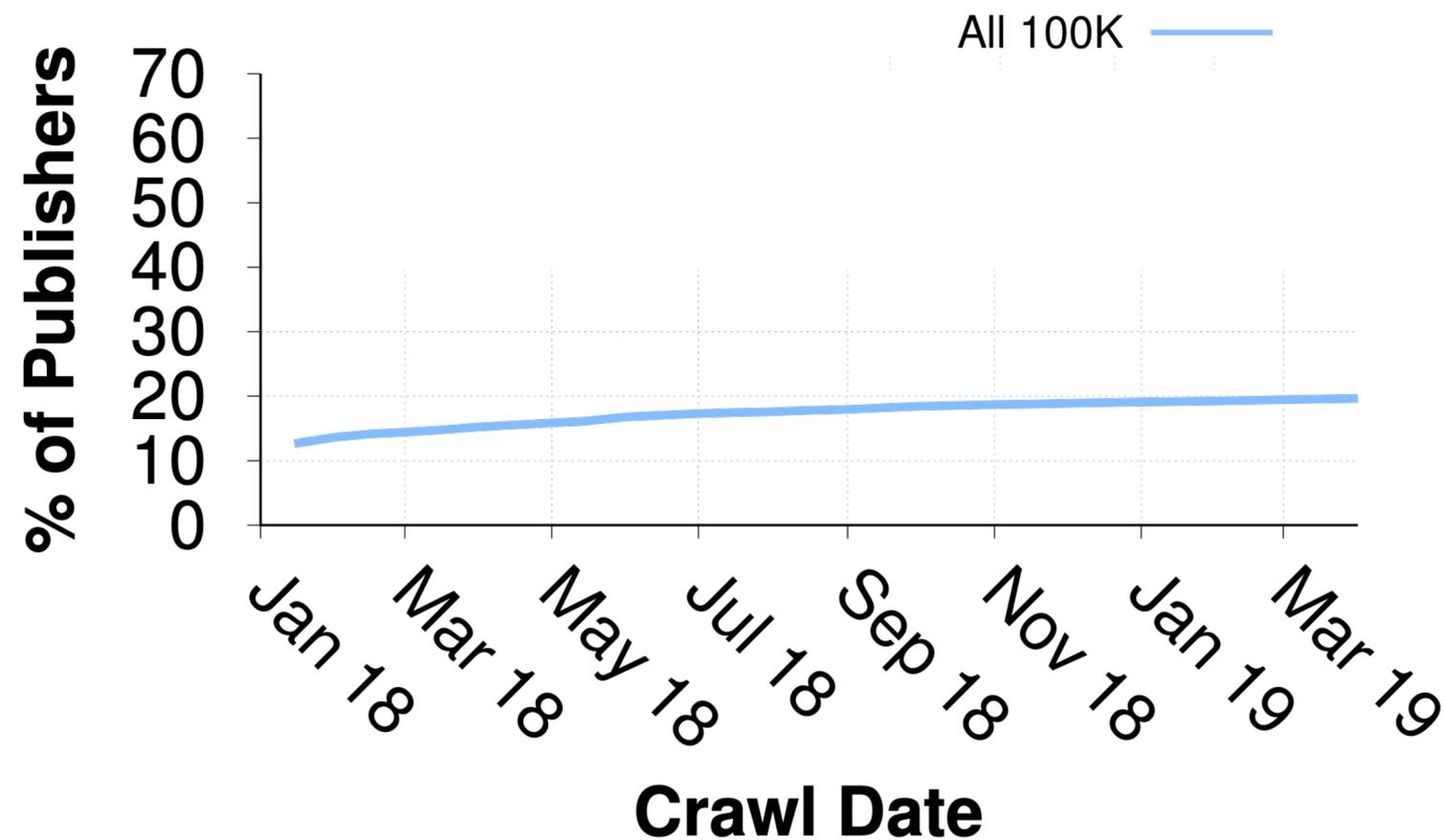
Talk Overview

1. RTB background & motivation for the ads.txt standard

2. Data collection

3. Adoption trends

4. Compliance of the ads.txt standard

5. Means of providing transparency

How Have Publishers Adopted?

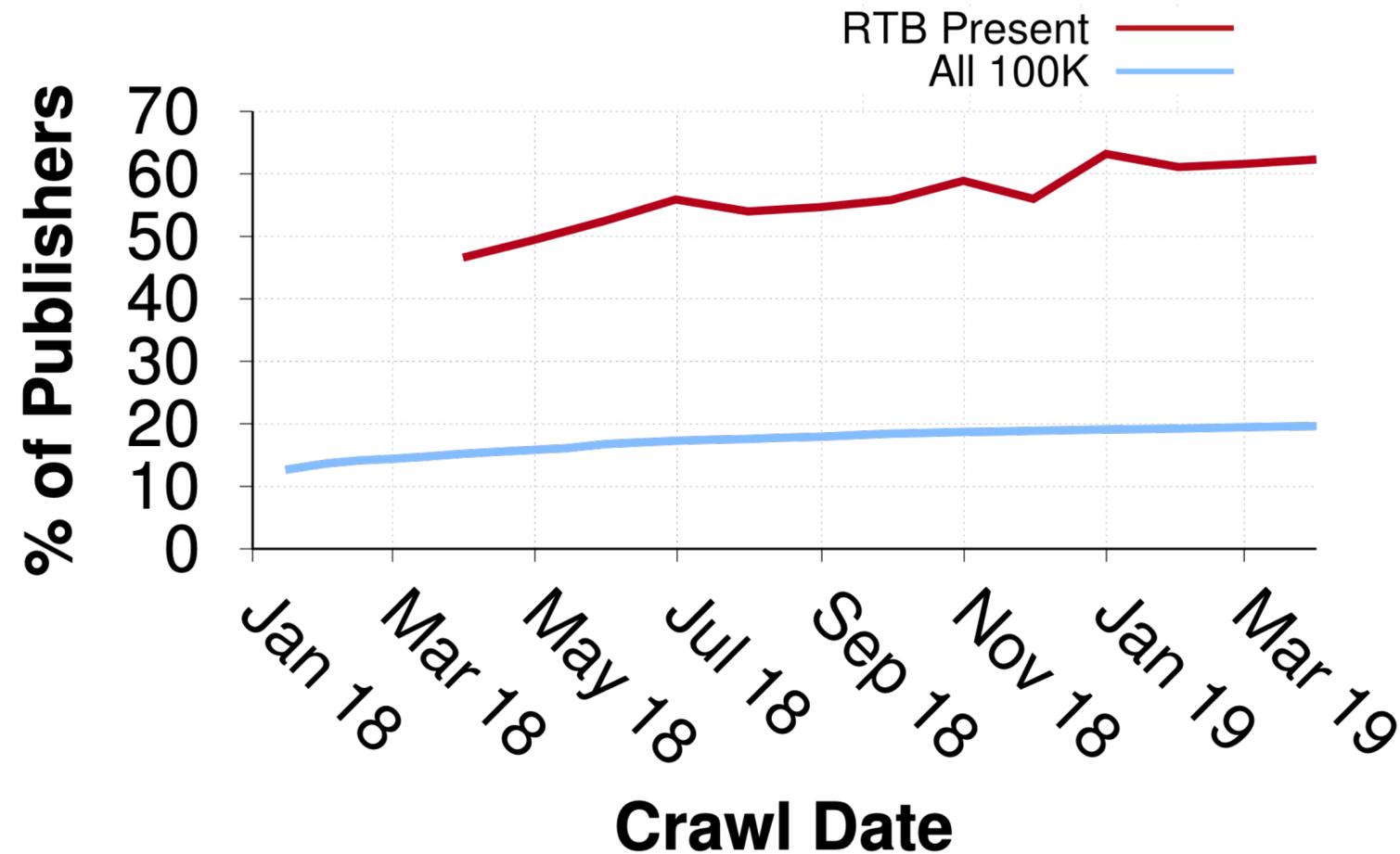ads.txt adoption by Alexa Top-100K
publishers over time

How Have Publishers Adopted?

All 100K

% of Publishers (y-axis: 0, 10, 20, 30, 40, 50, 60, 70)

Crawl Date (x-axis: Jan 18, Mar 18, May 18, Jul 18, Sep 18, Nov 18, Jan 19, Mar 19)

