HoPLA: a Honeypot PLatform to Lure Attackers

Elisa Chiapponi - EURECOM
Onur Catakoglu - Amadeus IT Group
Olivier Thonnard - Amadeus IT Group
Marc Dacier- EURECOM

EURECOM
Sophia Antipolis

amadeus
Global Security Operations

Who are we?

Elisa Chiapponi

- Ph.D. student at EURECOM, collaboration with Amadeus IT Group
- Analysis and mitigation of the new generation of botnet

Dr. Onur Catakoglu

- Information Security Architect, GSO Amadeus IT Group

Dr. Olivier Thonnard

- Senior Security Expert, Tech Lead GSO Amadeus IT Group

Prof. Marc Dacier

- Head of Digital Security Department, EURECOM

Who are we?

Elisa Chiapponi

- Ph.D. student at EURECOM, collaboration with Amadeus IT Group
- Analysis and mitigation of the new generation of botnet

Dr. Onur Catakoglu

- Information Security Architect, GSO Amadeus IT Group

Dr. Olivier Thonnard

- Senior Security Expert, Tech Lead GSO Amadeus IT Group

Prof. Marc Dacier

- Head of Digital Security Department, EURECOM

Our Agenda

Our Agenda

1. Introduction
and motivations

Our Agenda

1. Introduction and motivations

2. Experimental setup

Our Agenda

1. Introduction and motivations

2. Experimental setup

3. Results and discussion

Our Agenda

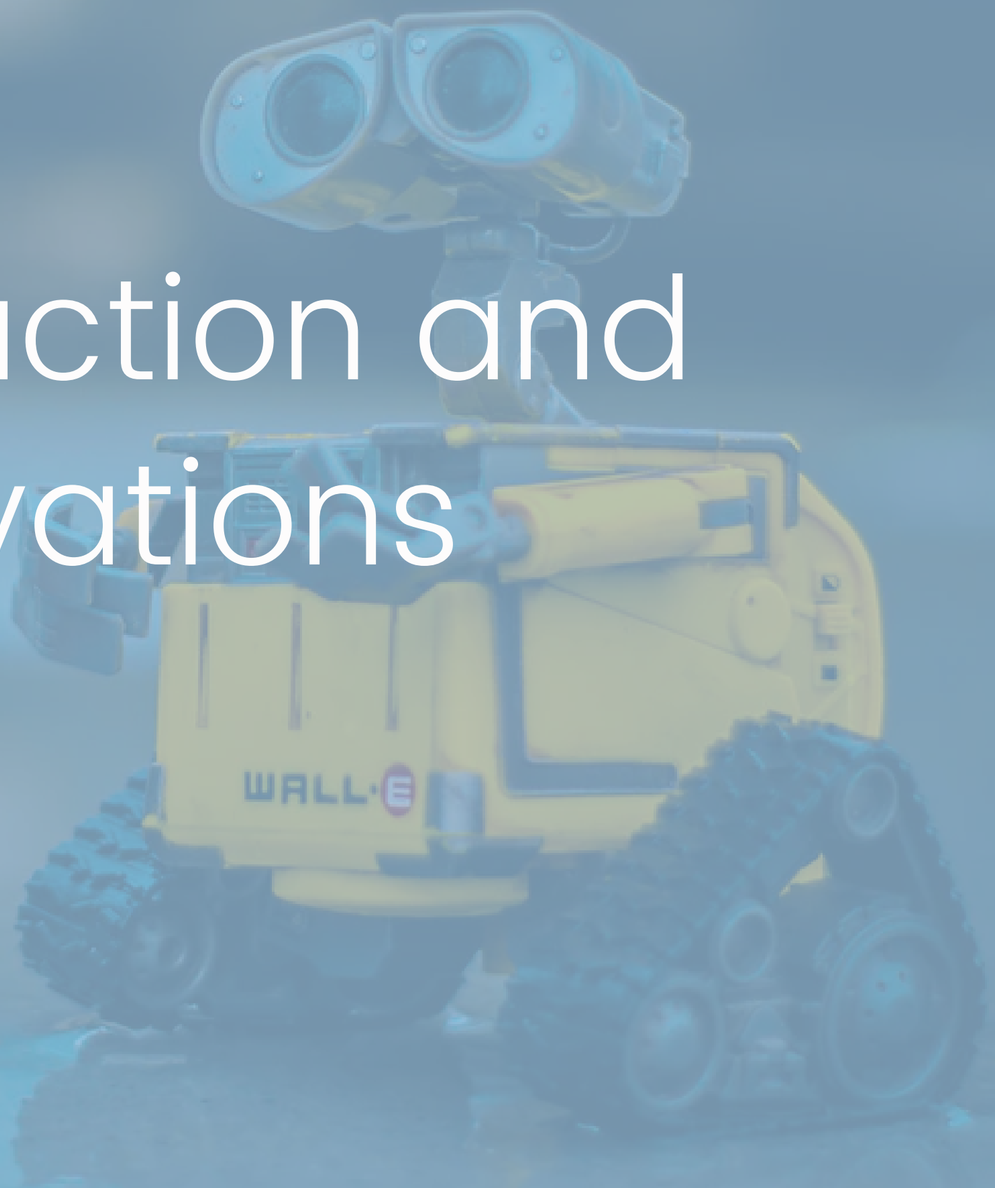
1. Introduction and motivations

2. Experimental setup

3. Results and discussion

4. Conclusions and future works

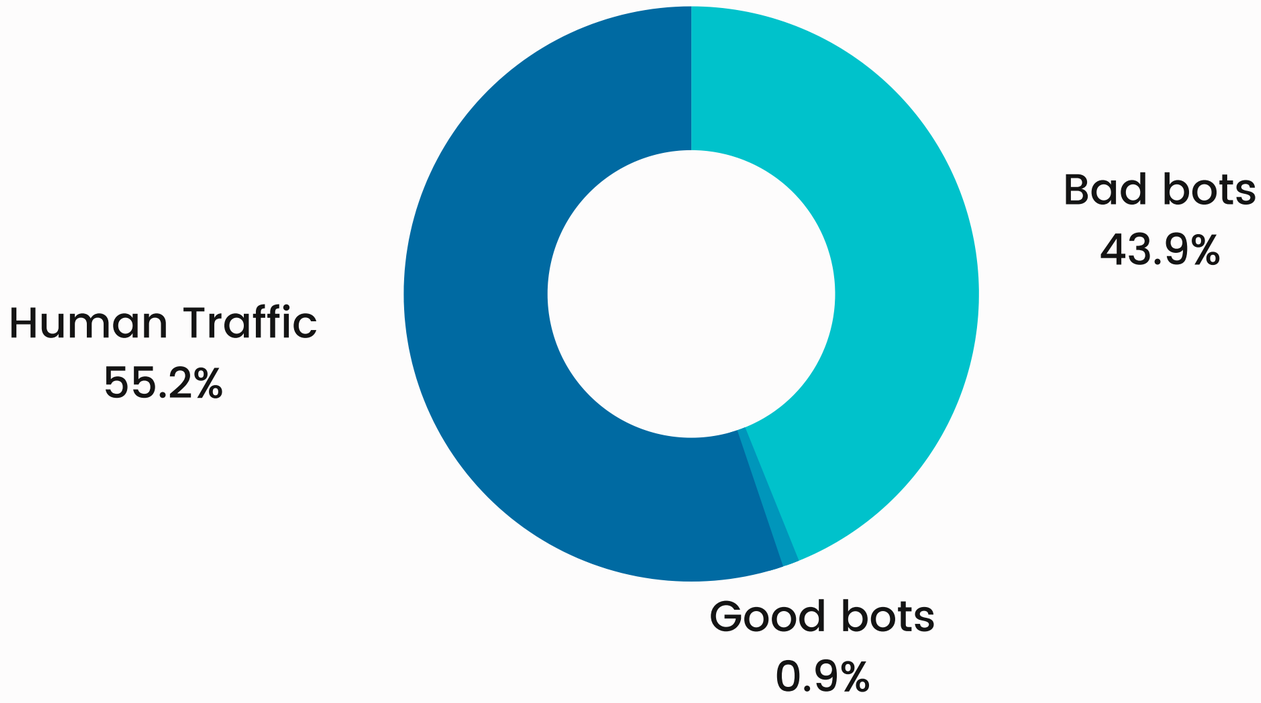
1. Introduction and motivations



Web scraping

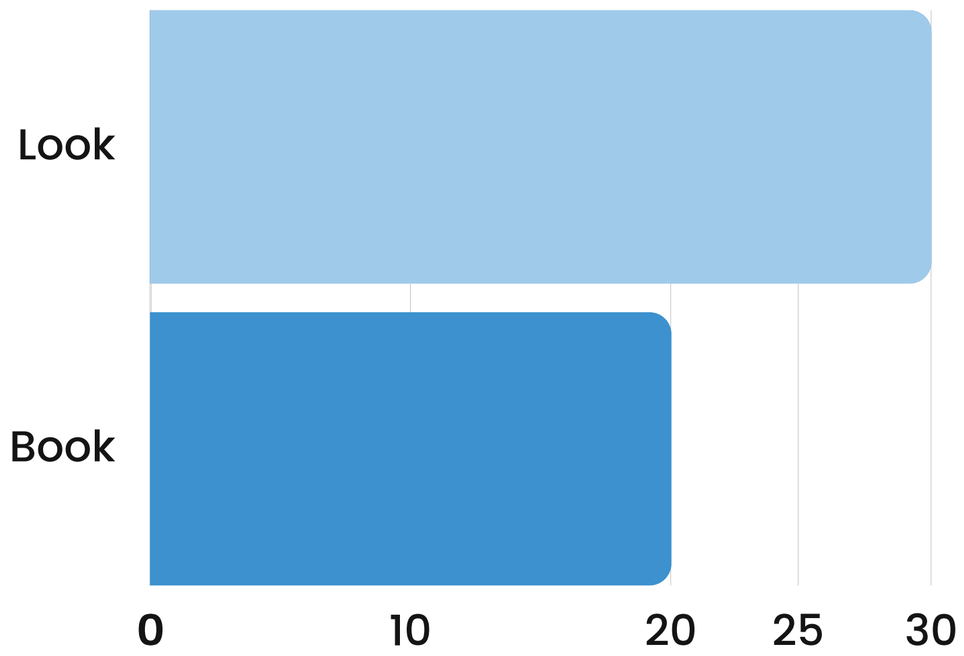
Web scraping is the periodical or continuous retrieval of accessible data and/or processed output contained in web pages.

