

Luminati Provides Web-Transparency

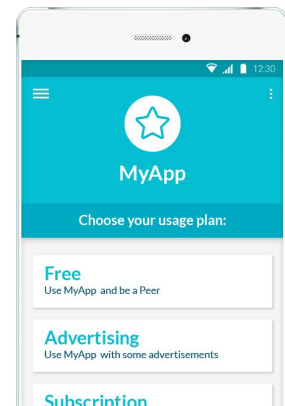


Web Scraping Proxy Management Workshop

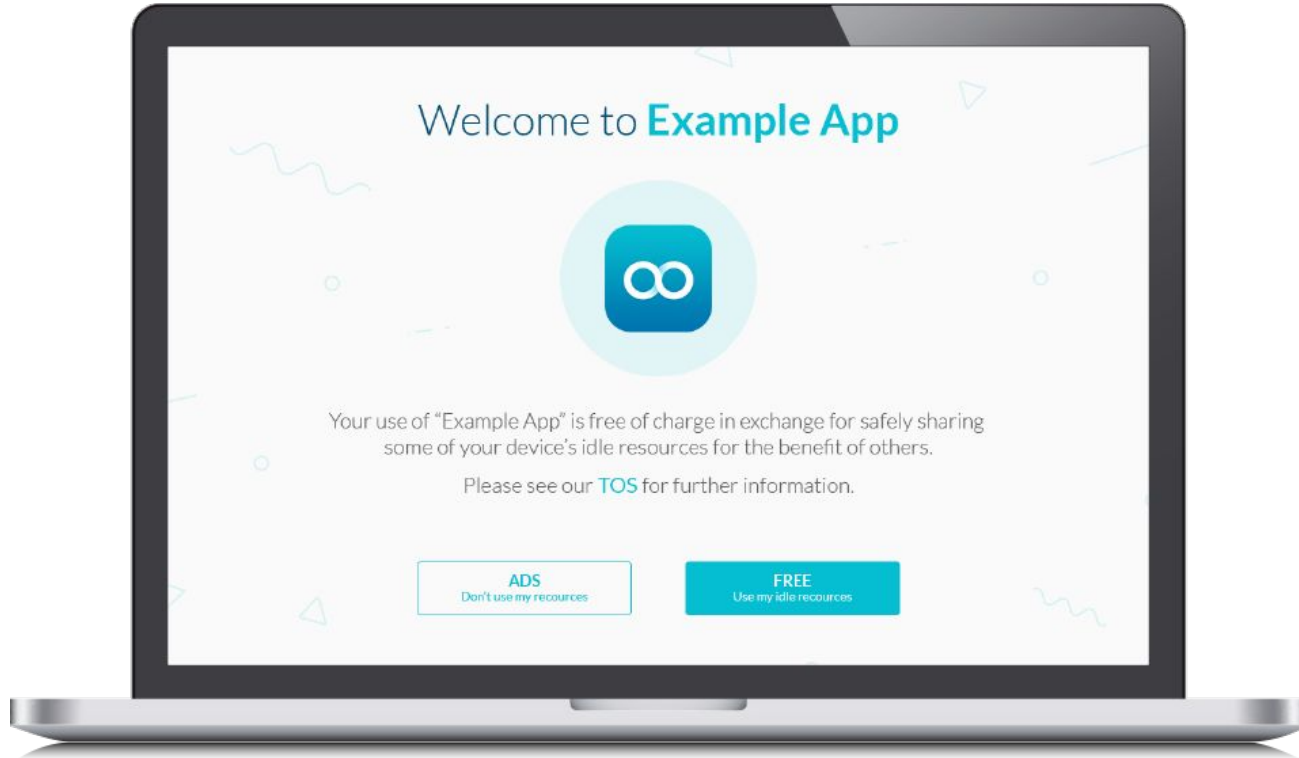
Luminati developed a global P2P network 35M+ consumers willing to help



Consumers opt-in to the network in return for free partner's application usage



How do we get users active consent?



How does it work?