ads.txt adoption by Alexa Top-100K publishers over time

Adoption has been modest
Increase: 12% —> 20%

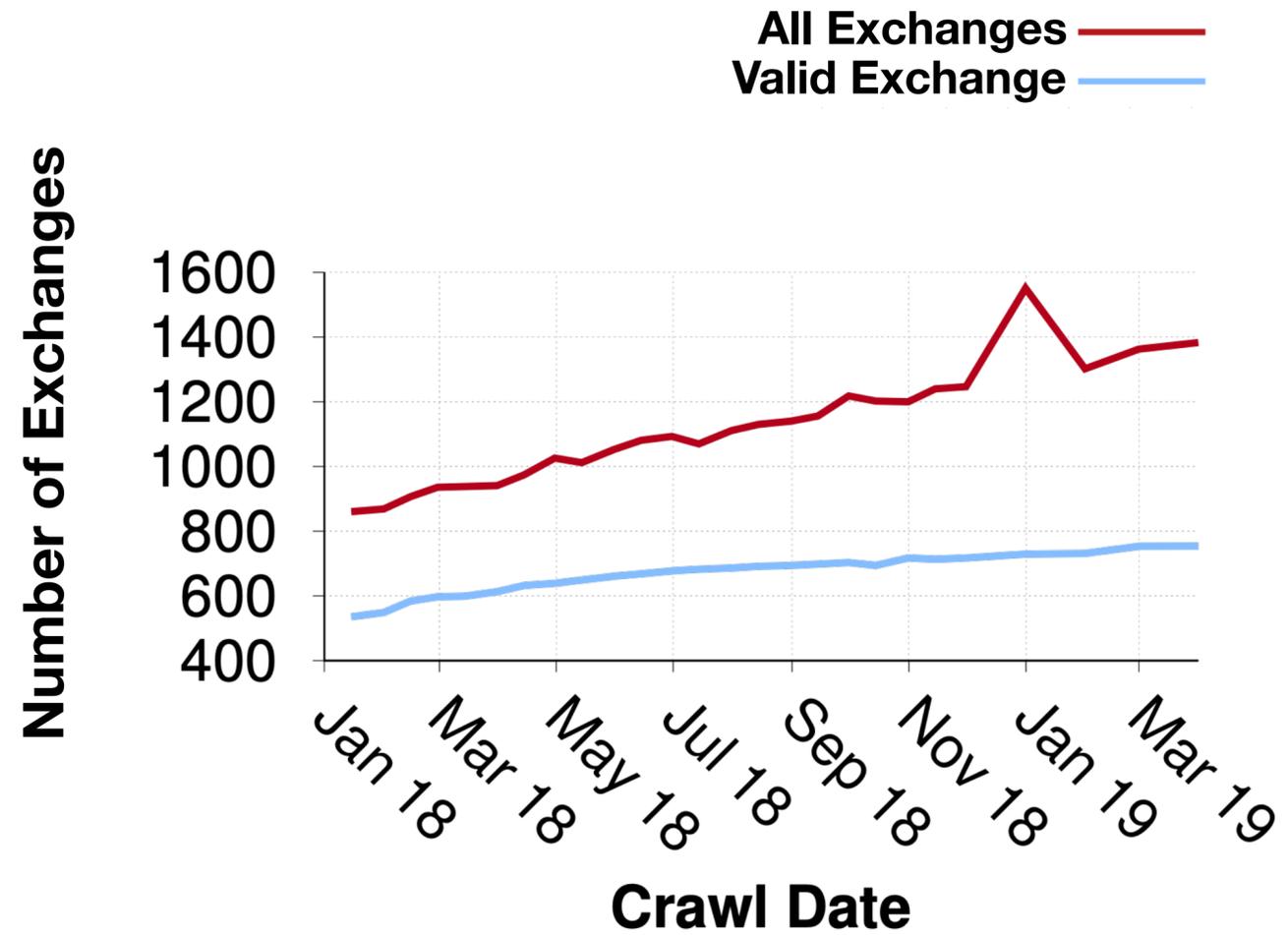# How Have Publishers Adopted?

ads.txt adoption by Alexa Top-100K
publishers over time

**Adoption has been modest
Increase: 12% —> 20%**

**Publishers that run RTB ads
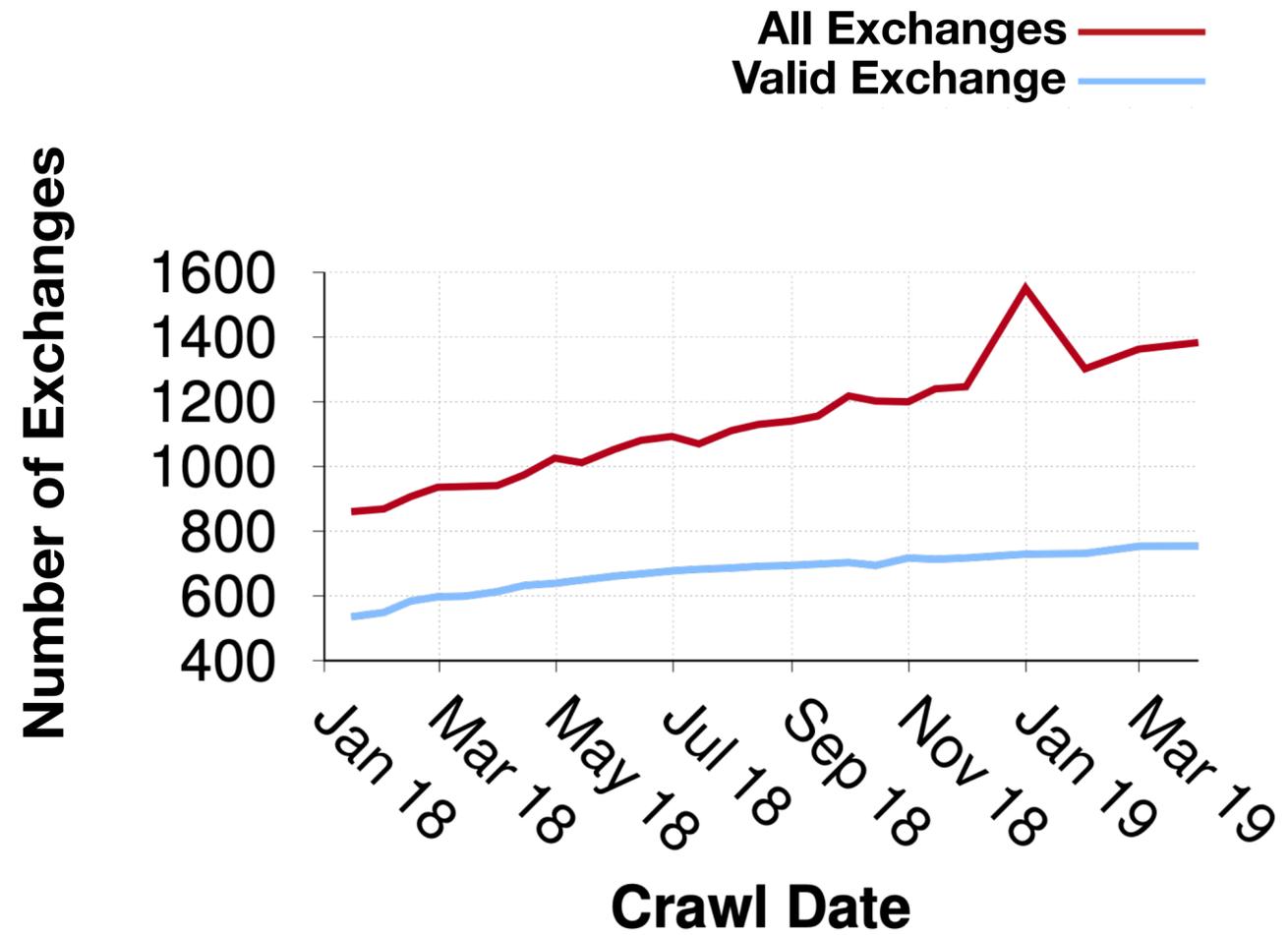Adoption is pretty decent. ~64%**

Authorized Exchanges Discovered

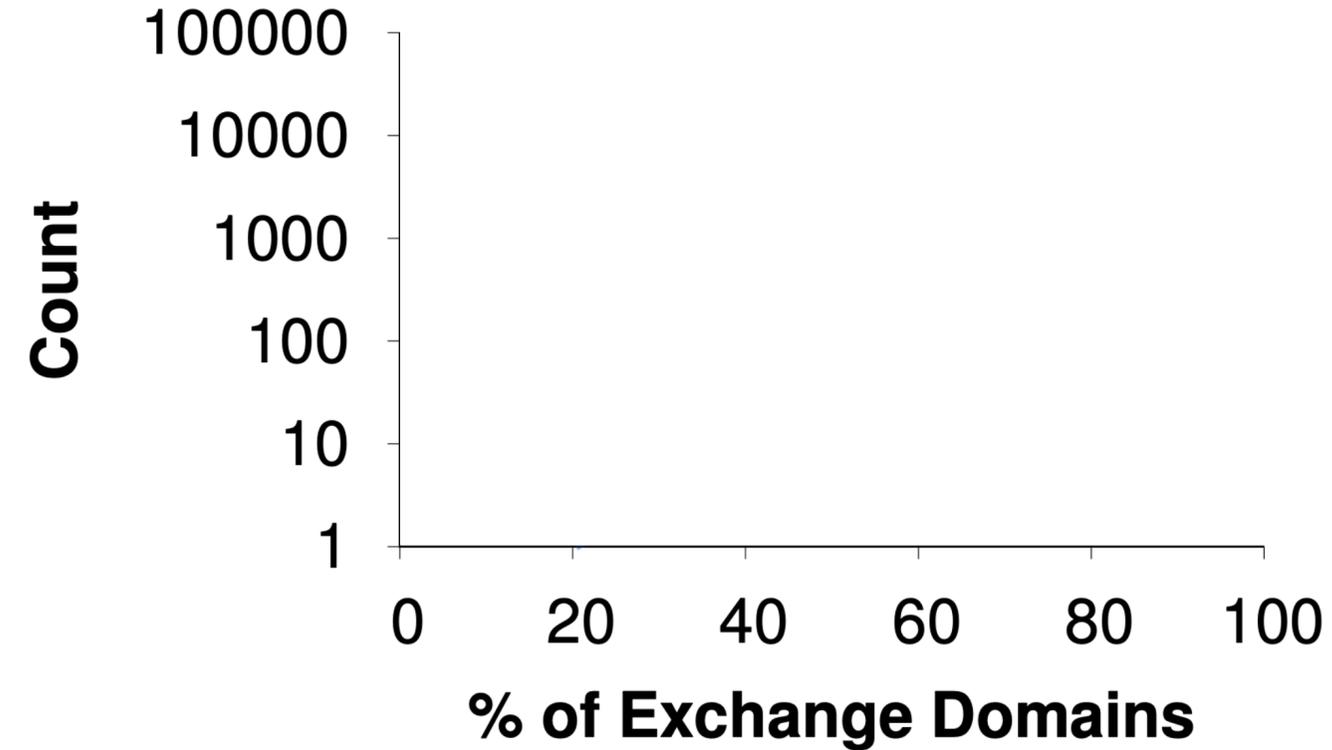Number of authorized exchange domains over time