Scraping bots and airlines

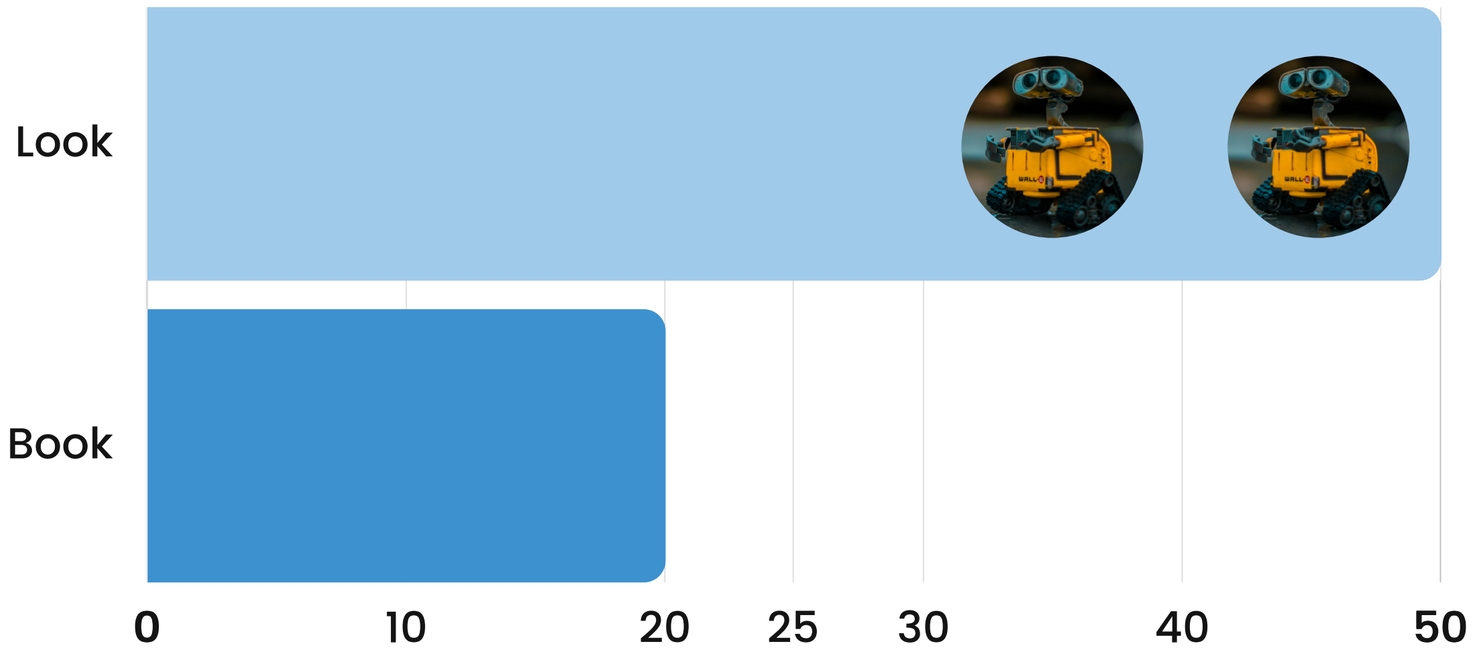


How bots affect airlines - Imperva 2019

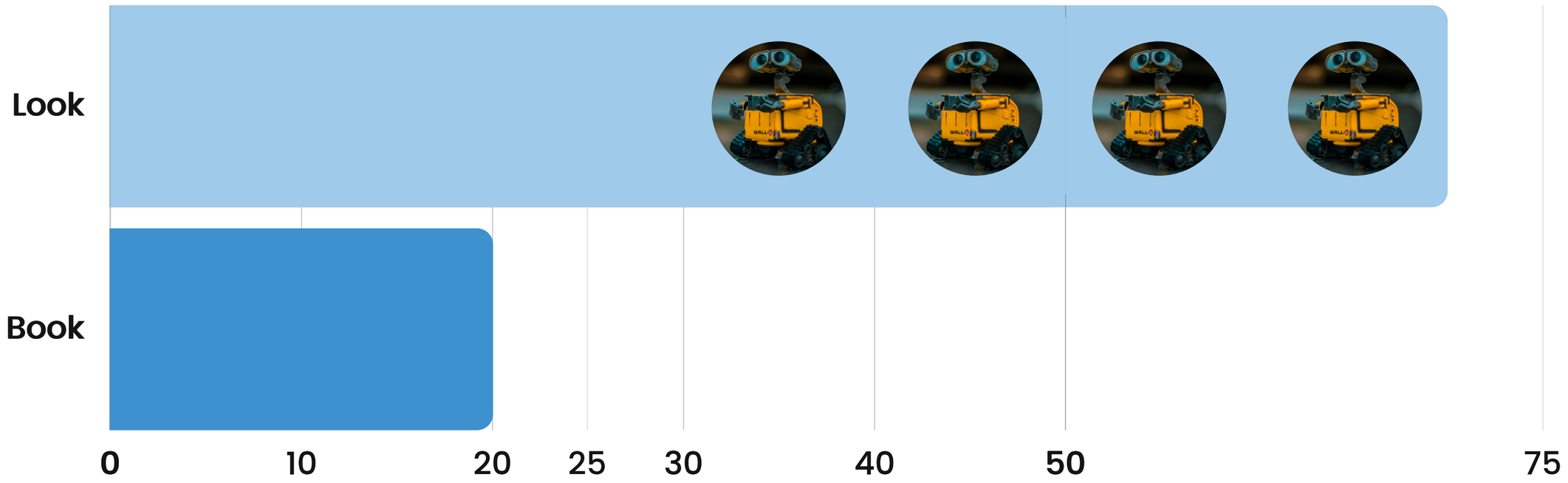
Look-to-book ratio



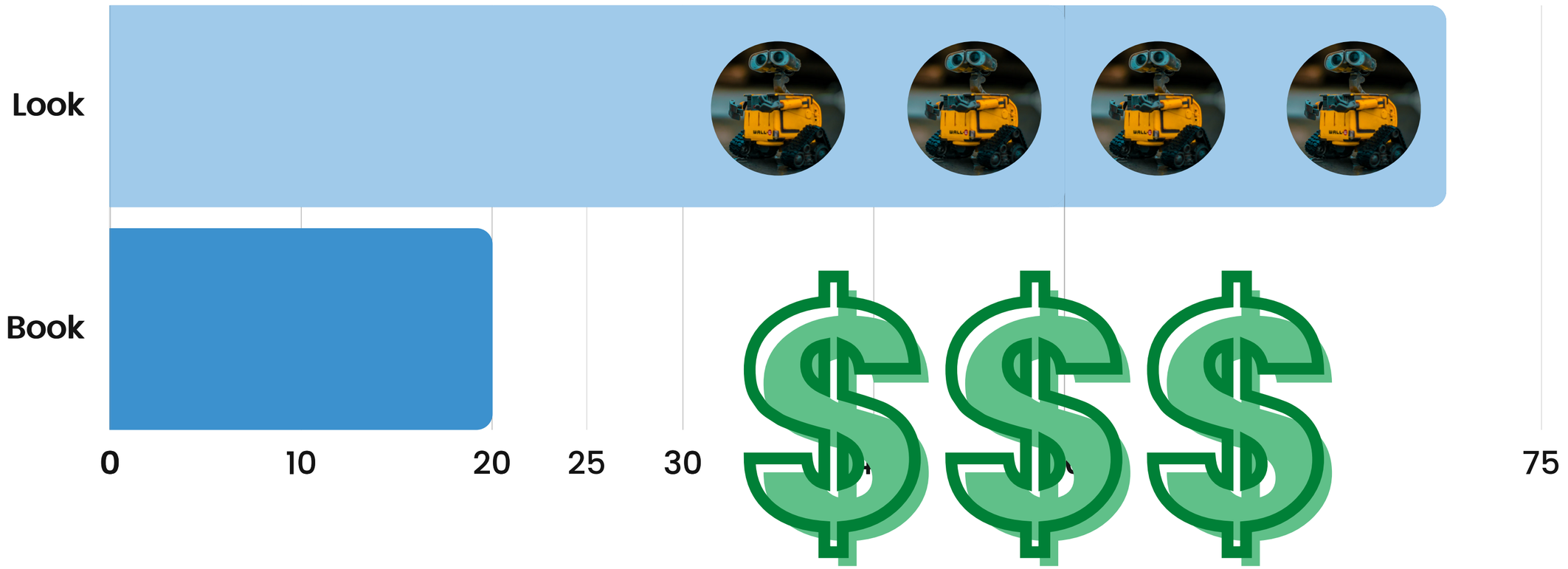
Look-to-book ratio



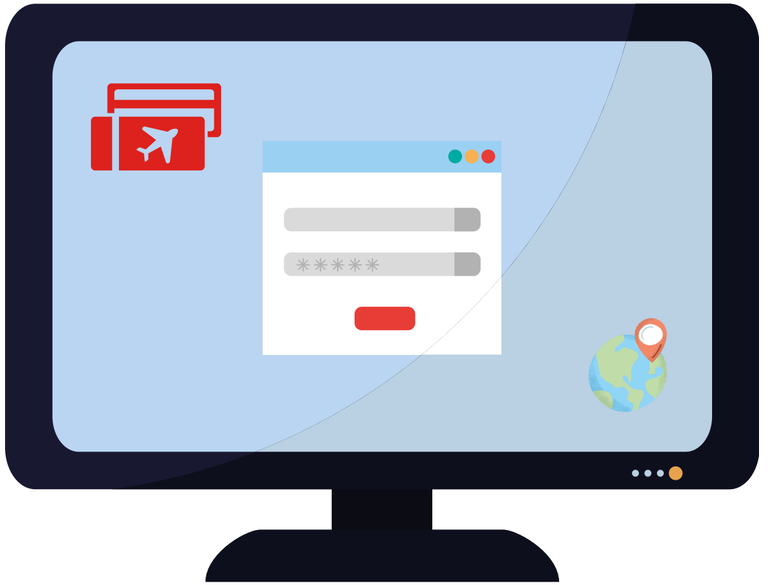
Look-to-book ratio



Look-to-book ratio

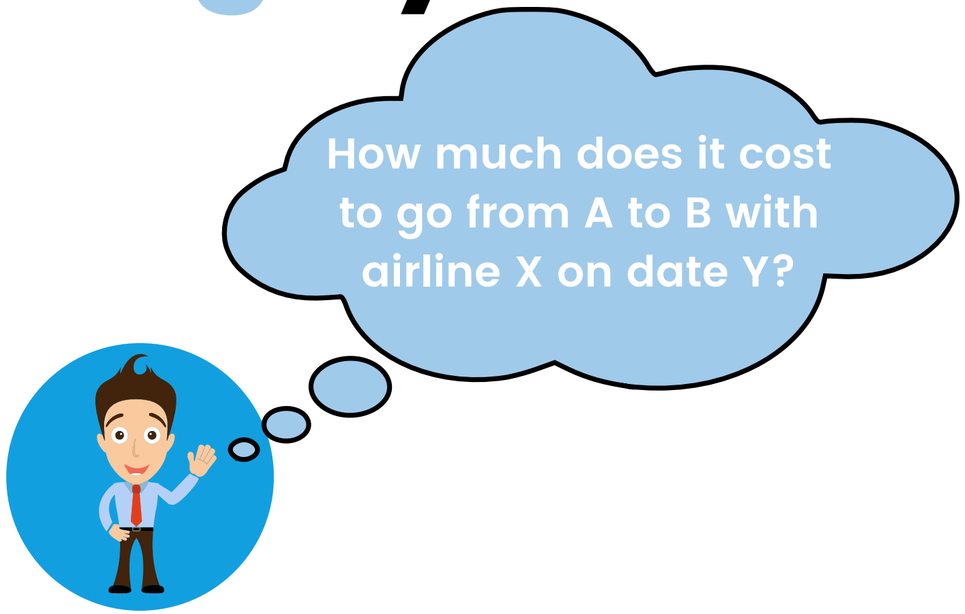


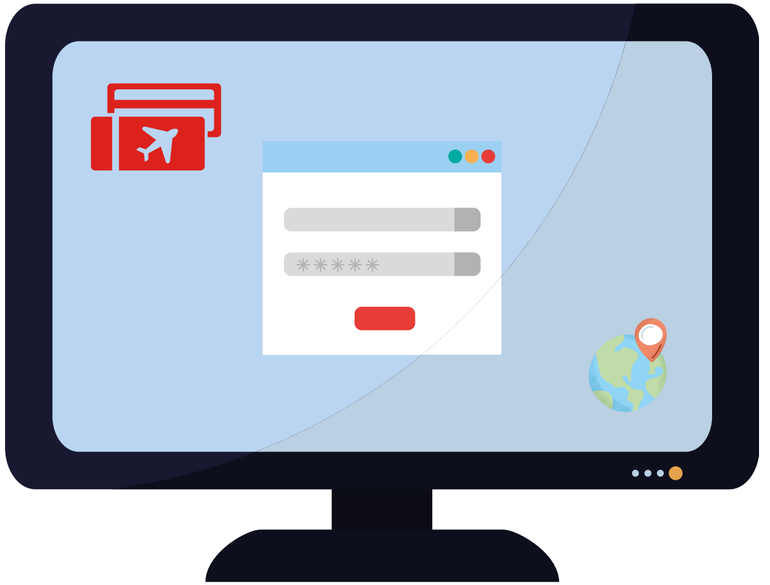
Booking systems



Booking domain
of airline X

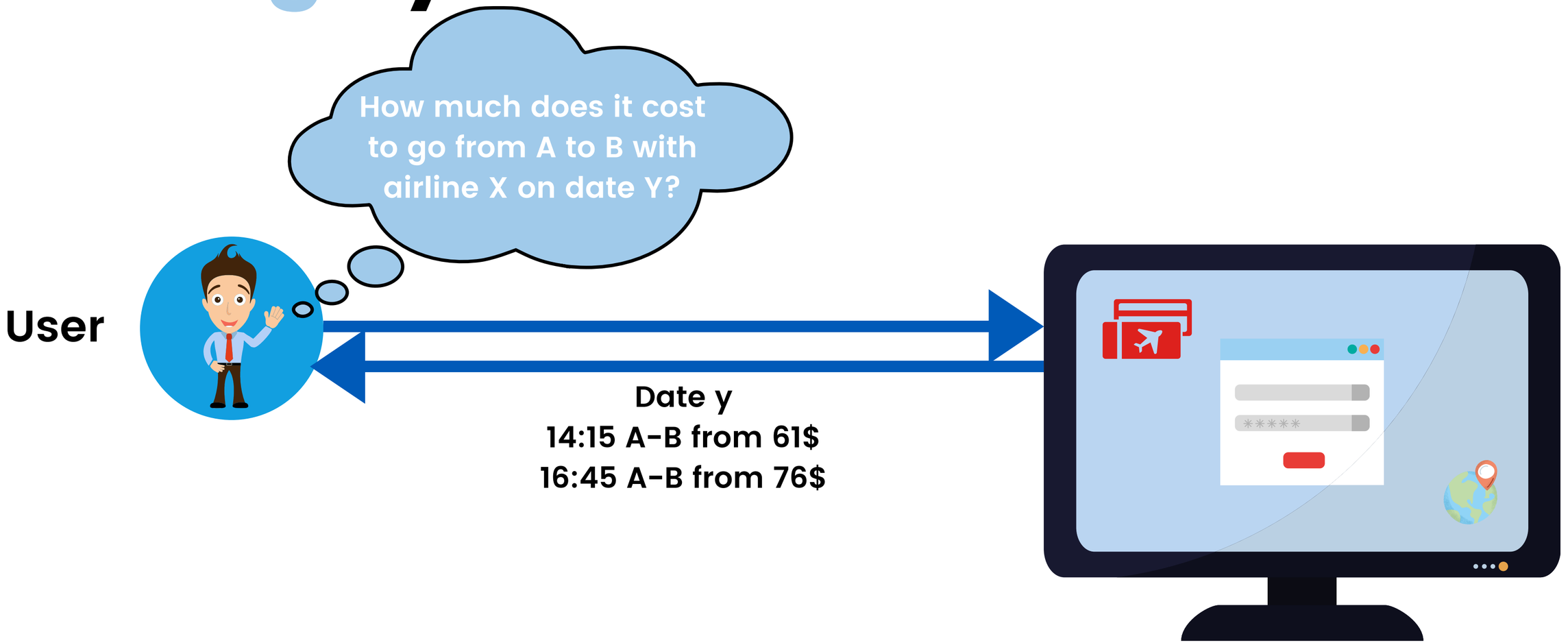
Booking systems





Booking domain of airline X

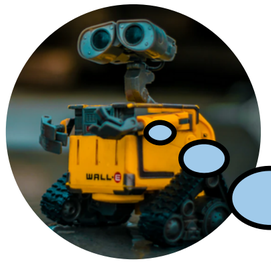
Booking systems



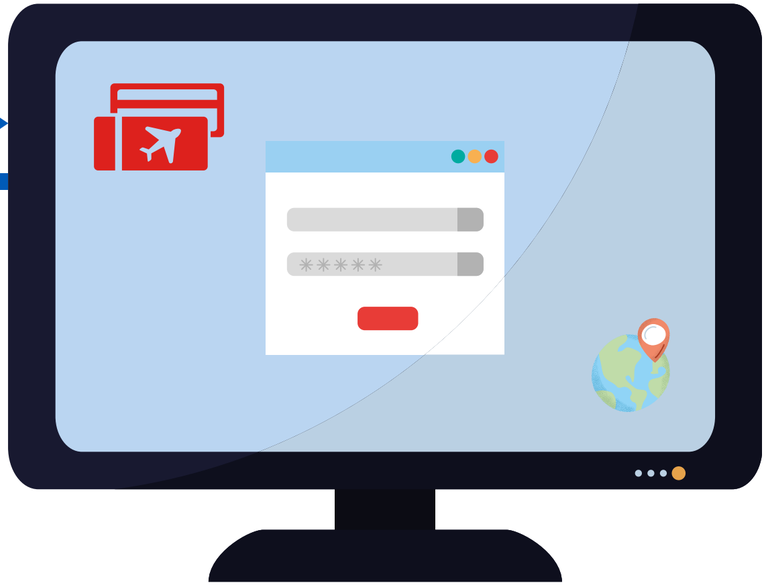
Booking systems

User

Bot

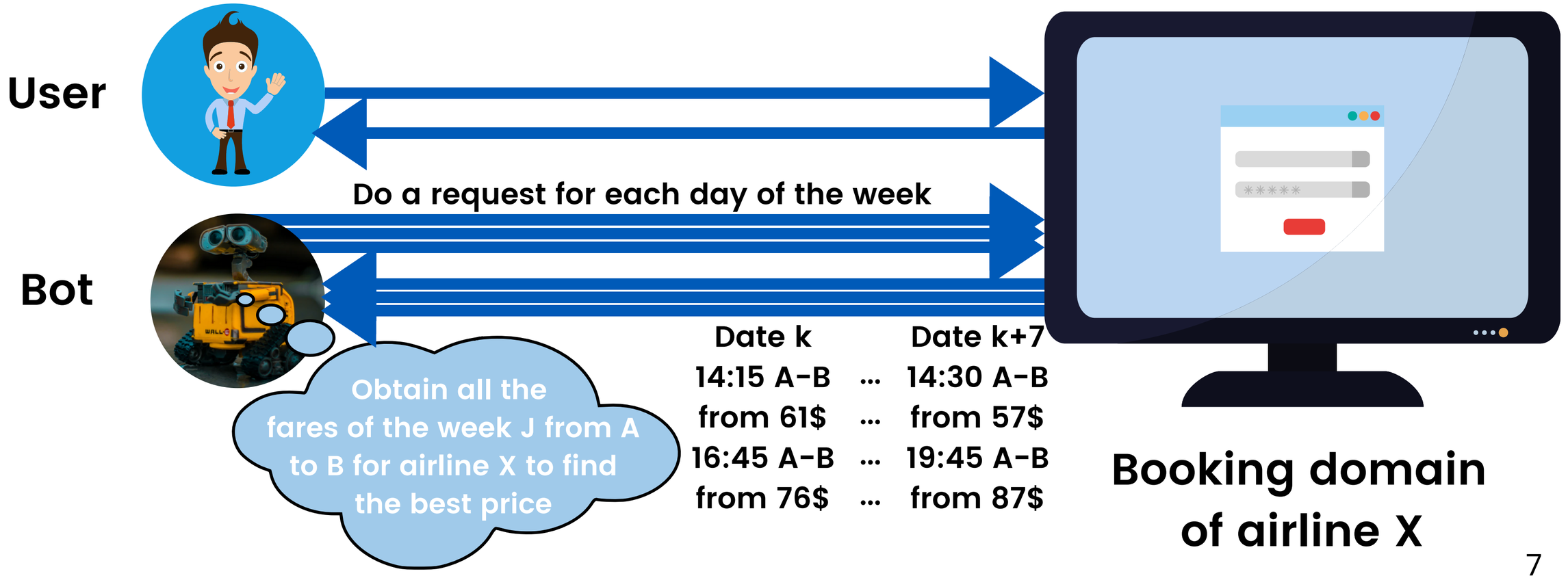


Obtain all the fares of the week J from A to B for airline X to find the best price