We use a peer's IP address only when a device meets 3 conditions:



Sits idle

We route traffic through the device only when it's absolutely not in use



Connected to internet

We prefer to use WiFi and may use very limited cellular data



Connected to power

Device is plugged in or has enough battery power



**Businesses can now view the web, as these
35M global consumers can see it**

Luminati Proxy Networks Available



Rotating Residential Network

- ✓ 35+ million IPs
- ✓ 195 countries
- ✓ Country, City, ASN and Mobile Carrier Targeting
- ✓ 99.99% uptime



Mobile Network

- ✓ 7.5 million IPs
- ✓ 3g/4g connections
- ✓ 195 countries
- ✓ Country, City, ASN and Mobile Carrier Targeting



Datacenter Network

- ✓ 500,000 IPs
- ✓ 95 countries
- ✓ Available with country and city targeting
- ✓ Multiple IP types in a shared or dedicated pool



Static Residential (ISP) IPs

- ✓ 65k IPs
- ✓ 30+ countries
- ✓ Non-rotating
- ✓ Directly from an ISP

Crawling Network Architecture



Luminati Proxy Manager



Auto retry

Retry the request based on unwanted status codes, timeouts or failures



Limit requests

Limit the number of requests sent per hour to maximize scraping rates without getting blocked



IP rotation

Track campaigns with the click of your mouse Create rules for automated rotation of IPs



Ban IPs

Blacklist unwanted IPs to improve success rates



Multiple vendors

With the central point of control being 100% at the client side work with multiple proxy vendors with easy integration



Speed

With no limit on the number of concurrent connections, perform scraping jobs faster



Luminati

Luminati Proxy Manager



Bypass proxy

Control which requests go to the proxy provider and which bypass the proxy



Reduce bandwidth

Use the LPM's regex option to filter specific URLs and reduce traffic



Waterfall routing

Route requests through Residential, Data-center and Mobile IP networks based on custom rules to get the most cost-effective blend



Logging

Detailed logging of all requests



Success ratio metrics

Measure the success of proxies



SSL decryption

Allow debugging of SSL requests



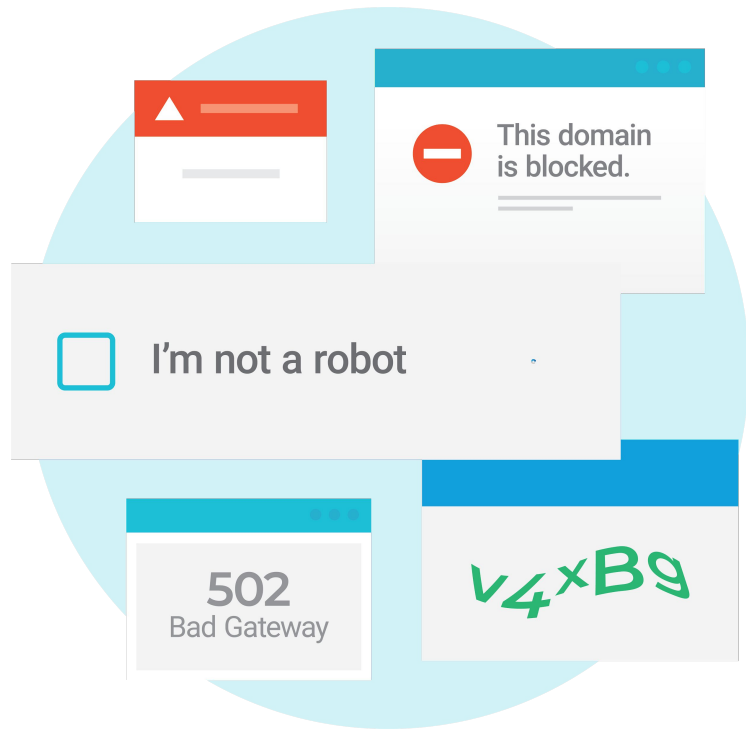
Luminati



Robot Detection



- IP reputation
- Browser headers and cookies
- Device fingerprints
- User behaviour and history
- IP leaks



- Type
- Request rate
- Account association
- Blacklisted IPs
- Inconsistencies



- User uniqueness on the web
- Users become more unique as the entropy level increases

$$H(x) = - \sum_x P(x) \cdot \log P(x)$$

$$= \sum_x \underbrace{P(x)} \cdot \underbrace{\log\left(\frac{1}{P(x)}\right)}$$