1035 valid domains
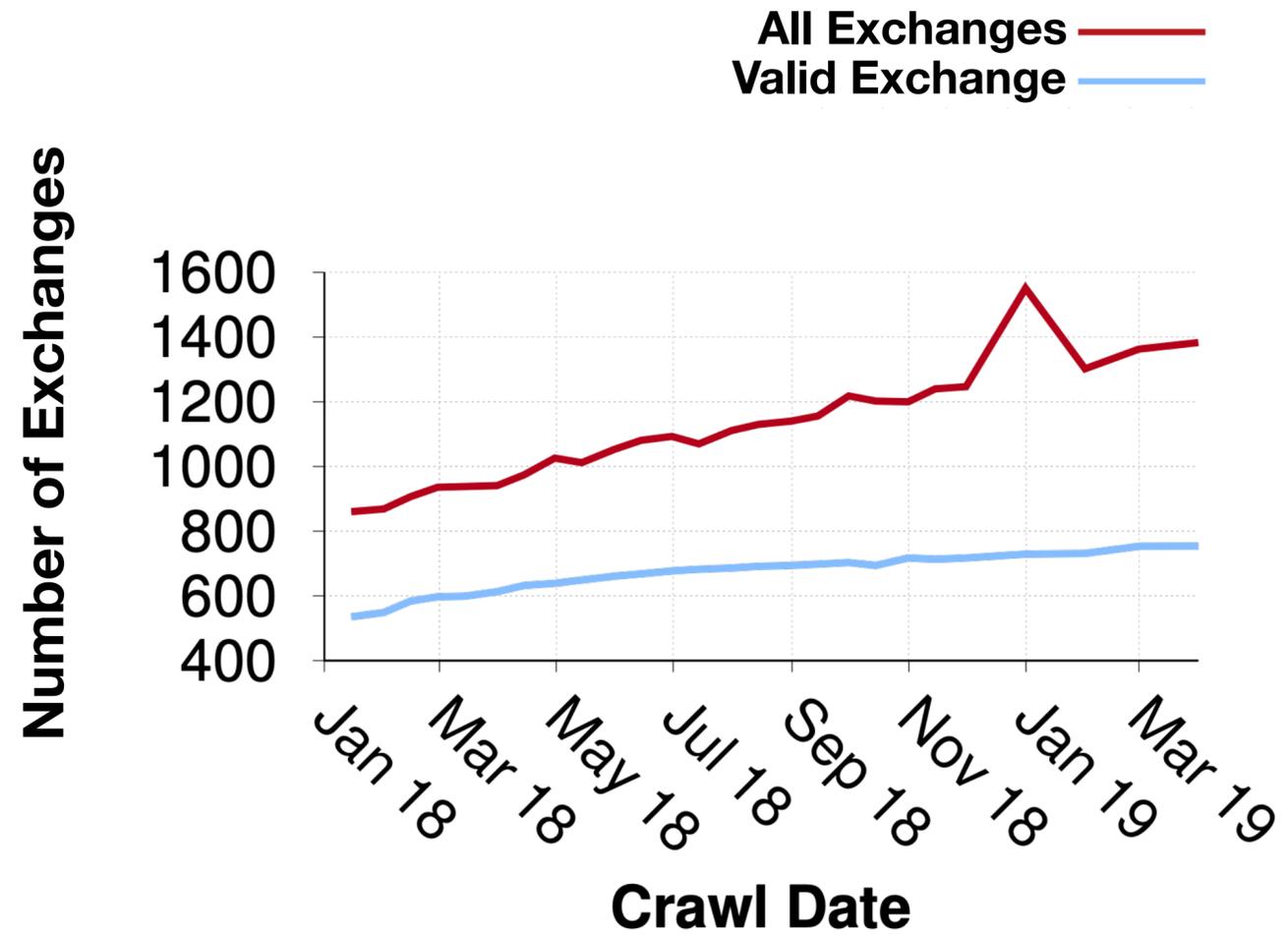
Authorized Exchanges Discovered

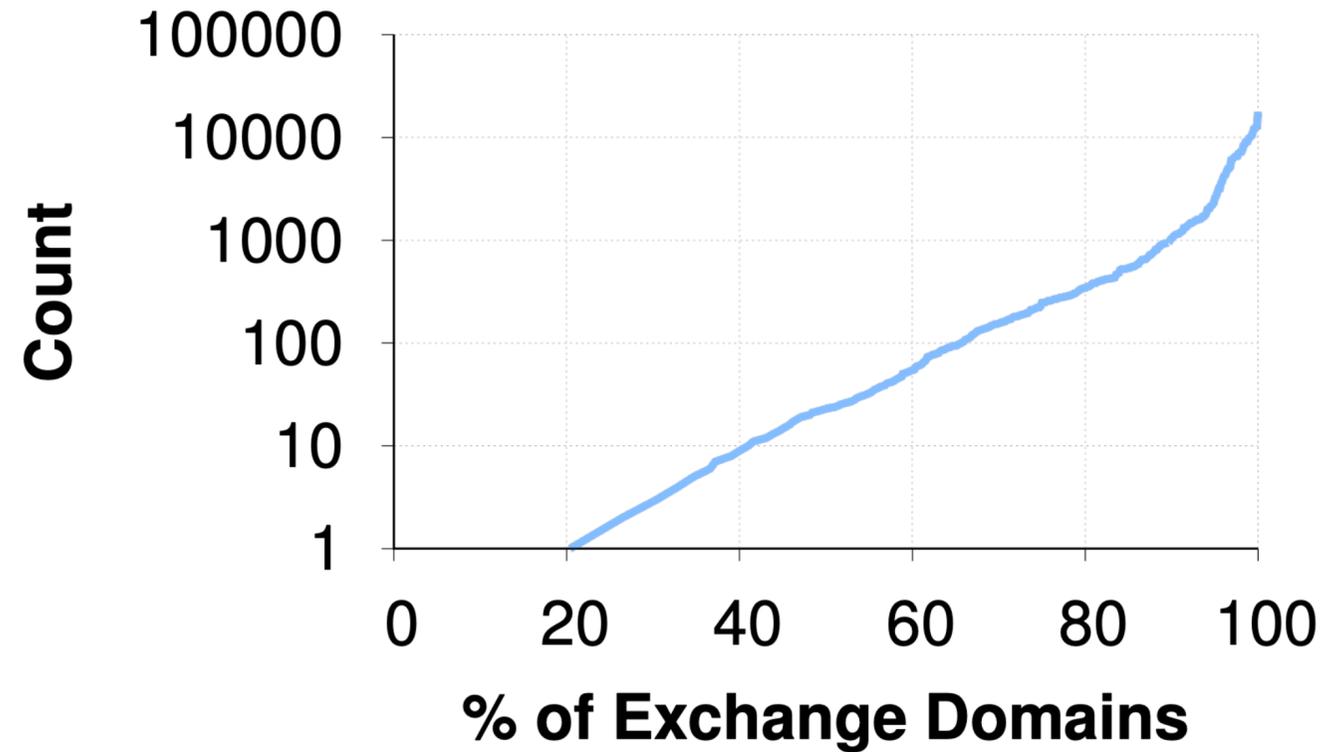Number of authorized exchange domains over time

1035 valid domains

Number of unique publishers per authorized exchange domain

Authorized Exchanges Discovered

All Exchanges
Valid Exchange

Number of authorized exchange domains over time

1035 valid domains

Number of unique publishers per authorized exchange domain

Out of 1035 domains, **20%** appear only on a **single ads.txt** file

Adoption Trends are Encouraging

- 64% of the publishers from Alexa top-100K who run RTB ads.

- Major A&A domains support ads.txt

Adoption Trends are Encouraging

- 64% of the publishers from Alexa top-100K who run RTB ads.

- Major A&A domains support ads.txt

Talk Overview

1. RTB background & motivation for the ads.txt standard

2. Data collection

3. Adoption trends

4. Compliance of the ads.txt standard

5. Means of providing transparency

A Simple Check for Compliance

Observable via browser

Not observable

Publisher

Ad Exchange

DSP (Bidder)

A Simple Check for Compliance

A Simple Check for Compliance

Observable via browser

Not observable

Publisher

Ad Exchange

DSP (Bidder)

https://cnn.com/ads.txt

google.com, pub−787, DIRECT

rubiconproject.com, 11078, DIRECT

openx.com, 183753,RESELLER

A Simple Check for Compliance

Publisher   Ad Exchange   DSP (Bidder)

Observable via browser
Not observable

https://cnn.com/ads.txt

google.com, pub−787, DIRECT

rubiconproject.com, 11078, DIRECT

openx.com, 183753, RESELLER

https://cnn.com/ads.txt

google.com, pub−787, DIRECT

rubiconproject.com, 11078, DIRECT

A Simple Check for Compliance

Publisher **Ad Exchange** **DSP (Bidder)**

https://cnn.com/ads.txt

google.com, pub−787, DIRECT

rubiconproject.com, 11078, DIRECT

openx.com, 133753,RESELLER

https://cnn.com/ads.txt
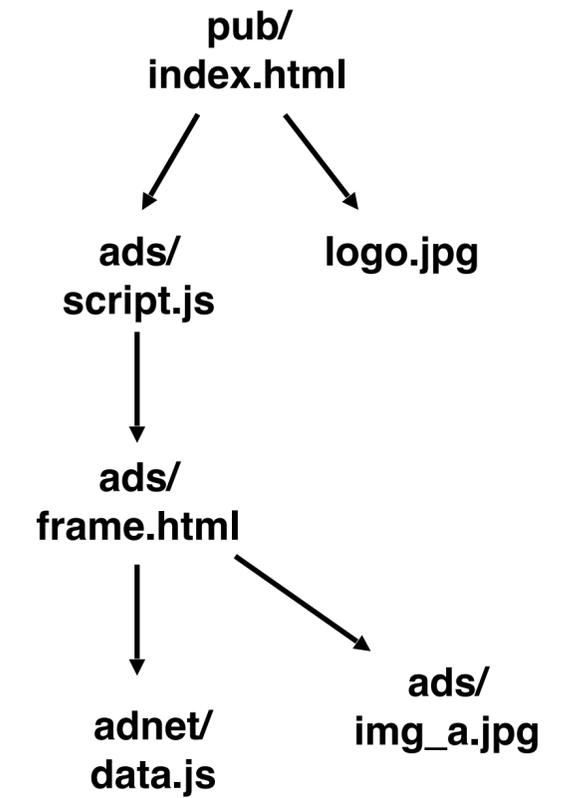
google.com, pub−787, DIRECT