Booking domain of airline X

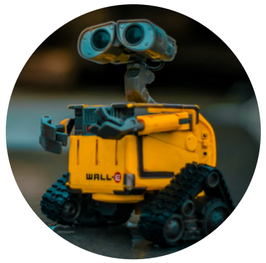
Booking systems

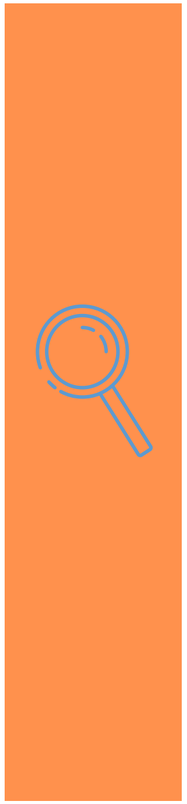


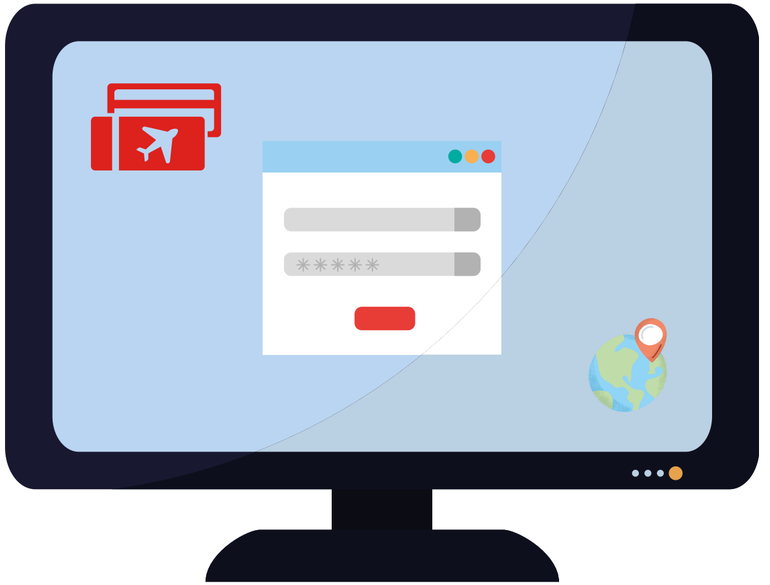
Anti-bot solutions

User

Bot

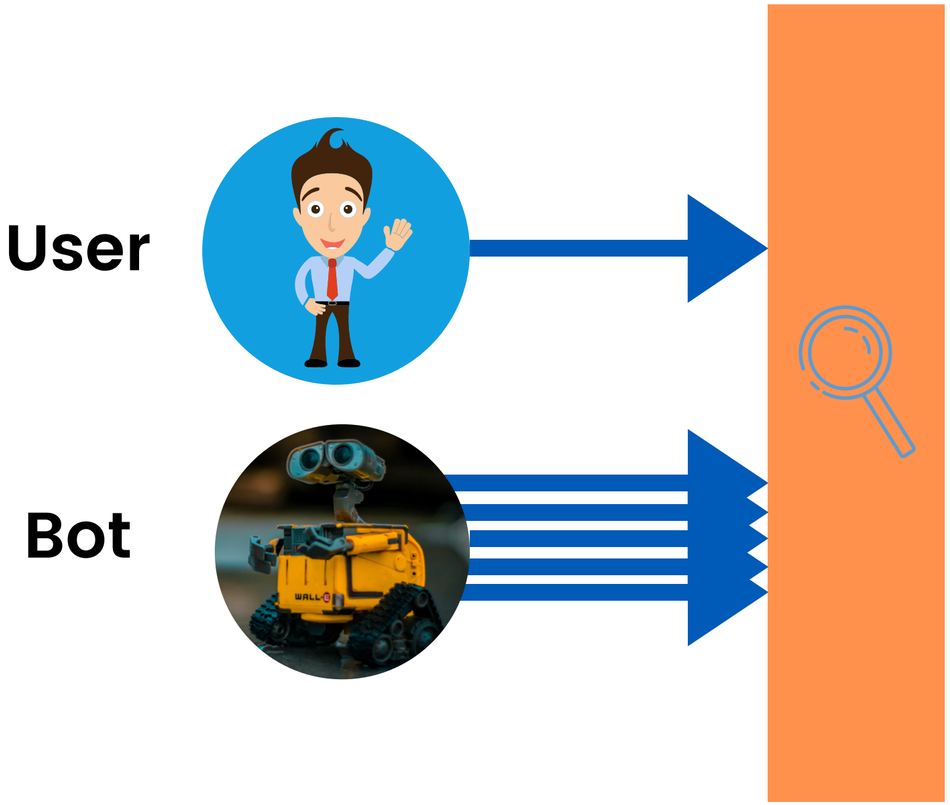


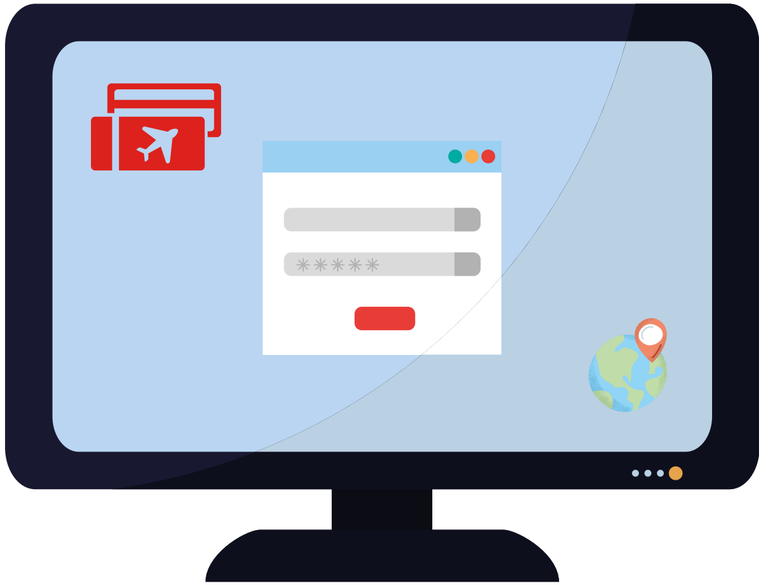




Booking domain
of airline X

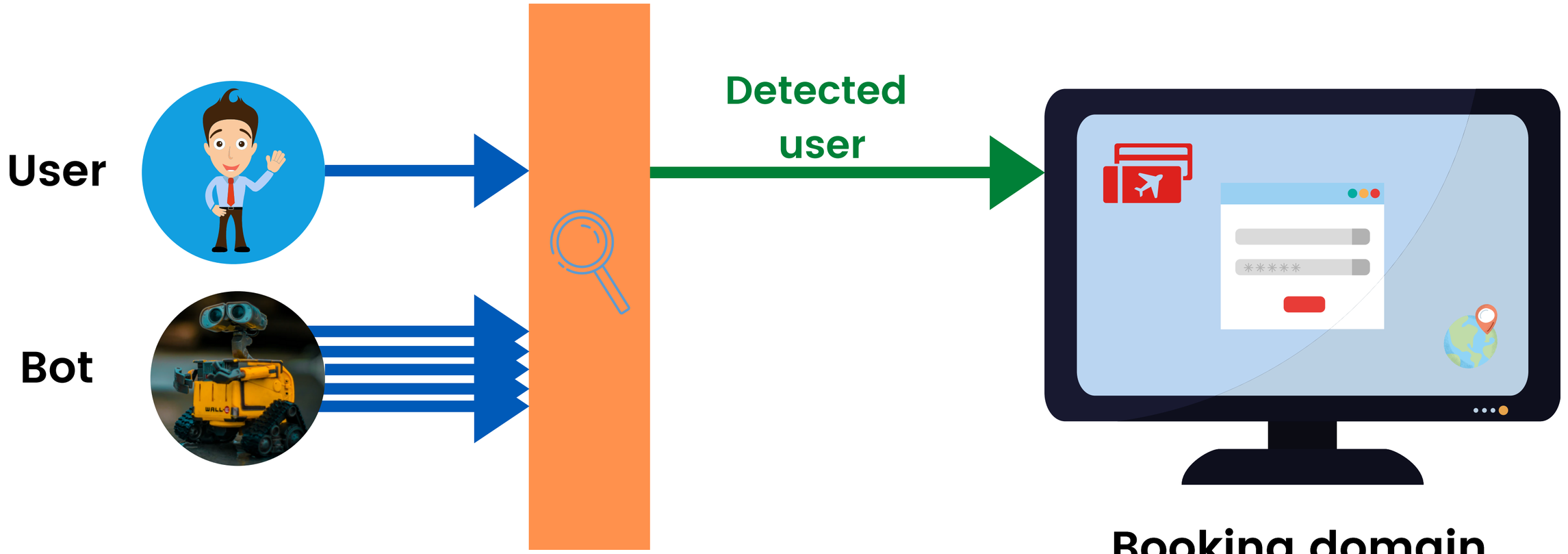
Anti-bot solutions



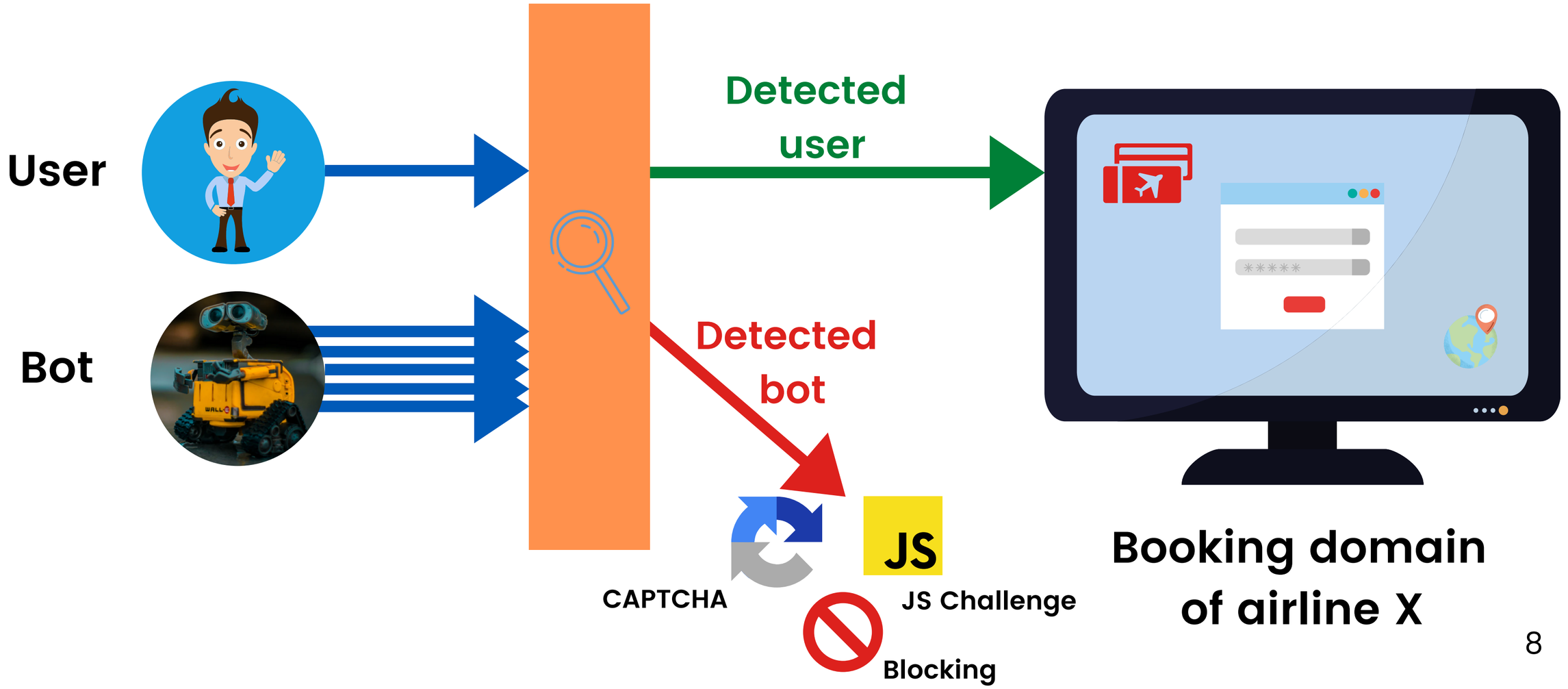


Booking domain
of airline X

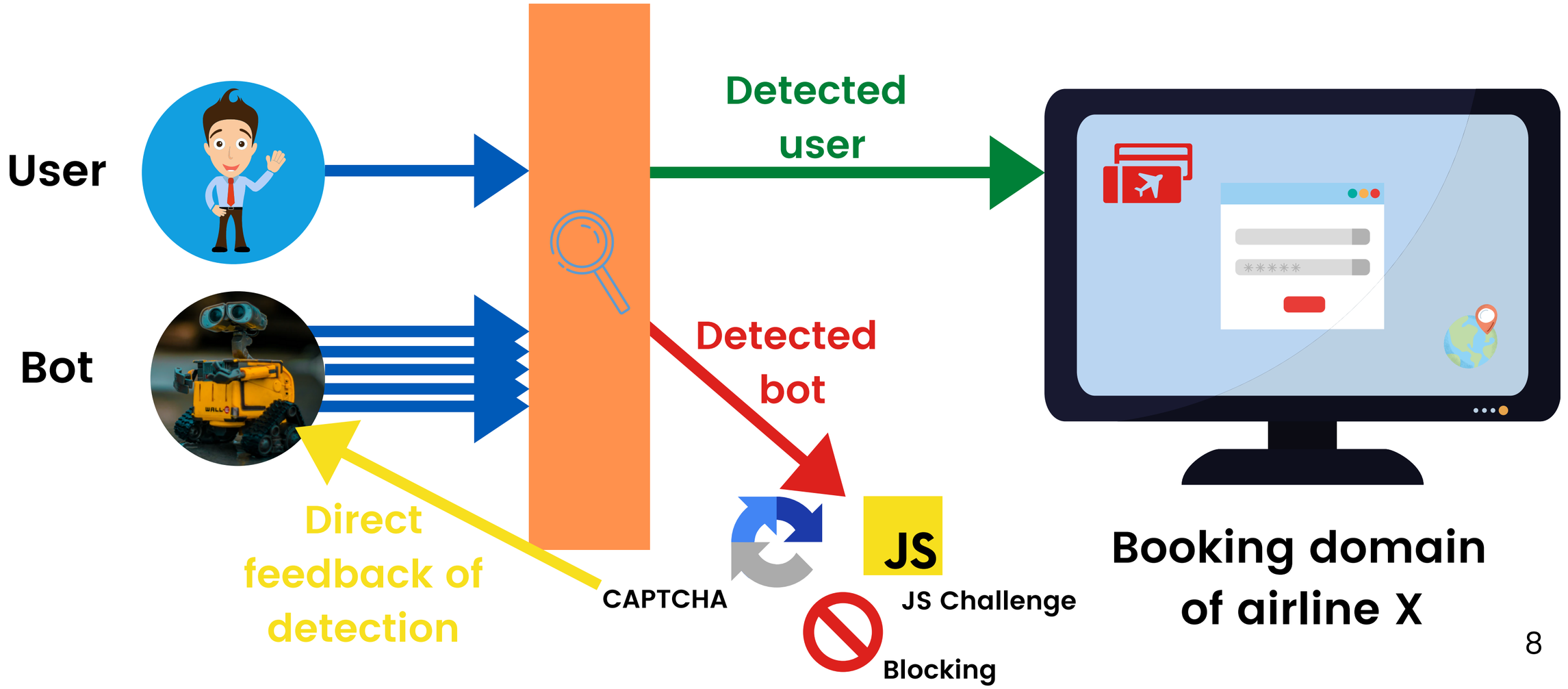
Anti-bot solutions



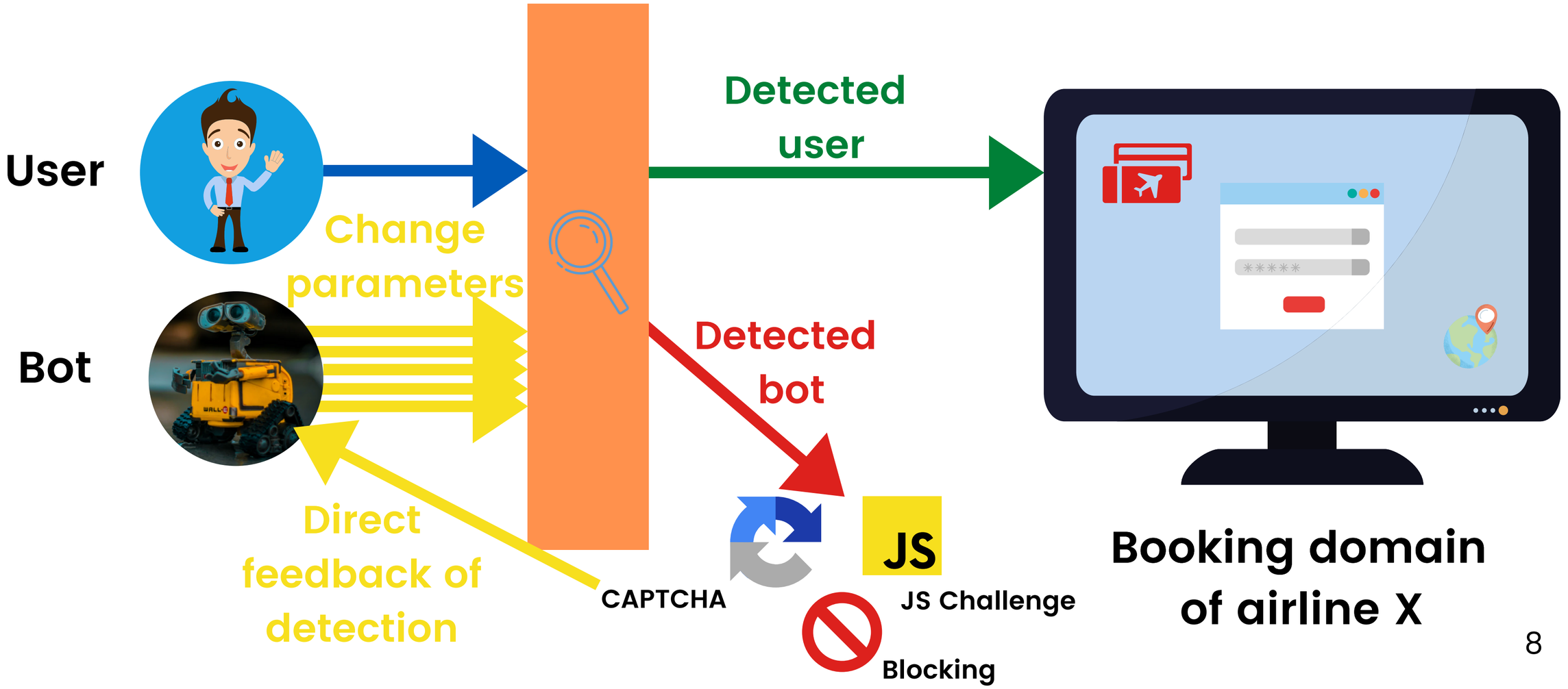
Anti-bot solutions



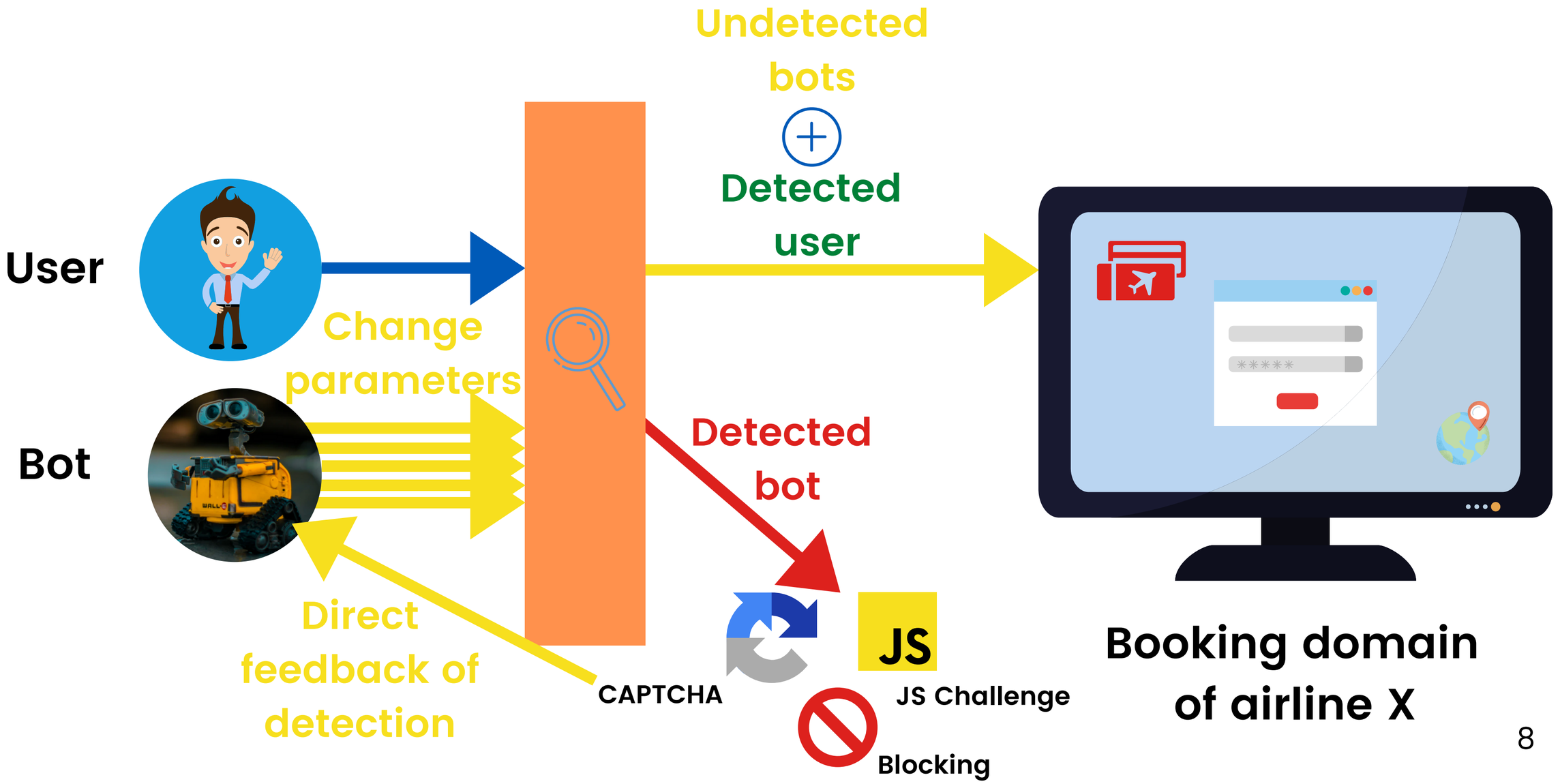
Anti-bot solutions



Anti-bot solutions



Anti-bot solutions



What 3rd parties works tell us

Blocked bots die

- Continuous verification of stealthiness and efficacy of the bots
- Rapid modification to avoid detection

Harnessed information is verified

- Continuous verification of the correctness of the information