rubiconproject.com, 11078, DIRECT

Checking Compliance at Scale

Example Inclusions

pub/
index.html

ads/
script.js

logo.jpg

ads/
frame.html

adnet/
data.js

ads/
img_a.jpg

Checking Compliance at Scale

Checking Compliance at Scale

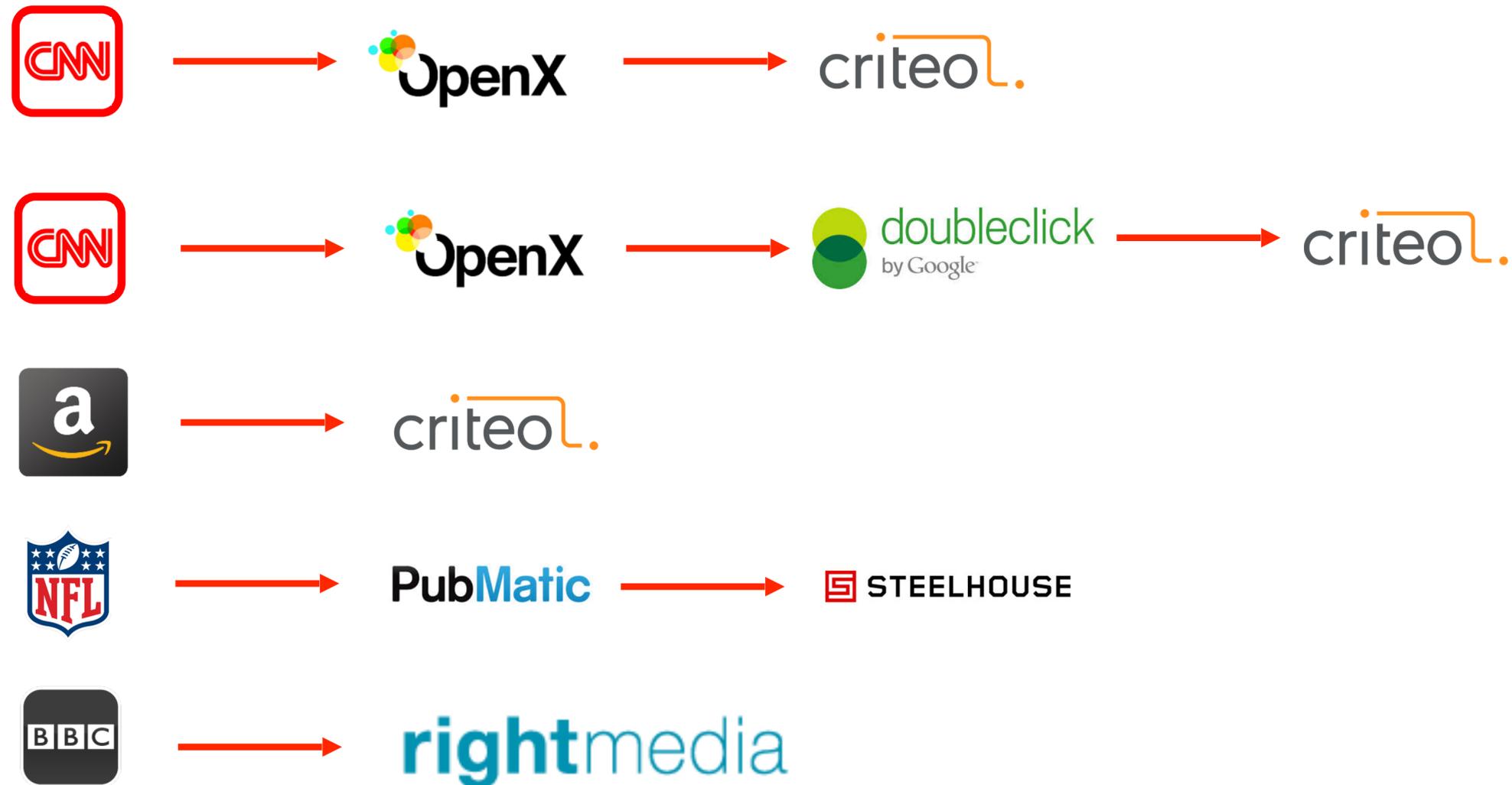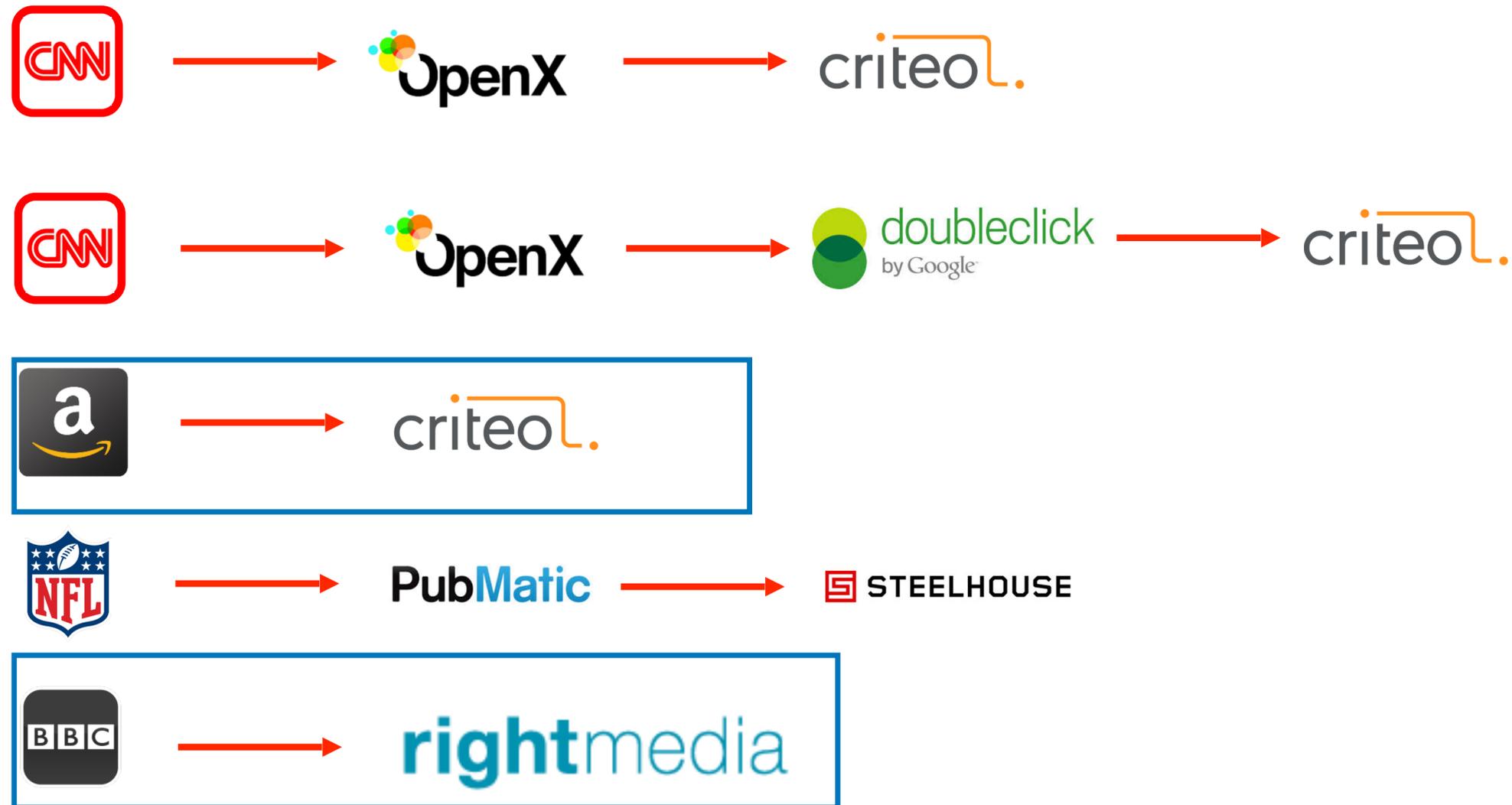# Checking Compliance at Scale

Checking Compliance at Scale

Advertisements
(Detected Using EasyList)
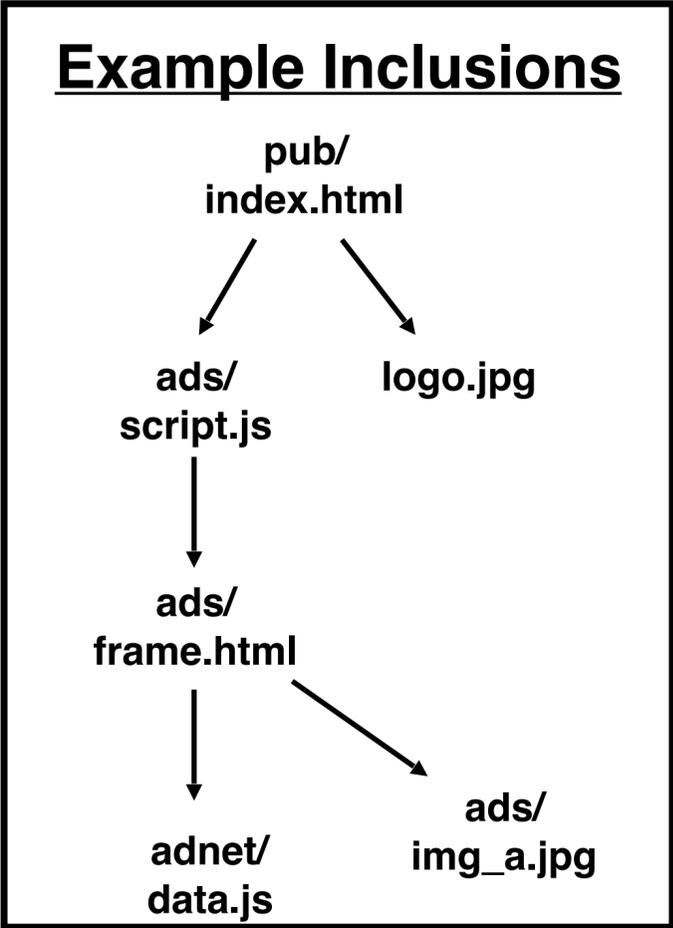
Example Inclusions

pub/
index.html

ads/
script.js → logo.jpg

ads/
frame.html

adnet/   ads/
data.js   img_a.jpg

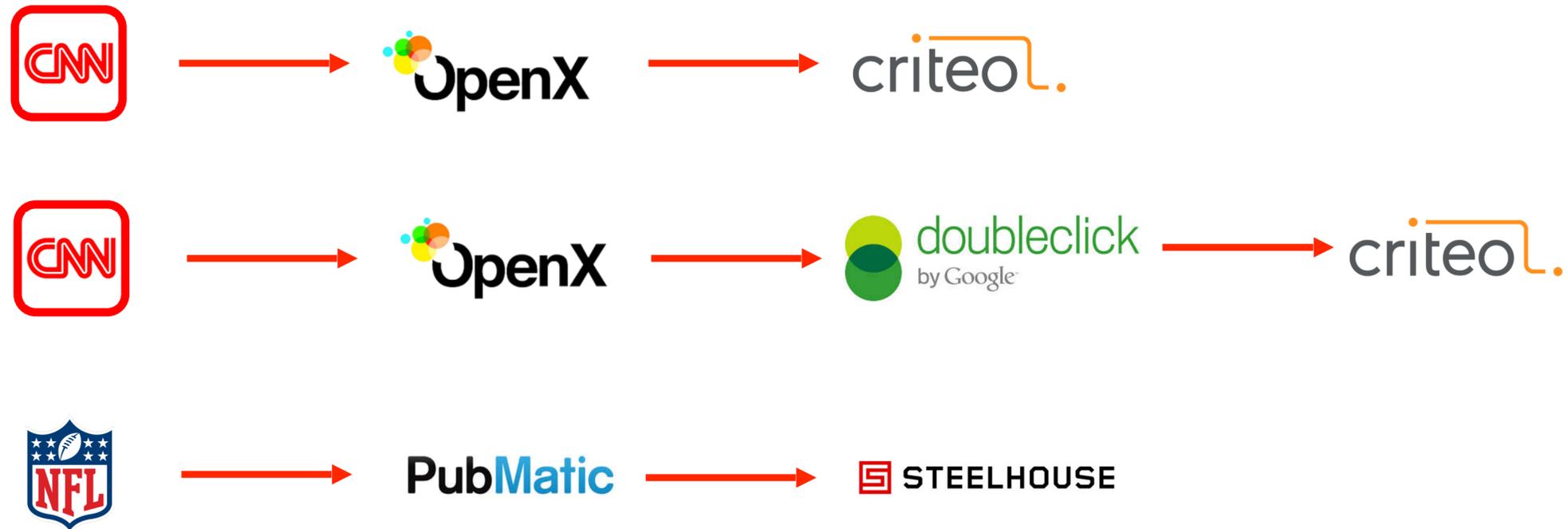**Chain must have at least 3 elements**

Extracting Tuples For Compliance Analysis

Advertisements
(Detected Using EasyList)

Chain must have at least 3 elements

Extracting Tuples For Compliance Analysis

Extracting Tuples For Compliance Analysis

Advertisements
(Detected Using EasyList)

Associated Tuples

Chain must have at least 3 elements

Extracting Tuples For Compliance Analysis

Advertisements
(Detected Using EasyList)

Associated Tuples

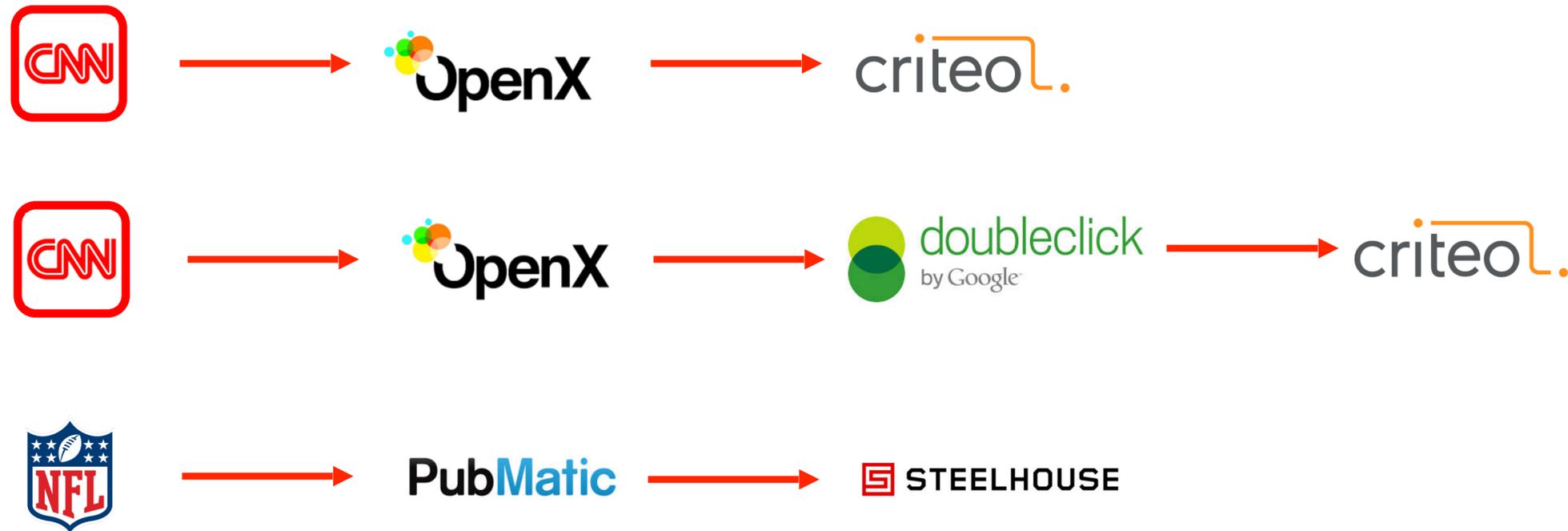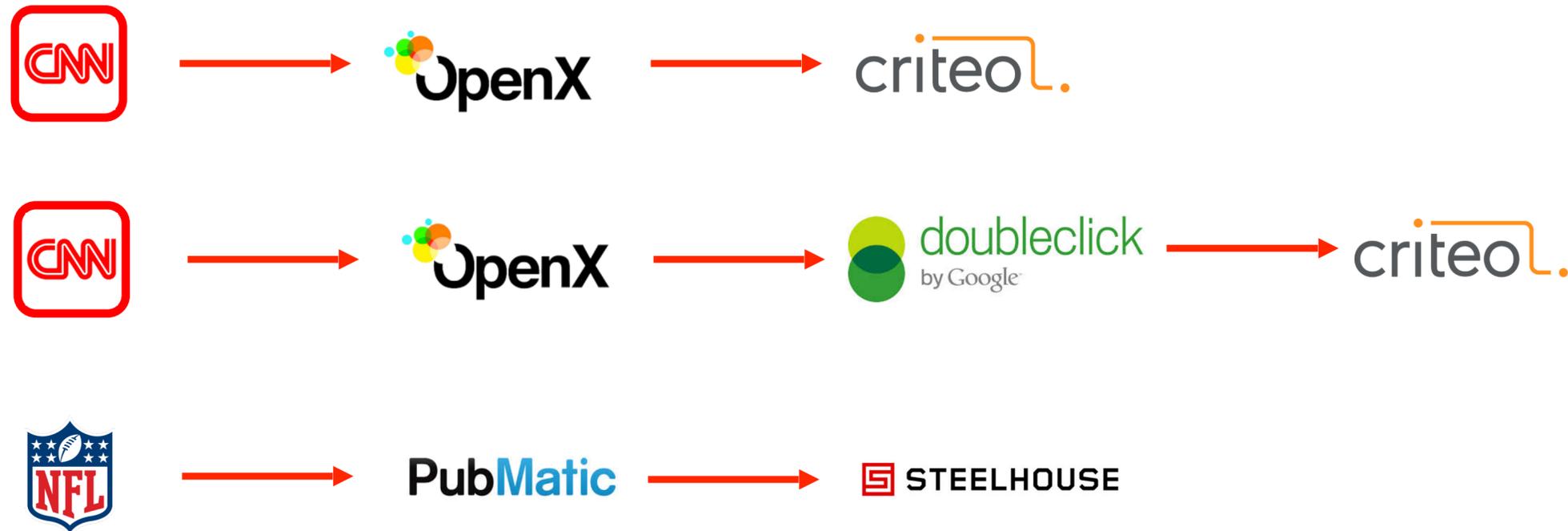**Chain must have at least 3 elements**