- Incorrect information is a direct feedback they have been detected

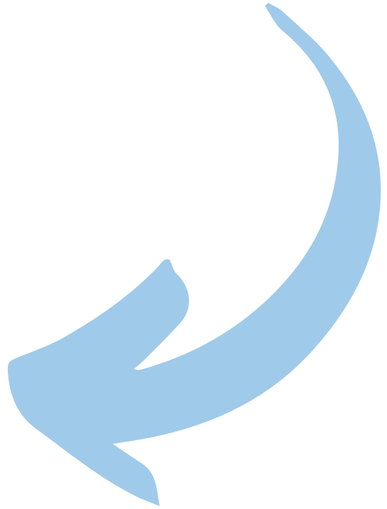
New approach

New approach

Prevent bots to know they have been detected & save costs for the provider

New approach

Prevent bots to know they have been detected & save costs for the provider



Provide bots incorrect but plausible answers

Our discoveries

1. Identification of specific super stealthy bots, with extreme distribution of activity (one request per IP in most cases)

Our discoveries

1. Identification of specific super stealthy bots, with extreme distribution of activity (one request per IP in most cases)

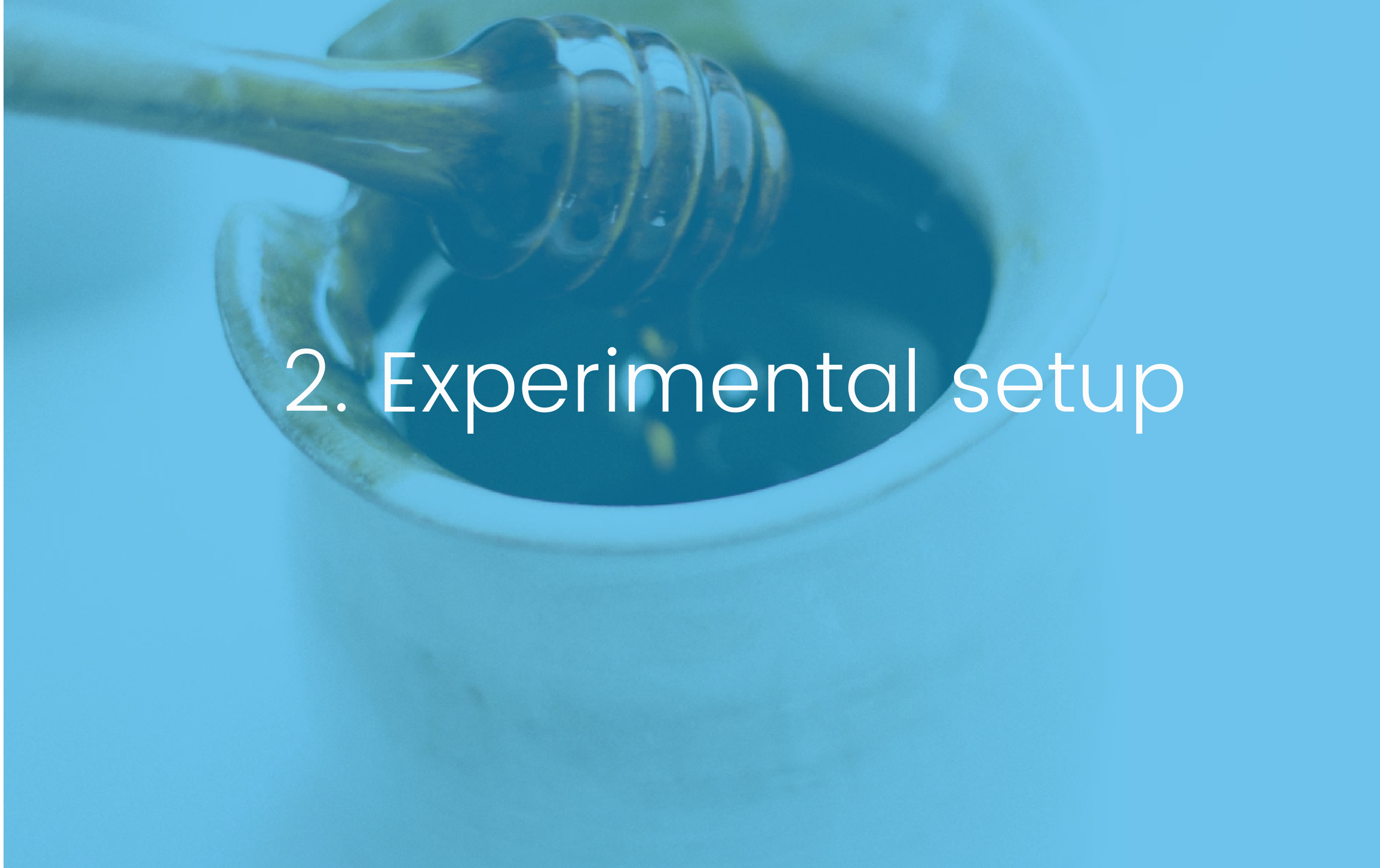
2. Identification of a behavioral pattern among these requests, which brings hope for another form of detection

Our discoveries

1. Identification of specific super stealthy bots, with extreme distribution of activity (one request per IP in most cases)

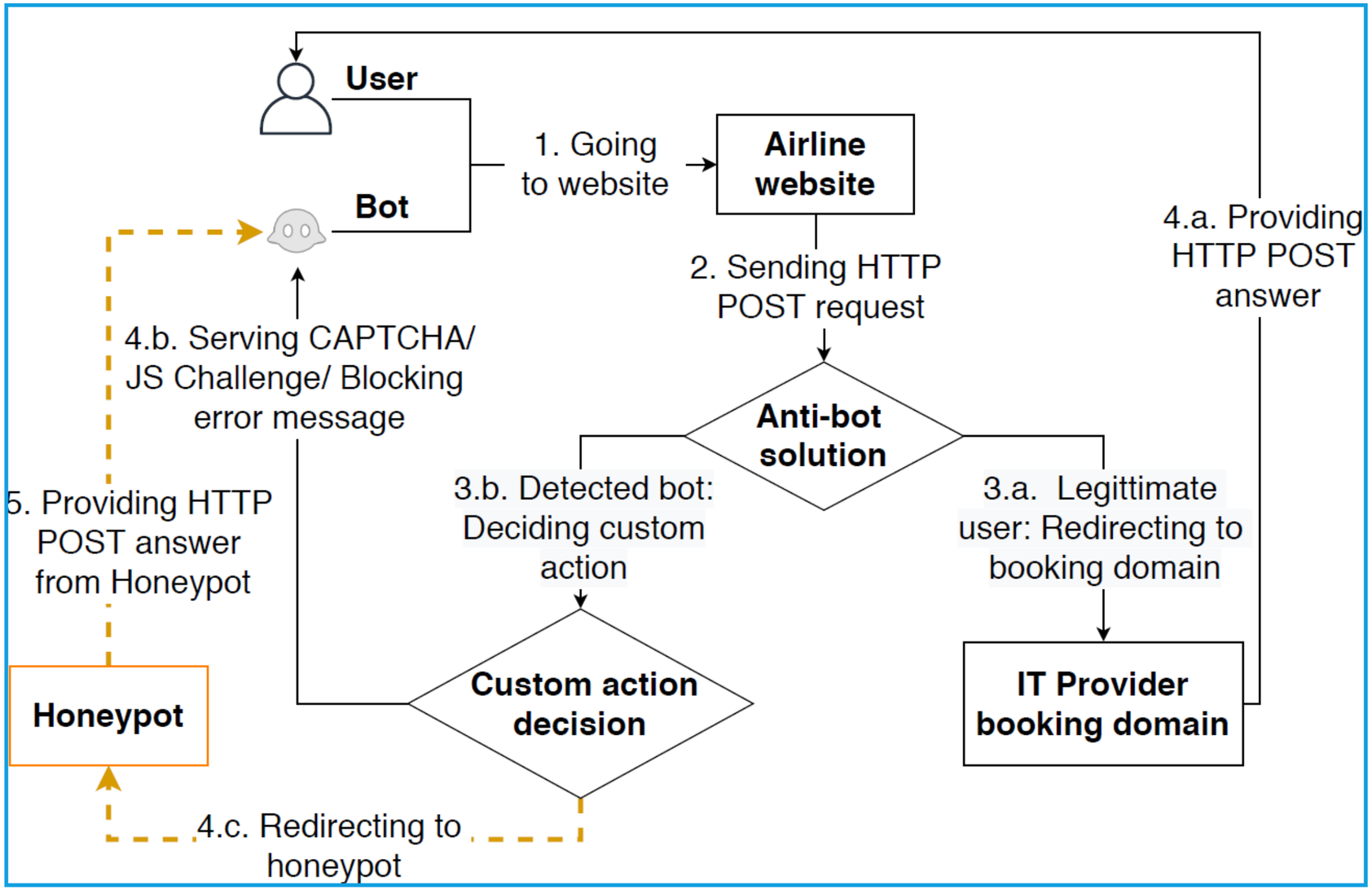
2. Identification of a behavioral pattern among these requests, which brings hope for another form of detection