Extracting Tuples For Compliance Analysis

Advertisements
(Detected Using EasyList)
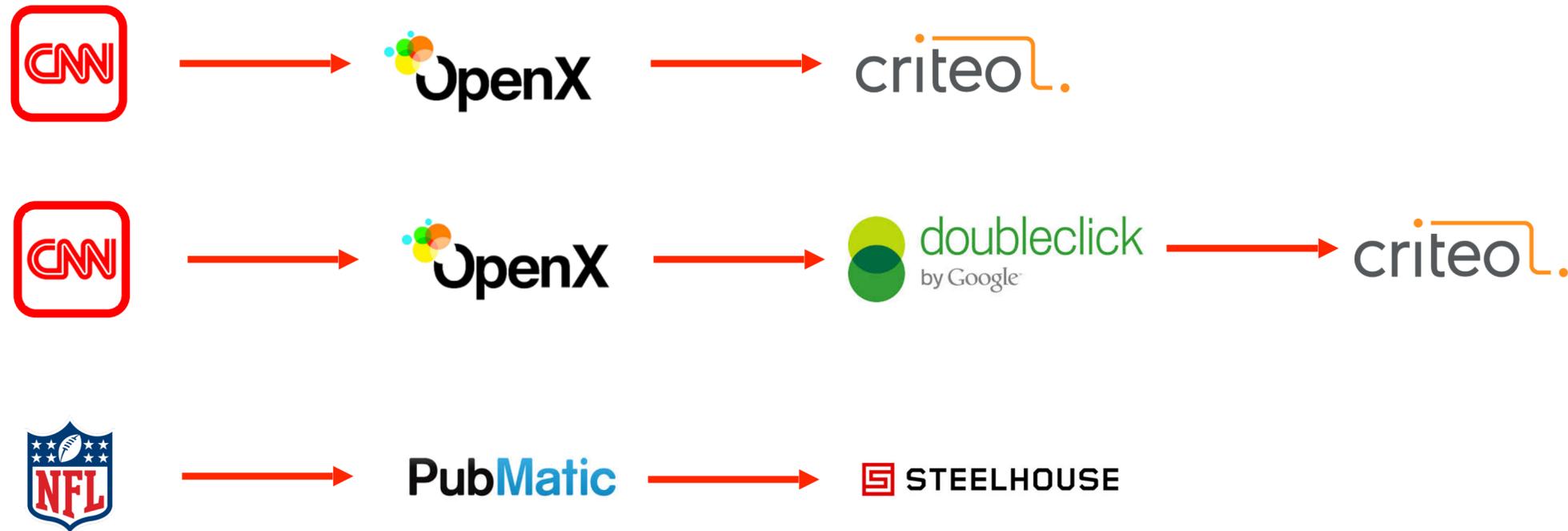
Associated Tuples
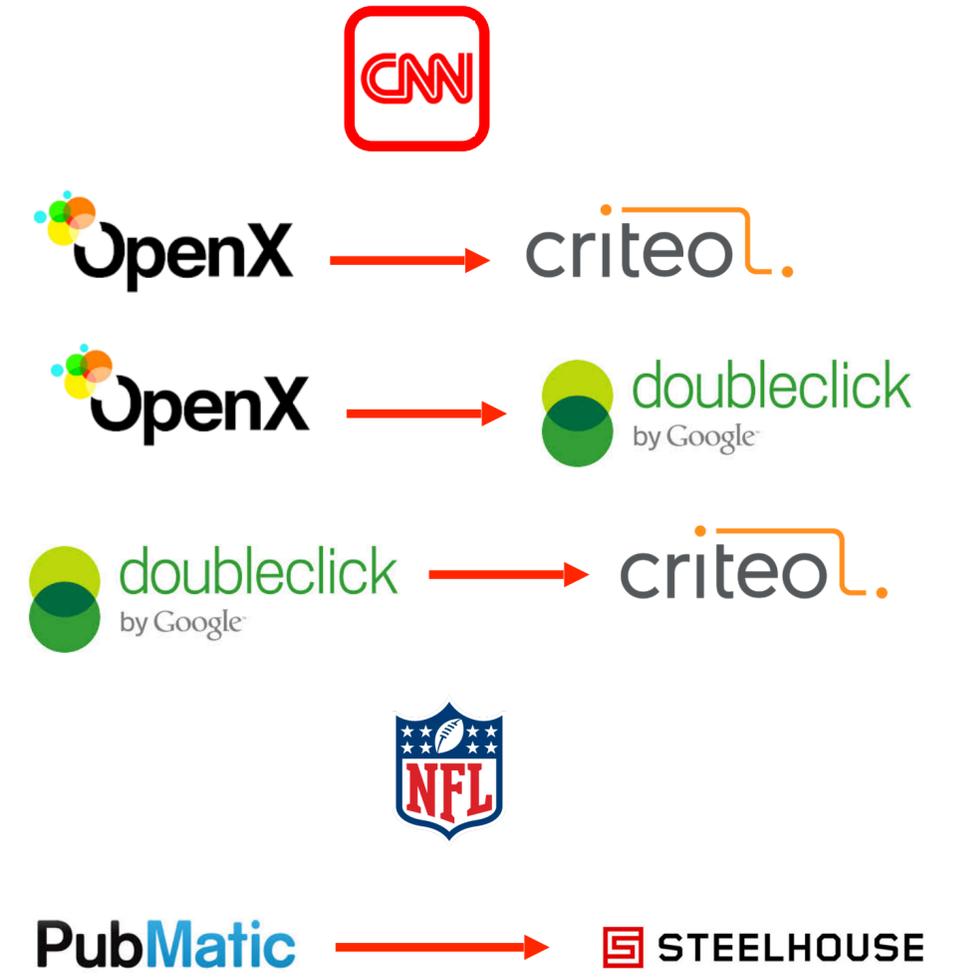
Chain must have at least 3 elements

Compliance Per Tuple

ads.txt

ads.txt
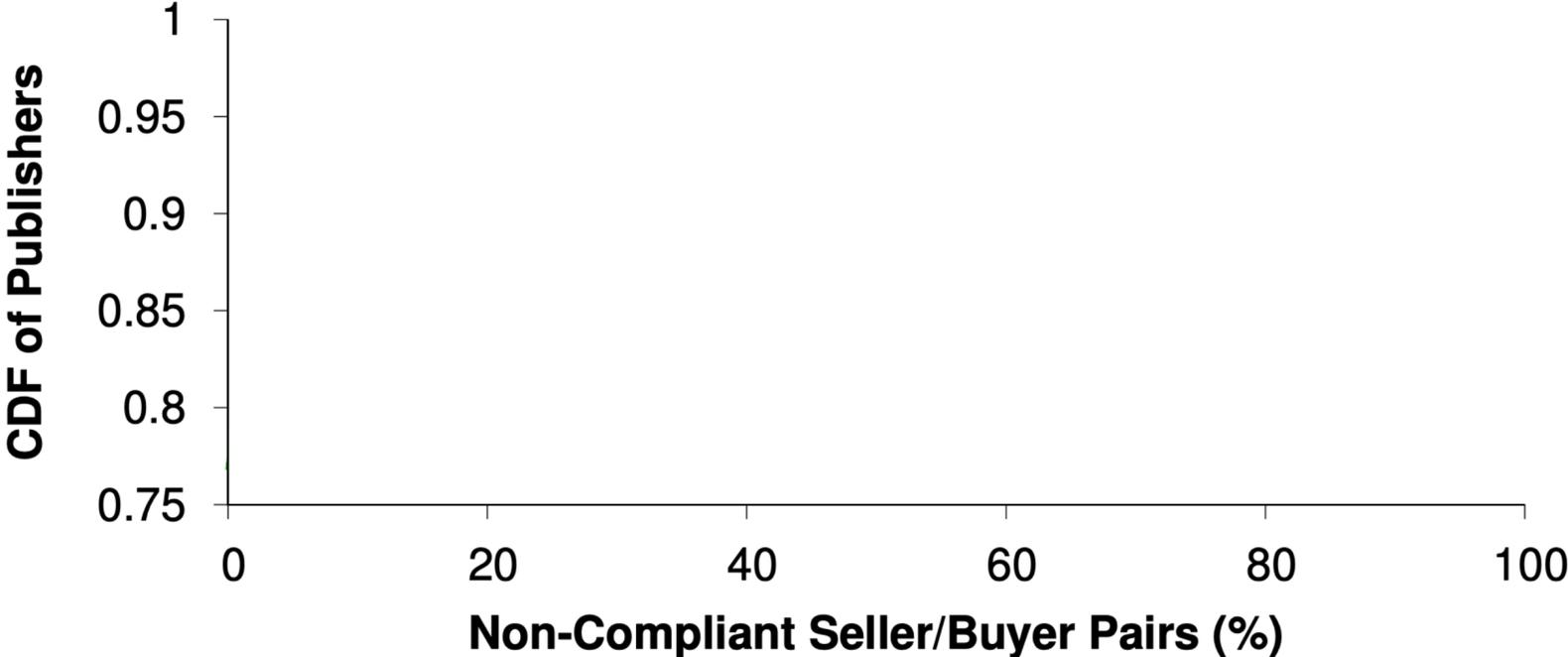
Compliance Per Tuple

ads.txt

ads.txt

Percentage of non-compliant
Exchange-DSP tuples per publisher
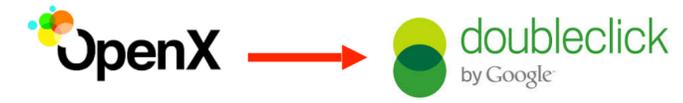
Compliance Per Tuple

Percentage of non-compliant
Exchange-DSP tuples per publisher

Compliance Per Tuple

ads.txt

ads.txt

75% publishers have 100% compliant inventory

CDF of Publishers

Non-Compliant Seller/Buyer Pairs (%)

Percentage of non-compliant Exchange-DSP tuples per publisher

Compliance Per Tuple

ads.txt

ads.txt

75% publishers have 100% compliant inventory

CDF of Publishers vs Non-Compliant Seller/Buyer Pairs (%)