3. Success in providing inaccurate information without being detected

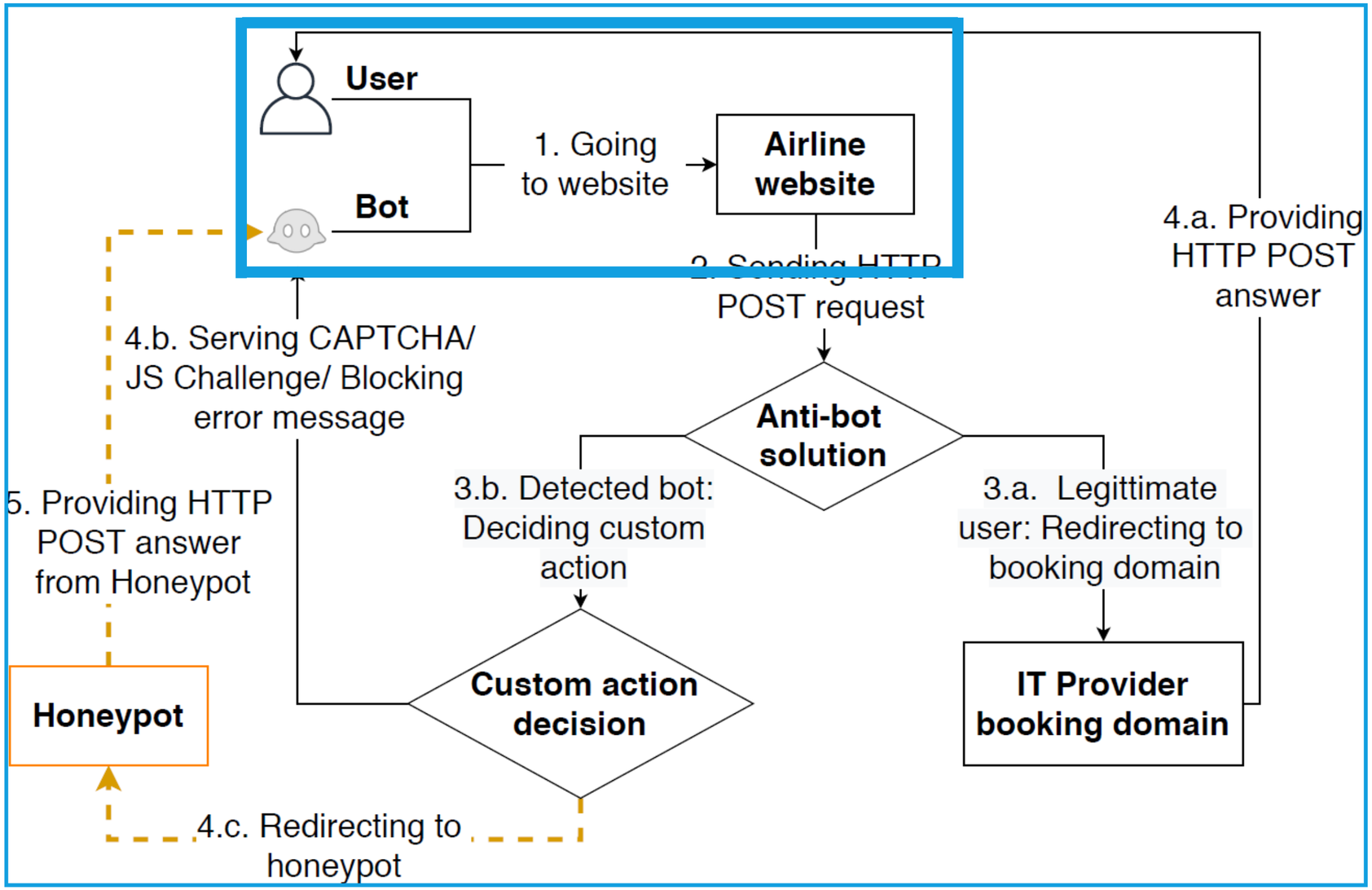


2. Experimental setup

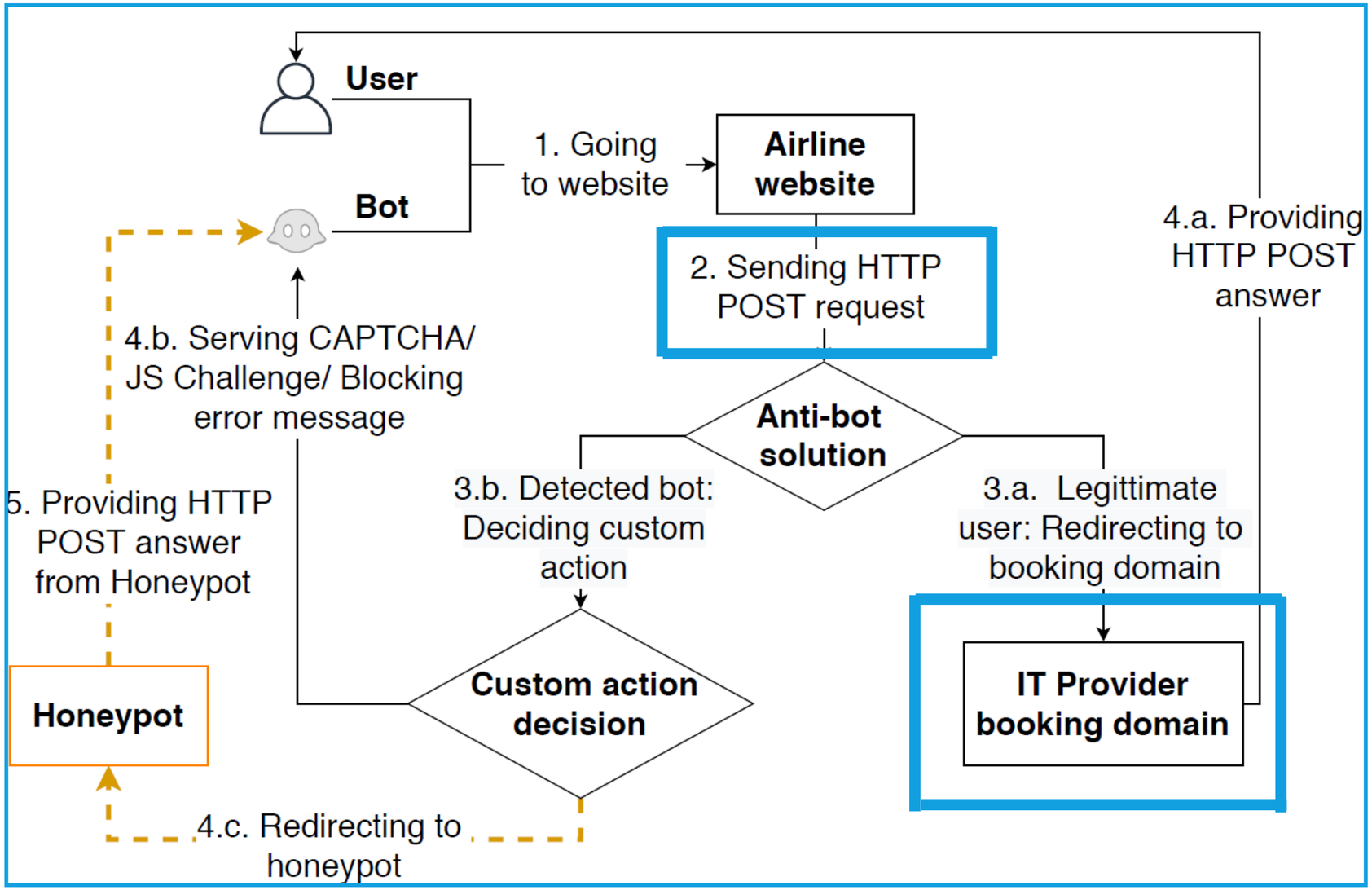
The architecture



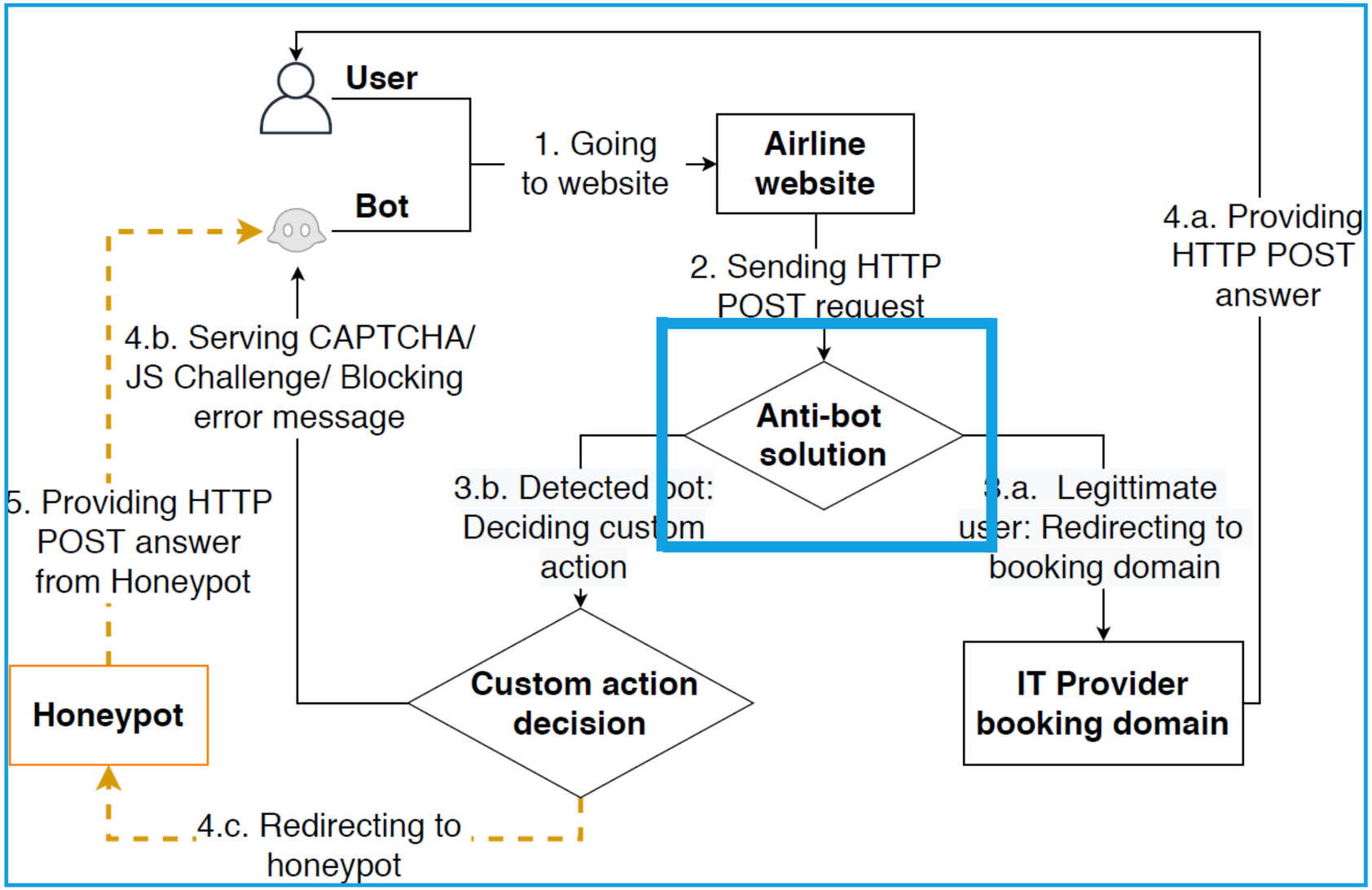
The architecture



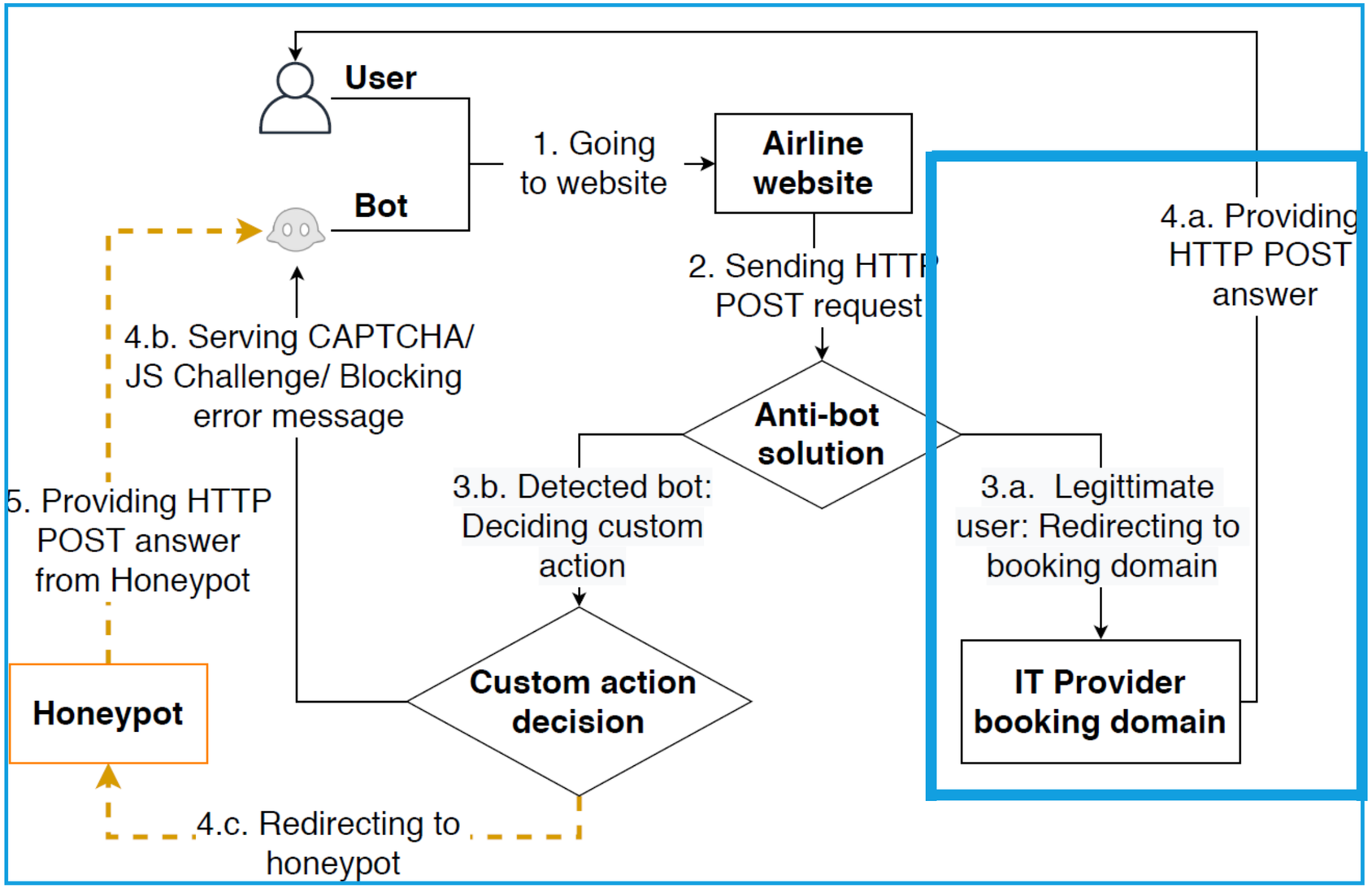
The architecture



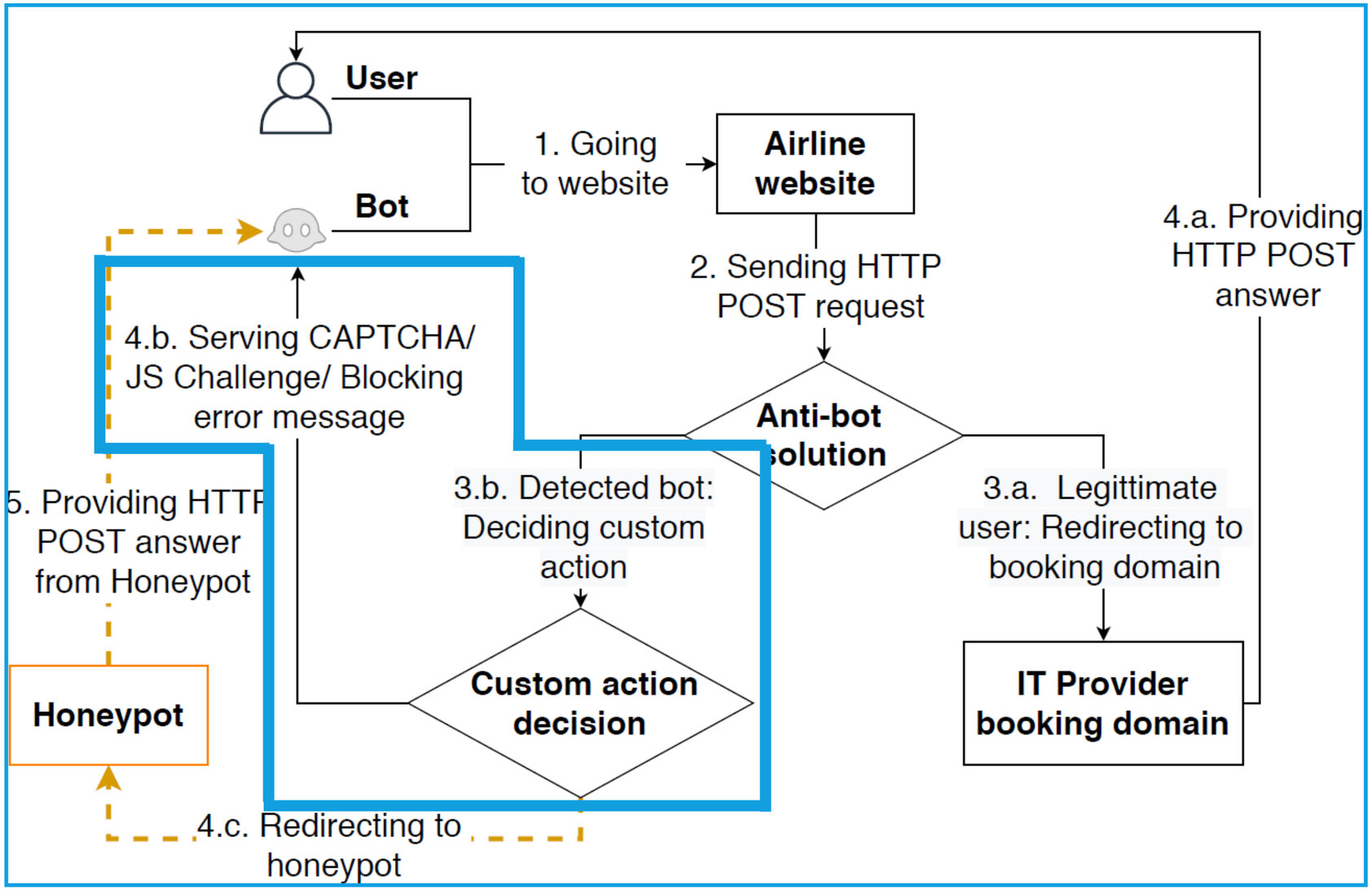
The architecture



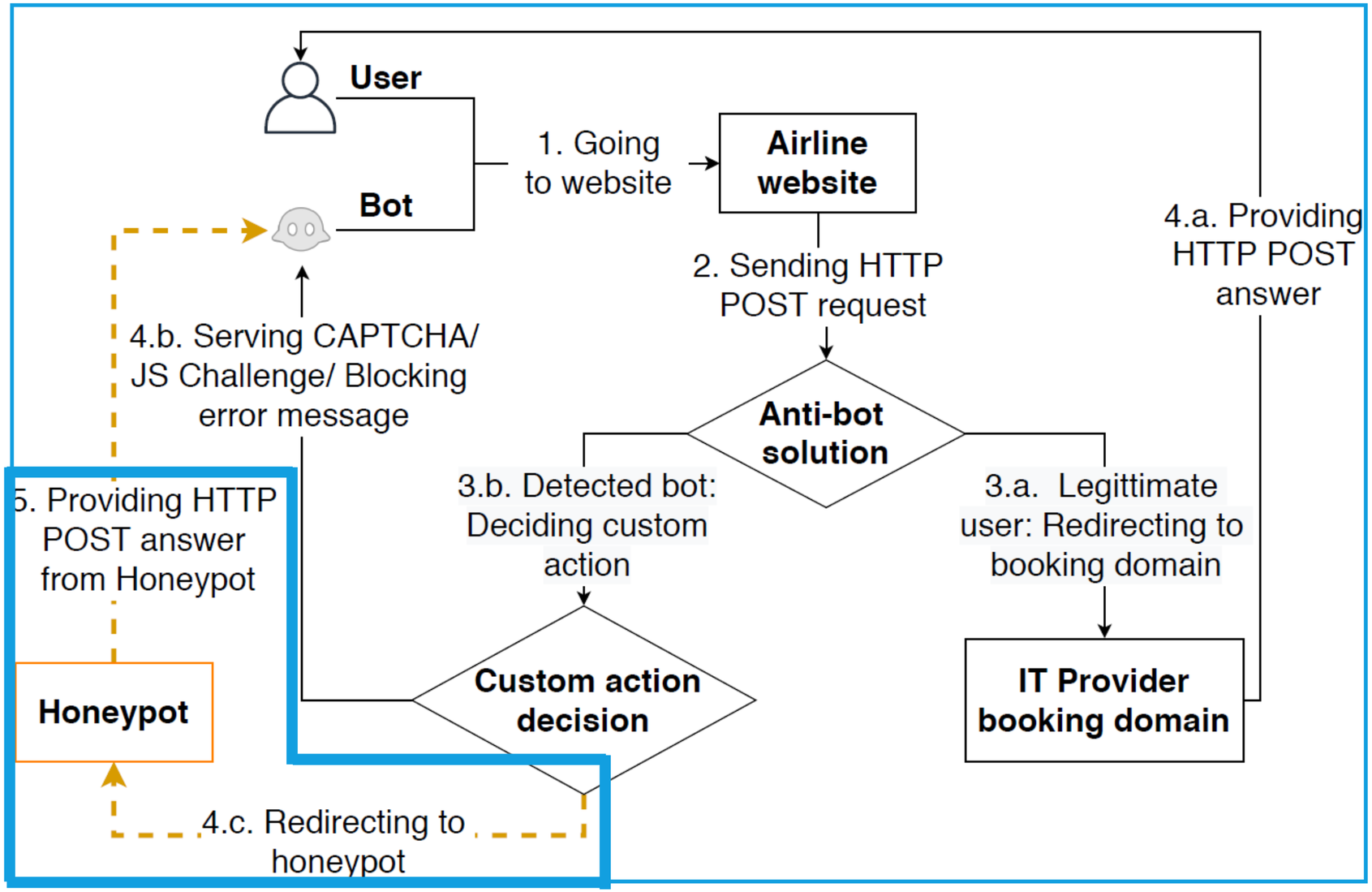
The architecture



The architecture



The architecture



Some numbers...

Pilot airline

Ticketing system: 1 million requests per day, 40% detected as coming from bad bots

Some numbers...

Pilot airline

Ticketing system: 1 million requests per day, 40% detected as coming from bad bots

Chosen bot

40-minutes daily time window of activity, no bookings, 400-500 requests per day

Pricing strategy

After 3 days, modification of fares: increase 10% of the requests by 5%

Goal: understanding if the bot master was checking for anomalies in the price

3. Results and discussion

General analysis

- Experiment running for 56 days (interruption linked with COVID-19 restrictions on flights)

General analysis

- Experiment running for 56 days (interruption linked with COVID-19 restrictions on flights)
- Reception of 22,991 HTTP request at the Honeypot

General analysis

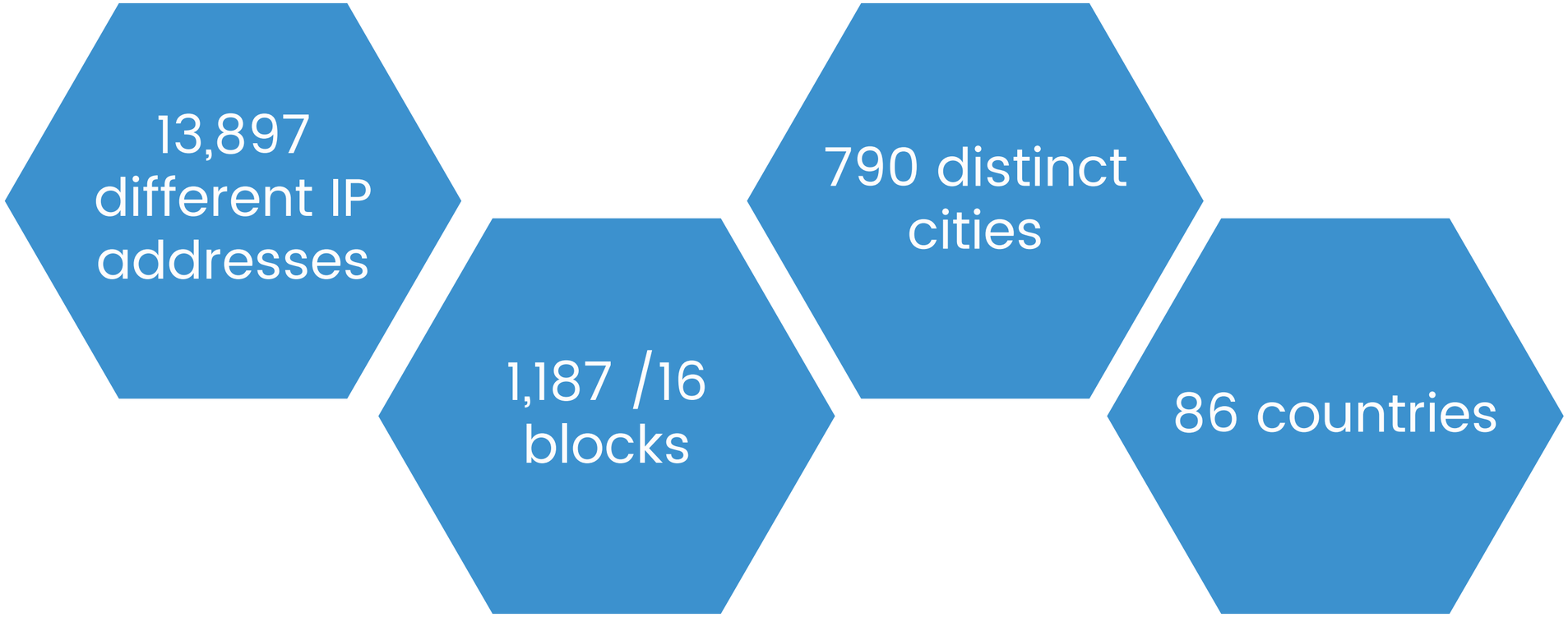
- Experiment running for 56 days (interruption linked with COVID-19 restrictions on flights)
- Reception of 22,991 HTTP requests at the Honeypot
- No change of behavior from before and during the experiment

Lessons learned modifying values

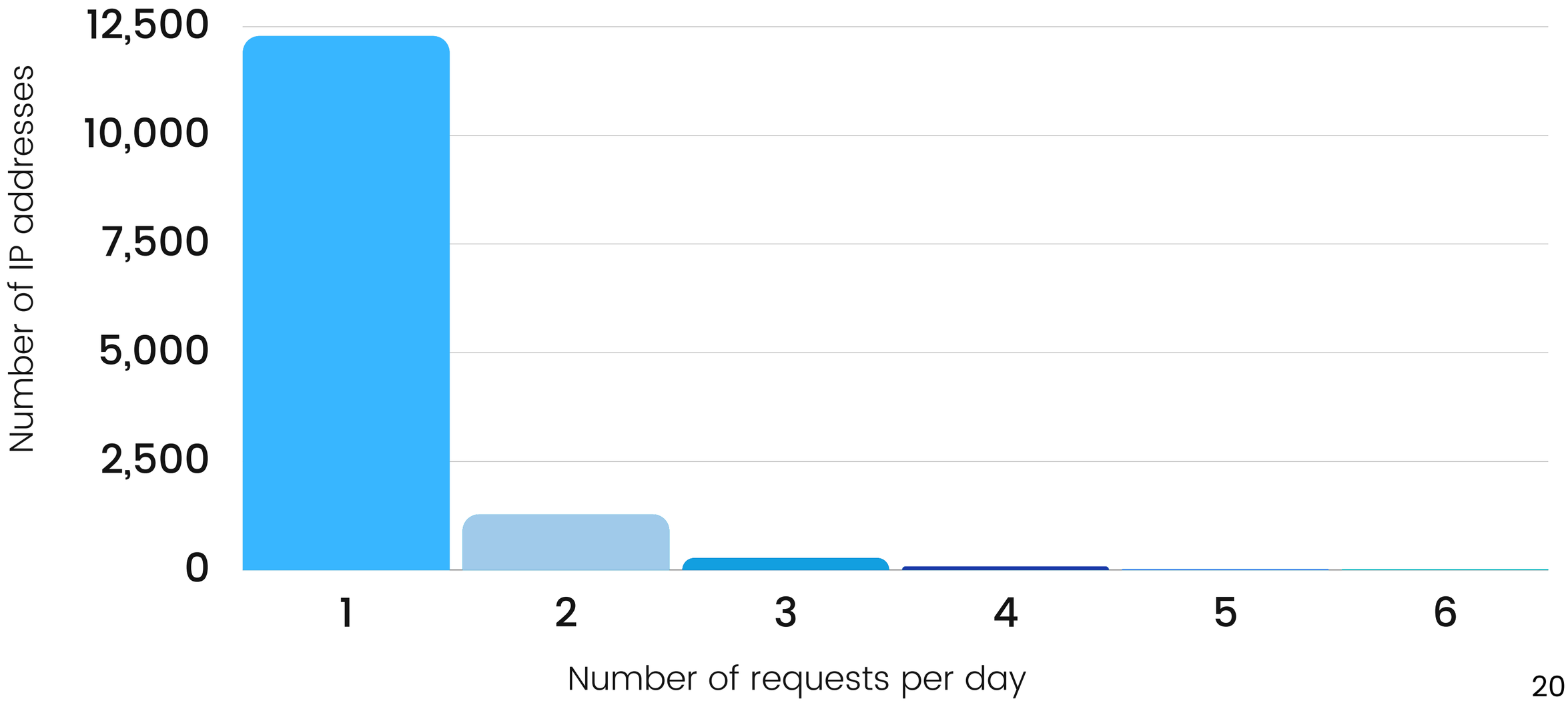
No ground truth to compared returned values

Plausibility check not sophisticated enough for small changes

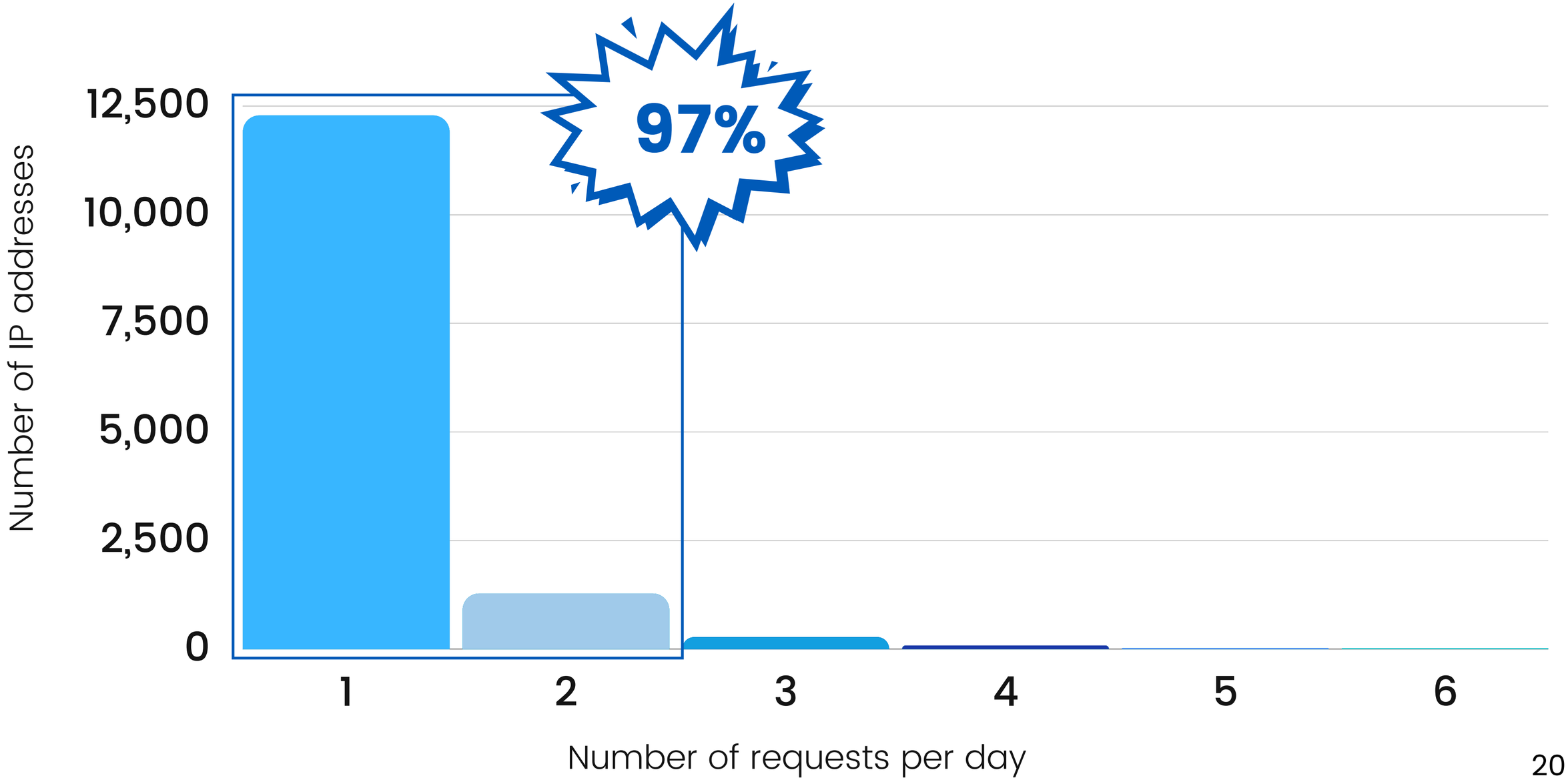
IP addresses study



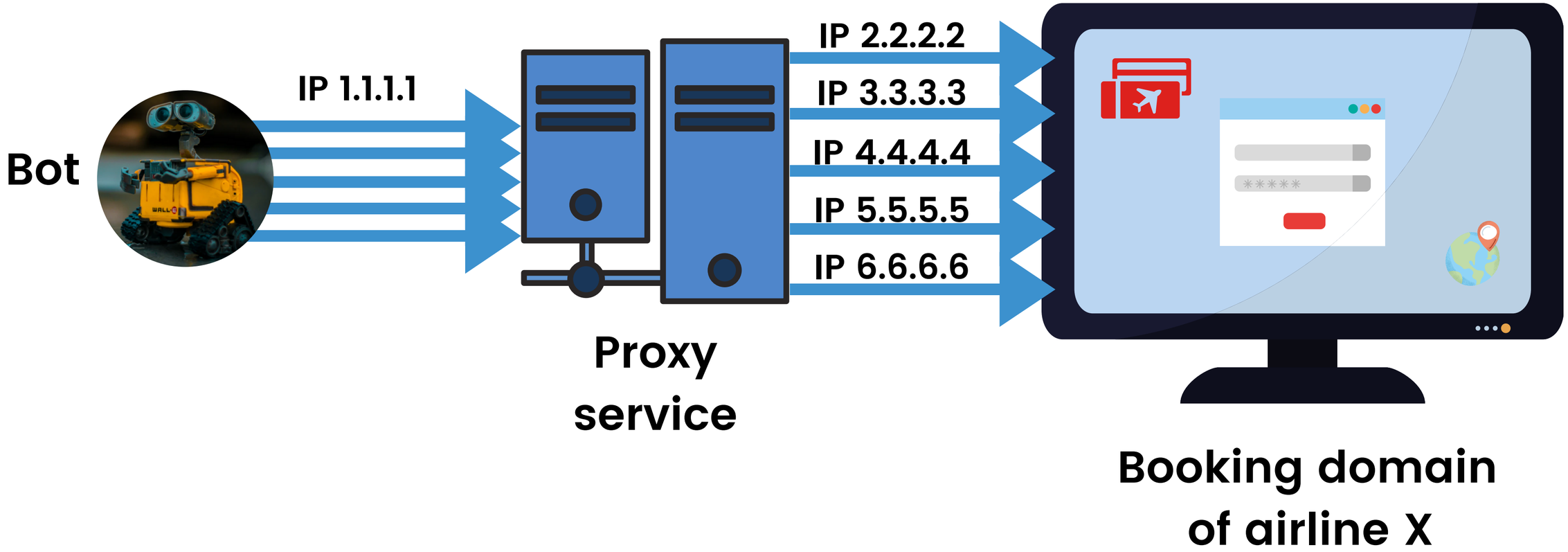
IP addresses study



IP addresses study



Residential proxy as a service



Proxy services IPs

Proxy services offer millions of IP addresses

Proxy services IPs

Proxy services offer millions of IP addresses

BUT

Proxy services IPs

Proxy services offer millions of IP addresses

BUT

32% of the IPs appeared on more than one day!

TOR network

Date	Number of TOR IPs
2020-02-14	12
2020-02-15	24
2020-02-18	20
2020-02-19	40
2020-02-24	12

TOR network

Date	Number of TOR IPs
2020-02-14	12
2020-02-15	24
2020-02-18	20
2020-02-19	40
2020-02-24	12

Bots tried to take advantage of TOR nodes but they did not continue

Behavioral analysis

Behavioral analysis

- 51,5% of requests for return flights

Behavioral analysis

- 51,5% of requests for return flights
- Return flights: 7 days period

Behavioral analysis

- 51,5% of requests for return flights
- Return flights: 7 days period
- Only 25 combination of departure and arrival airports, small fraction of the airline's offer

Time interval

- Time Interval=Departure date - Date of the request

Time interval

- Time Interval=Departure date - Date of the request
- Value between 0 and 14 days or 21, 30, 45, 60, 90, 120 days

Time interval

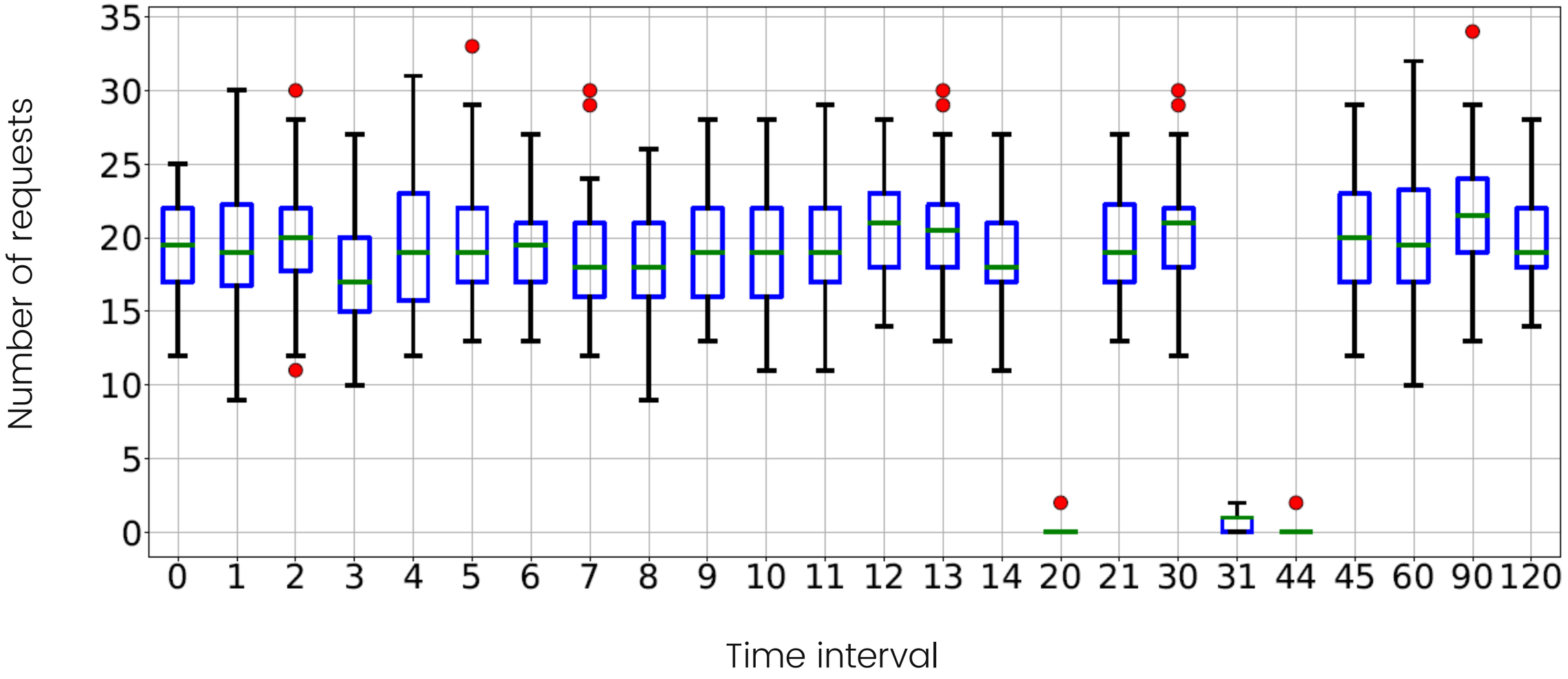
- Time Interval=Departure date - Date of the request
- Value between 0 and 14 days or 21, 30, 45, 60, 90, 120 days
- Only 0.2% of the requests exhibit different values but out of the 40 minutes daily time window