Percentage of non-compliant Exchange-DSP tuples per publisher

04-01-2018
10-01-2018
04-01-2019

CDF of Publishers vs Non-Compliant Seller/Buyer Pairs (%)
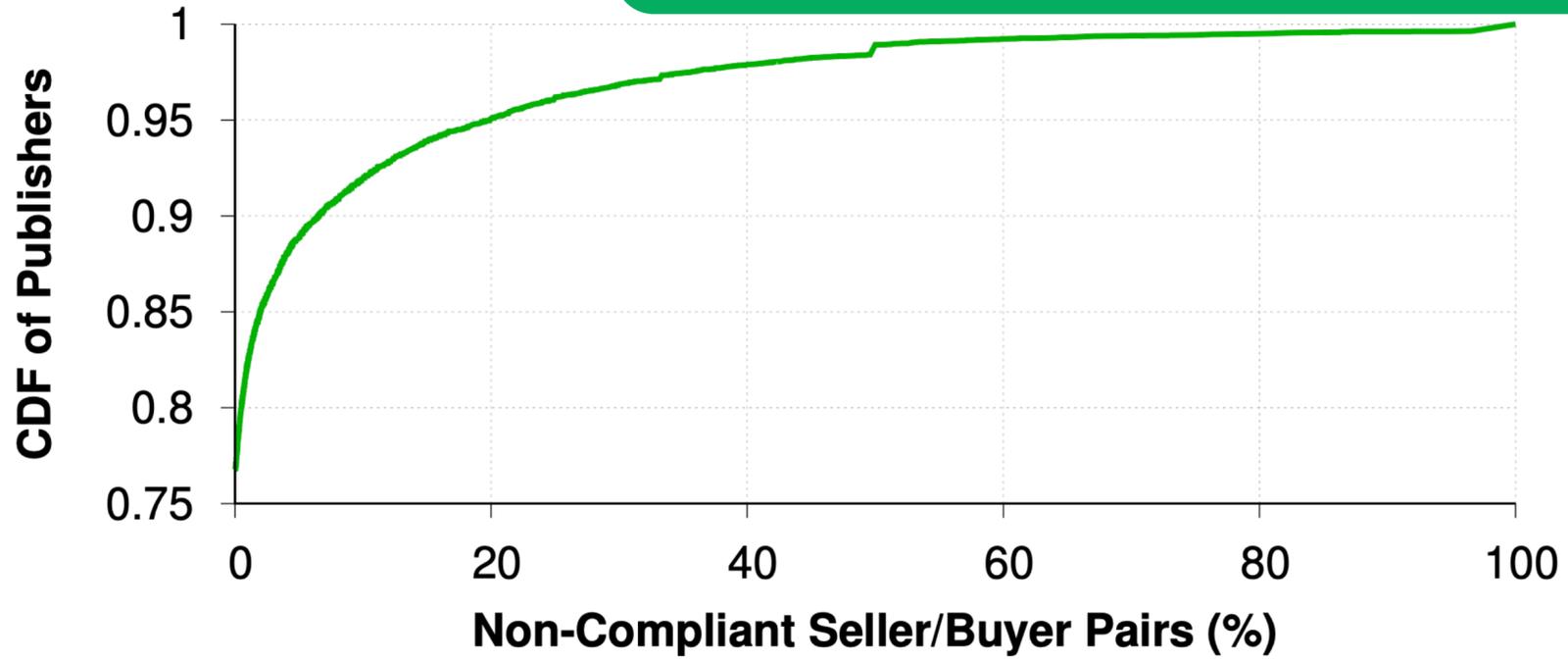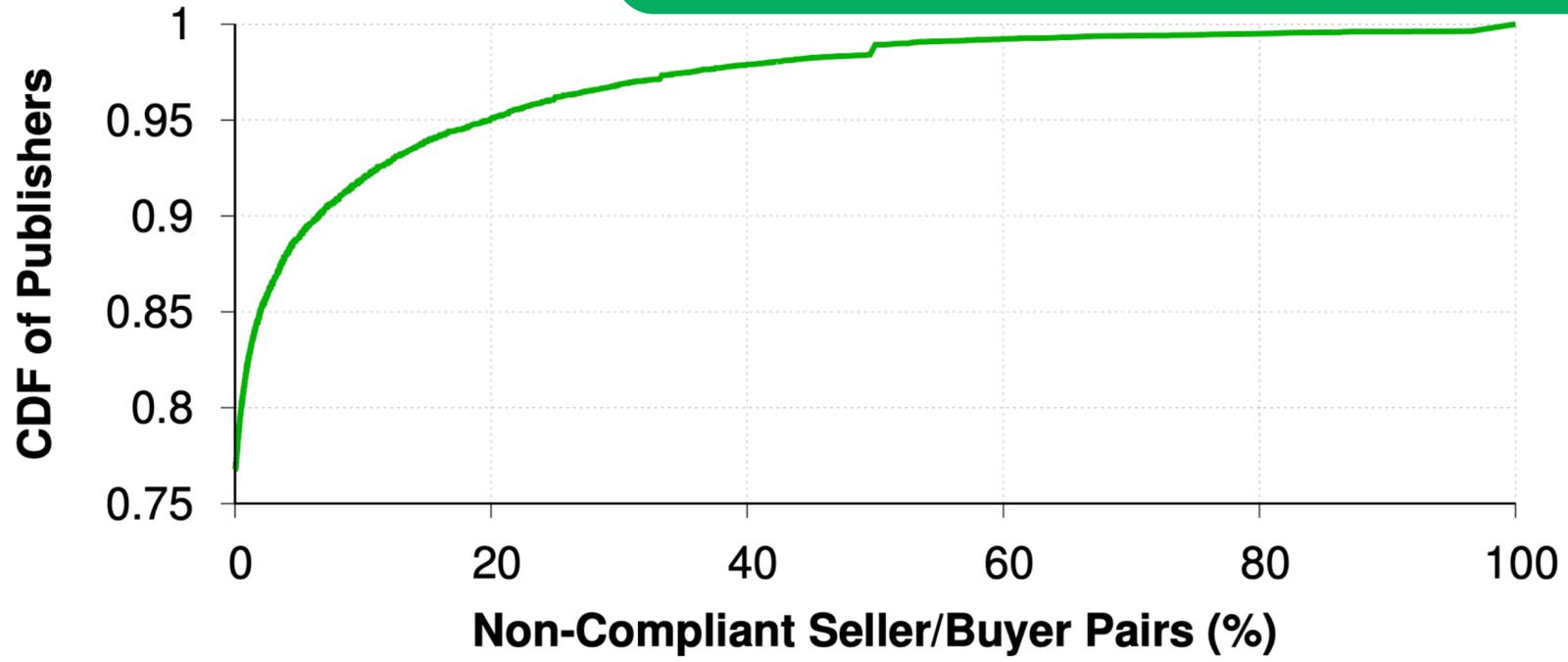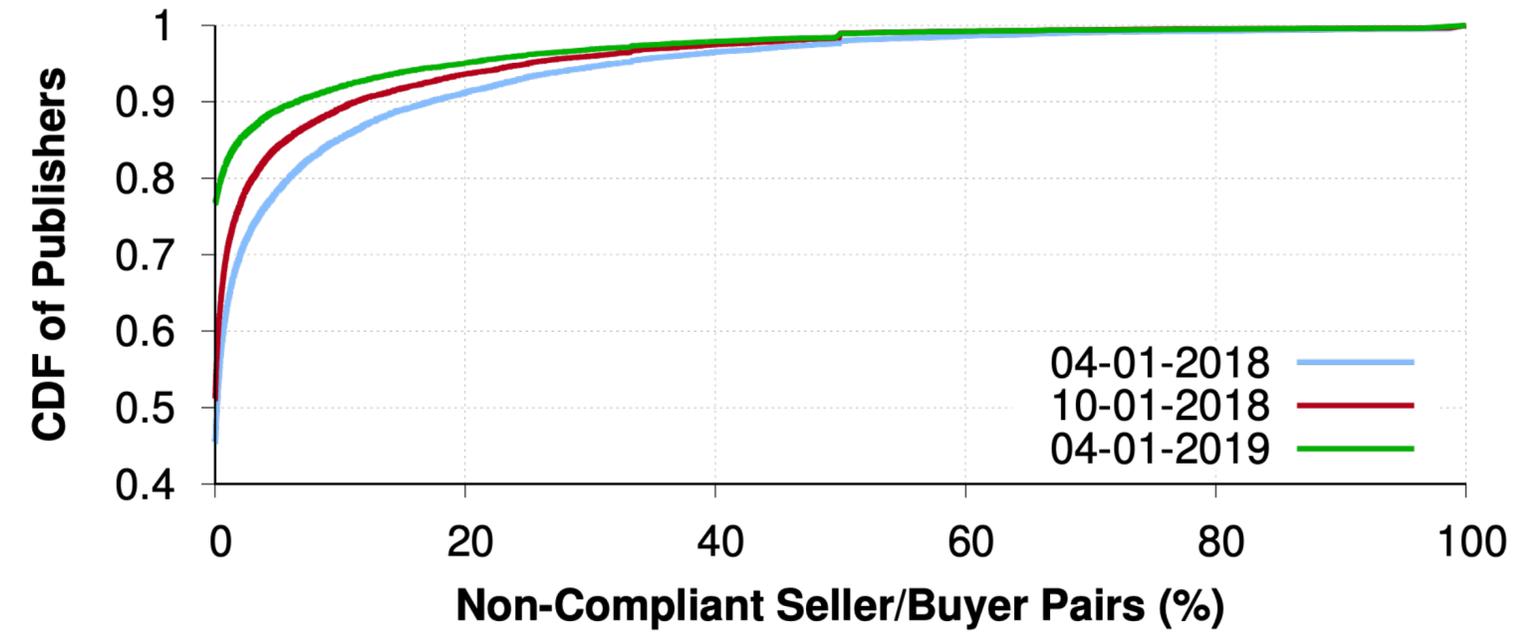
Compliance over time

Compliance Per Tuple

75% publishers have 100% compliant inventory
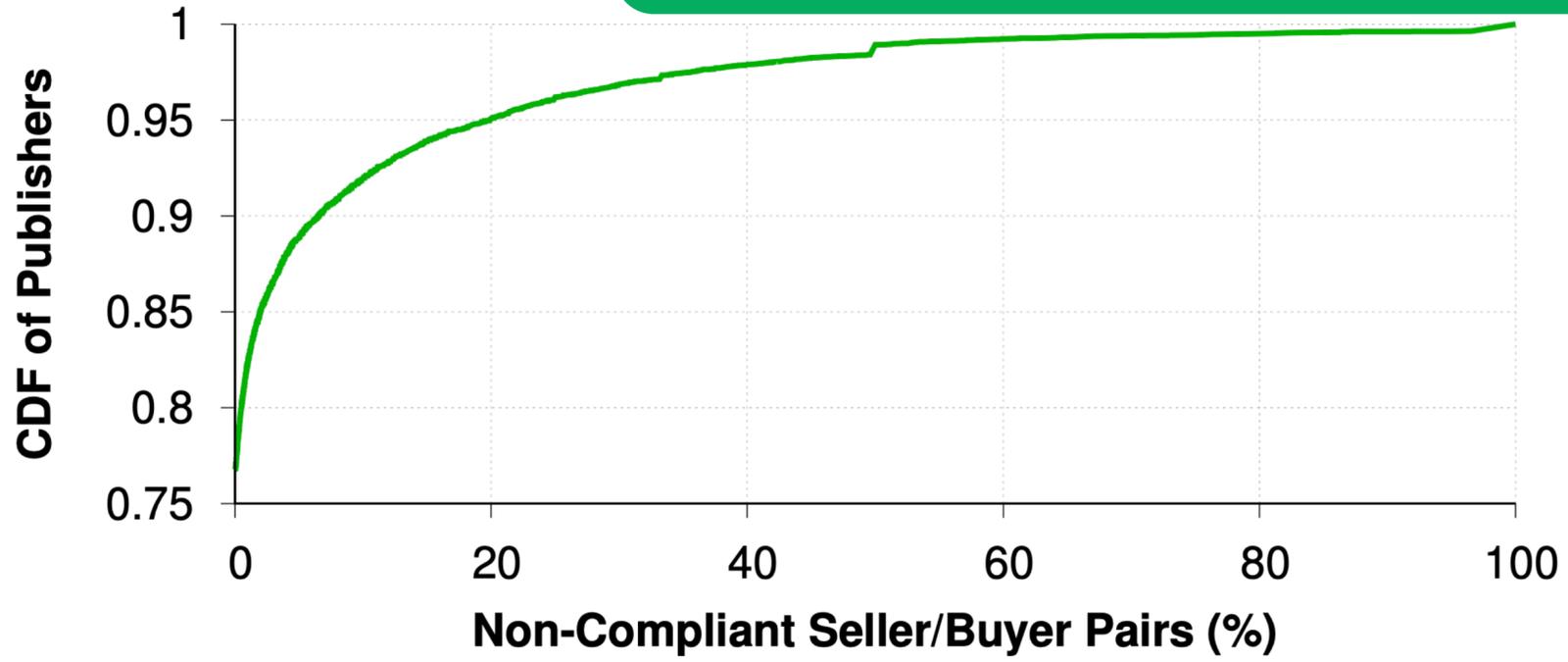
Compliance has increased over the year

04-01-2018
10-01-2018
04-01-2019

CDF of Publishers

Non-Compliant Seller/Buyer Pairs (%)

Percentage of non-compliant Exchange-DSP tuples per publisher
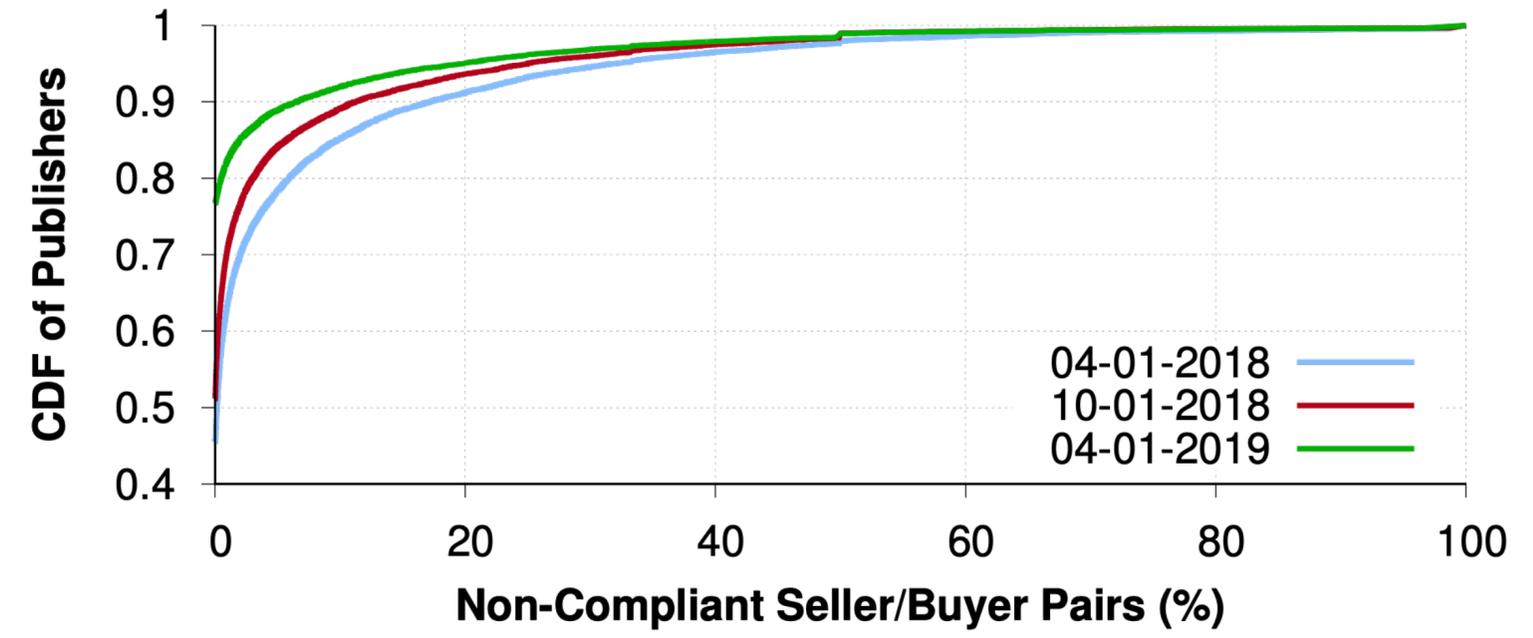
Compliance over time

Top Non-Compliant Tuples

Exchange / Seller	DSP / Buyer	% Across all pubs
gumgum	domdex	20.4
gumgum	appnexus	20.5
taboola	weborama	52.7
taboola	rubiconproject	11.6
gumgum	pubmatic	27.3
justpremium	openx	100
rubiconproject	yahoo	2.63
springserve	pubmatic	49.3
nativeroll	weborama	100
nativeroll	seedr	100

Top Non-Compliant Tuples

Exchange / Seller	DSP / Buyer	% Across all pubs
gumgum	domdex	20.4
gumgum	appnexus	20.5
taboola	weborama	52.7
taboola	rubiconproject	11.6
gumgum	pubmatic	27.3
justpremium	openx	100
rubiconproject	yahoo	2.63
springserve	pubmatic	49.3
nativeroll	weborama	100
nativeroll	seedr	100

Sellers

- **Some sellers are systematically non-compliant (e.g. NativeRoll, JustPremium).**

Top Non-Compliant Tuples

Exchange / Seller	DSP / Buyer	% Across all pubs
gumgum	domdex	20.4
gumgum	appnexus	20.5
taboola	weborama	52.7
taboola	rubiconproject	11.6
gumgum	pubmatic	27.3
justpremium	openx	100
rubiconproject	yahoo	2.63
springserve	pubmatic	49.3
nativeroll	weborama	100
nativeroll	seedr	100

Sellers

- **Some sellers are systematically non-compliant (e.g. NativeRoll, JustPremium).**

- **Top exchanges do not appear as non-compliant.**

Top Non-Compliant Tuples

Exchange / Seller	DSP / Buyer	% Across all pubs
gumgum	domdex	20.4
gumgum	appnexus	20.5
taboola	weborama	52.7
taboola	rubiconproject	11.6
gumgum	pubmatic	27.3
justpremium	openx	100
rubiconproject	yahoo	2.63
springserve	pubmatic	49.3
nativeroll	weborama	100
nativeroll	seedr	100

Sellers

- **Some sellers are systematically non-compliant (e.g. NativeRoll, JustPremium).**

- **Top exchanges do not appear as non-compliant.**

Buyers

- **Most are actually top exchanges (sellers).**

- **8/10 top sellers are buying non-compliant inventory.**

Compliance of ads.txt Standard

Compliance of ads.txt Standard

Top exchanges and DSPs appear to be honoring the standard

More than 75% of publishers that run RTB have 100% compliance

Major ad exchanges purchasing non-compliant inventory

Talk Overview

1. RTB background & motivation for the ads.txt standard

2. Data collection

3. Adoption trends

4. Compliance of the ads.txt standard

5. Means of providing transparency

Identification of A&A Domains

Efforts to identify Advertising and Analytics related domains.

- Filter lists (e.g. EasyList) are manually curated

- Machine learning based approaches

Identification of A&A Domains

Efforts to identify Advertising and Analytics related domains.

- Filter lists (e.g. EasyList) are manually curated

- Machine learning based approaches

https://cnn.com/ads.txt

```
google.com, pub-7439281311086140, DIRECT
rubiconproject.com, 11078, DIRECT
c.amazon-adsystem.com, 3159, DIRECT
openx.com, 183753,RESELLER RESELLER
```

1035 valid domains

https://washingtonpost.com/ads.txt

```
indexexchange.com, 197545, DIRECT
adtech.com, 232547, DIRECT
openx.com, 203799,RESELLER
```

Top Exchanges like Google, openX have adopted

New Standards Coming As Well

Have the potential to make the ecosystem even more transparent

- App-ads.txt (For mobile apps)

- Sellers.json (To reveal relationships of sellers in the ecosystem)

Conclusions / Takeaways

- A 15-month study to study the ads.txt standard

  - We observed heathy adoption (64% publishers have adopted)

  - Major exchanges & DSPs are part of it, and honoring the standard

- A step towards making the ecosystem transparent

  - Data needs to be cleaned

Dataset
https://personalization.ccs.neu.edu/Projects/Adstxt

Questions?
ahmad@ccs.neu.edu