Distribution among segments

Segment=combination of one way/return flight, departure and arrival location

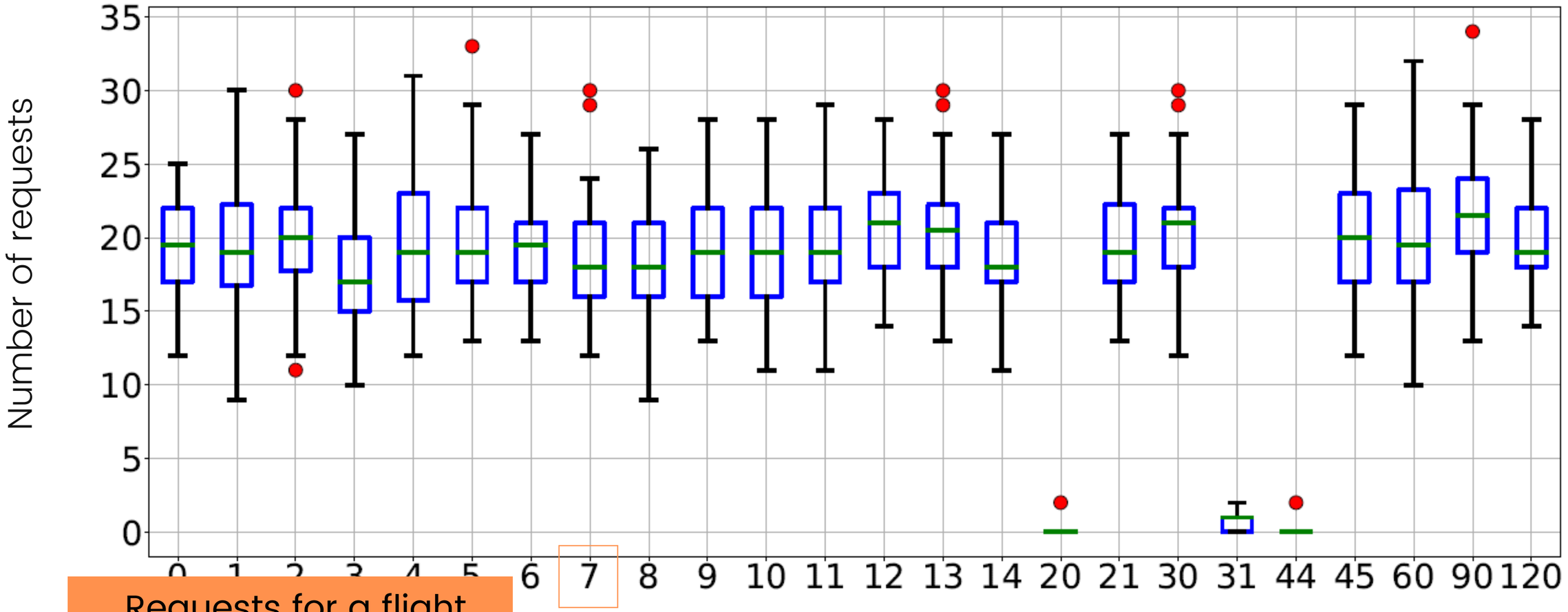
Distribution among segments

Segment=combination of one way/return flight, departure and arrival location



Distribution among segments

Segment=combination of one way/return flight, departure and arrival location

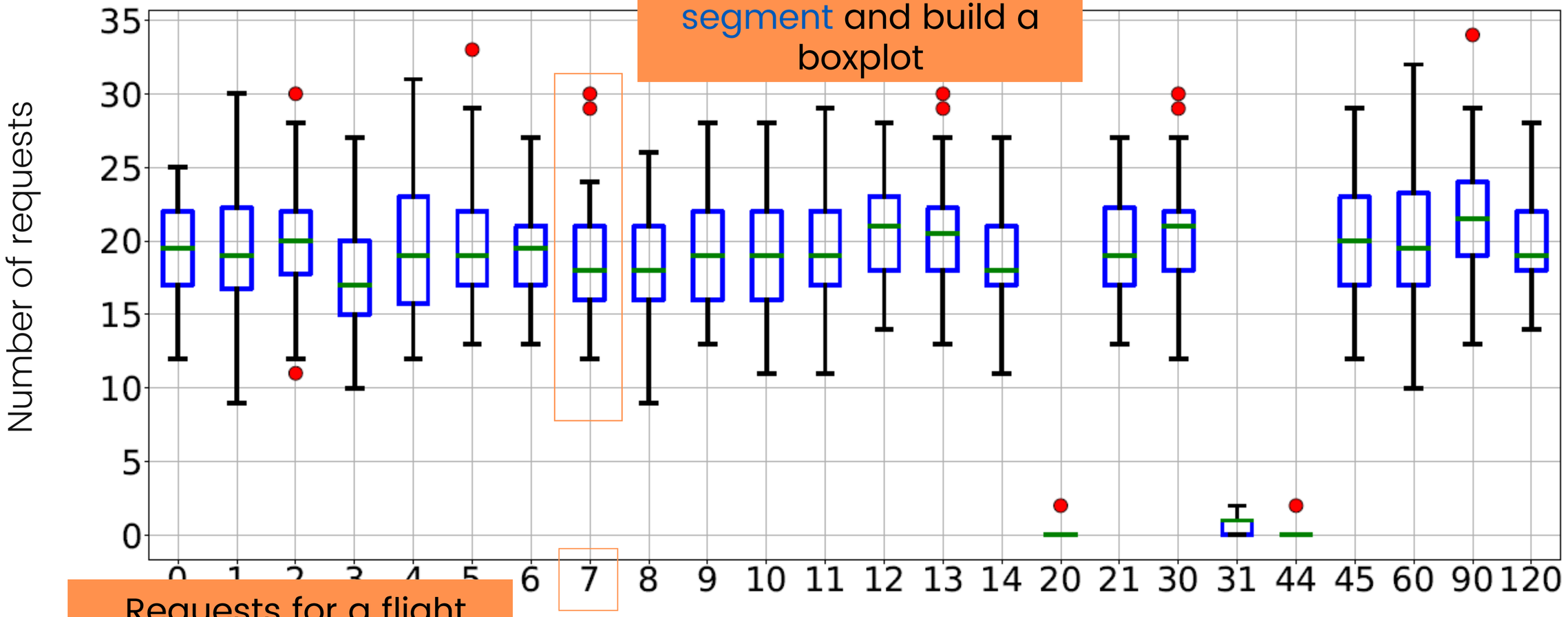


Requests for a flight departing 7 days after the request date

Distribution among segments

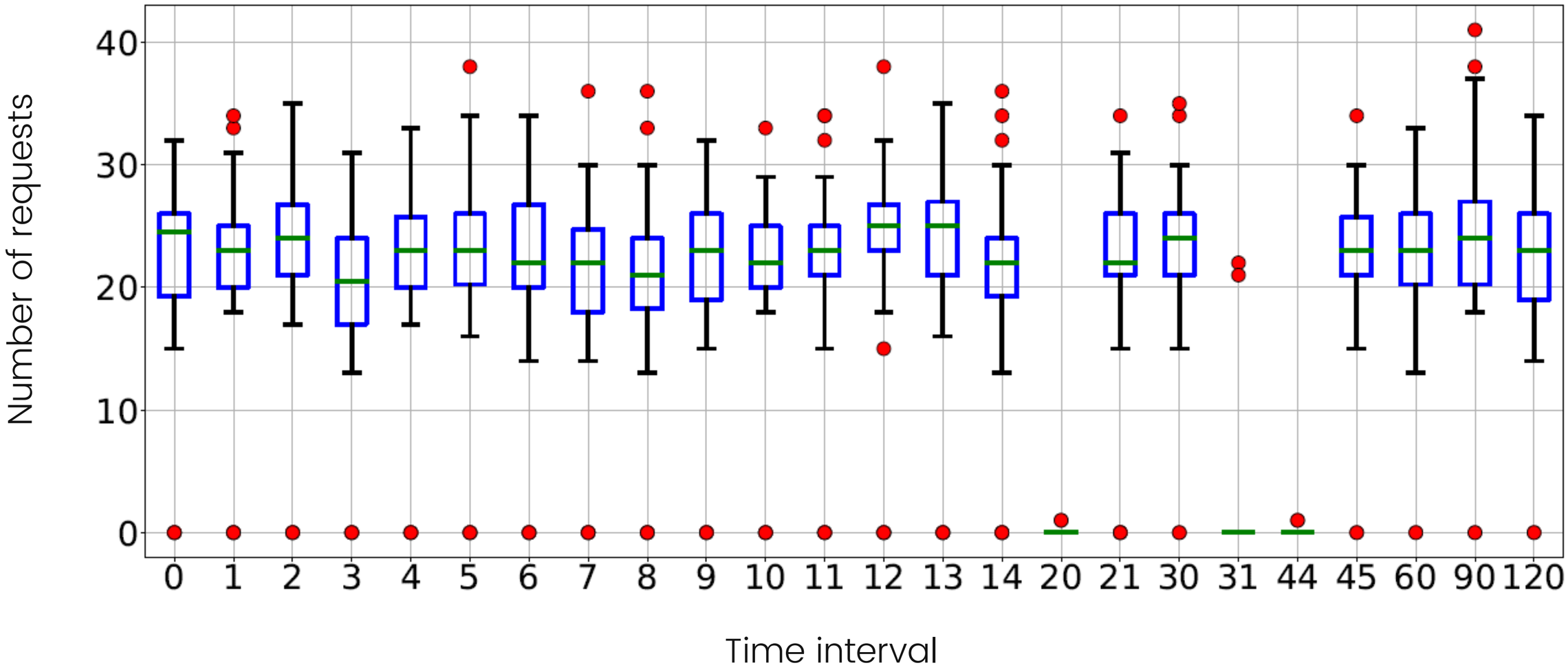
Segment=combination of one way/return flight, departure and arrival location

Divide them for each segment and build a boxplot



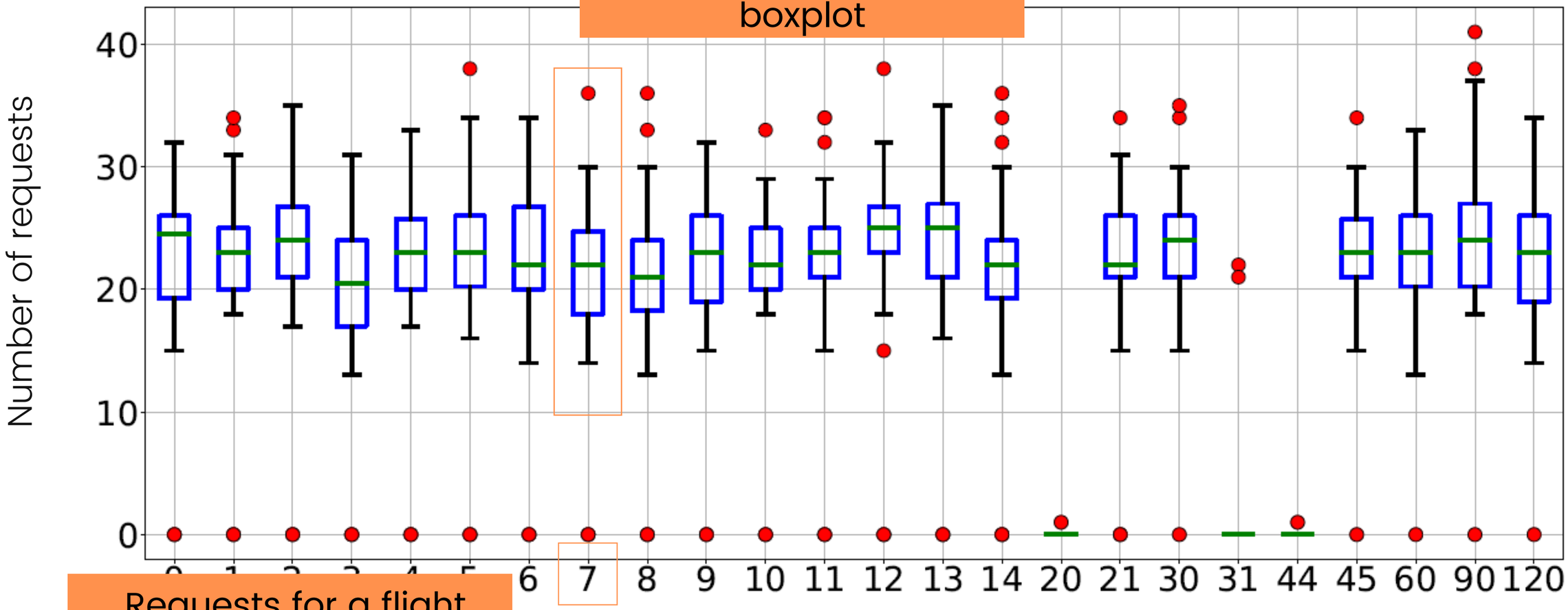
Requests for a flight departing 7 days after the request date

Distribution for request date



Distribution for request date

Divide them for date in which the request was made and build a boxplot



Requests for a flight departing 7 days after the request date

Tuples statistics

- 4-tuples made of:
  - Departure airport
  - Arrival airport
  - Time interval
  - Type of flight (one way/return)

Tuples statistics

- 4-tuples made of:
  - Departure airport
  - Arrival airport
  - Time interval
  - Type of flight (one way/return)
- 982 distinct tuples vs 410 average daily requests

Tuples statistics

- Each tuple is asked on average 23.41 times

Tuples statistics

- Each tuple is asked on average 23.41 times
- Average number of days a tuple was requested: 22.85 days

Tuples statistics

- Each tuple is asked on average 23.41 times
- Average number of days a tuple was requested: 22.85 days
- Generally tuples are asked once a day at most

Tuples statistics

- Each tuple is asked on average 23.41 times
- Average number of days a tuple was requested: 22.85 days
- Generally tuples are asked once a day at most
- 20% of tuples are asked at least more than once a day
  - All requests done in little span of time
  - Maximum difference: 5 minutes and 30 seconds

Tuples statistics

- Each tuple is asked on average 23.41 times
- Average number of days a tuple was requested: 22.85 days
- Generally tuples are asked once a day at most
- 20% of tuples are asked at least more than once a day
  - All requests done in little span of time
  - Maximum difference: 5 minutes and 30 seconds

Check for consistency, but not sophisticated enough

4. Conclusions and future work

What we have done

What we have done

Shown bots which send one request per day

What we have done

Shown bots which send one request per day

Served modified prices without being detected for 56 days

What we have done

Shown bots which send one request per day

Seen 32% of IPs were reused

Served modified prices without being detected for 56 days

What we have done

Shown bots which send one request per day

Served modified prices without being detected for 56 days

Seen 32% of IPs were reused

Found behavioral pattern that gives confidence they are all under the same bot master

Future works

Serve bot with cached prices

Because of:

- Repetition of requests
- Loose verification

To do:

- Sensitivity analysis

Future works

Serve bot with cached prices

Because of:

- Repetition of requests
- Loose verification

To do:

- Sensitivity analysis

New method to block IPs

To do:

- Longitudinal large-scale analysis of suspicious IPs exhibiting the same behavior

Thank You

How to reach us

Elisa Chiapponi - elisa.chiapponi@eurecom.fr

Onur Catakoglu - onur.catakoglu@amadeus.com

Olivier Thonnard - olivier.thonnard@amadeus.com

Marc Dacier - marc.dacier@eurecom.fr