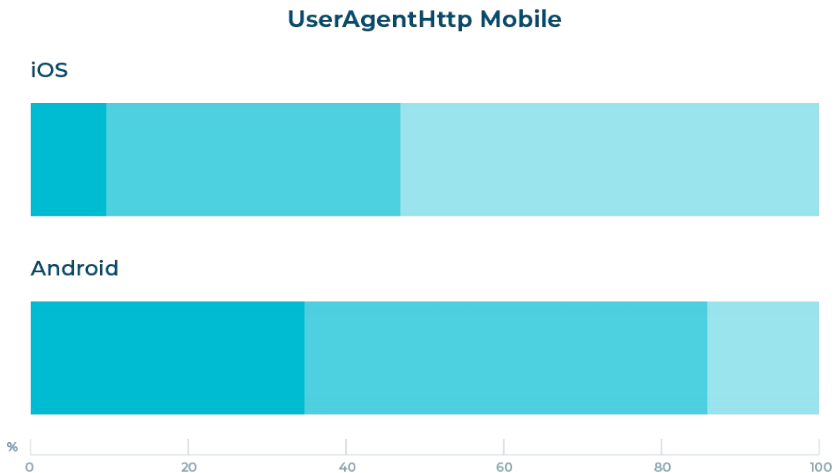
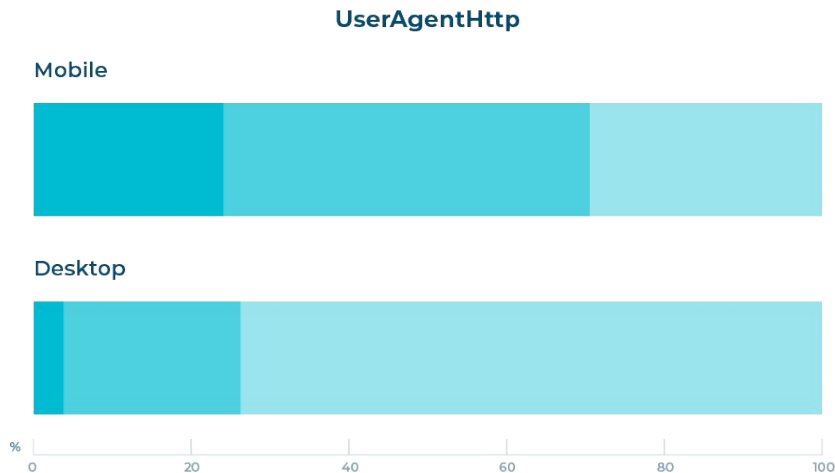
The prob. of event x

WHAT IS THIS?

Attribute	Source	Example
User agent	HTTP header	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/64.0.3282.119 Safari/537.36
Accept	HTTP header	text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
Content encoding	HTTP header	gzip, deflate, br
Content language	HTTP header	en-US,en;q=0.9
List of plugins	JavaScript	Plugin 1: Chrome PDF Plugin. Plugin 2: Chrome PDF Viewer. Plugin 3: Native Client. Plugin 4: Shockwave Flash...
Cookies enabled	JavaScript	yes
Use of local/session storage	JavaScript	yes
Timezone	JavaScript	-60 (UTC+1)
Screen resolution and color depth	JavaScript	1920x1200x24
List of fonts	Flash or JS	Abyssinica SIL,Aharoni CLM,AR PL UMing CN,AR PL UMing HK,AR PL UMing TW...
List of HTTP headers	HTTP headers	Referer X-Forwarded-For Connection Accept Cookie Accept-Language Accept-Encoding User-Agent Host
Platform	JavaScript	Linux x86_64
Do Not Track	JavaScript	yes
Canvas	JavaScript	Cwm fjordbank glyphs vext quiz, ☹ Cwm fjordbank glyphs vext quiz, 😊
WebGL Vendor	JavaScript	NVIDIA Corporation
WebGL Renderer	JavaScript	GeForce GTX 650 Ti/PCIe/SSE2
Use of an ad blocker	JavaScript	yes

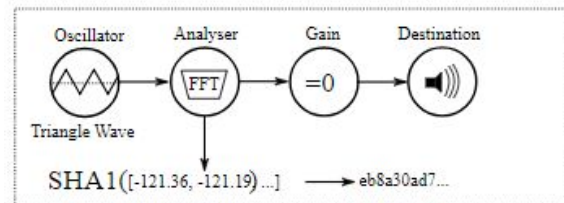
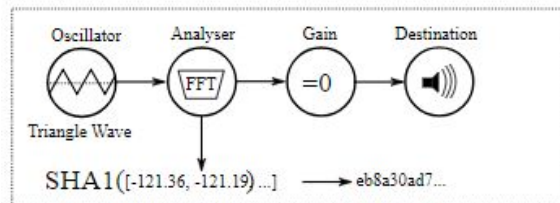
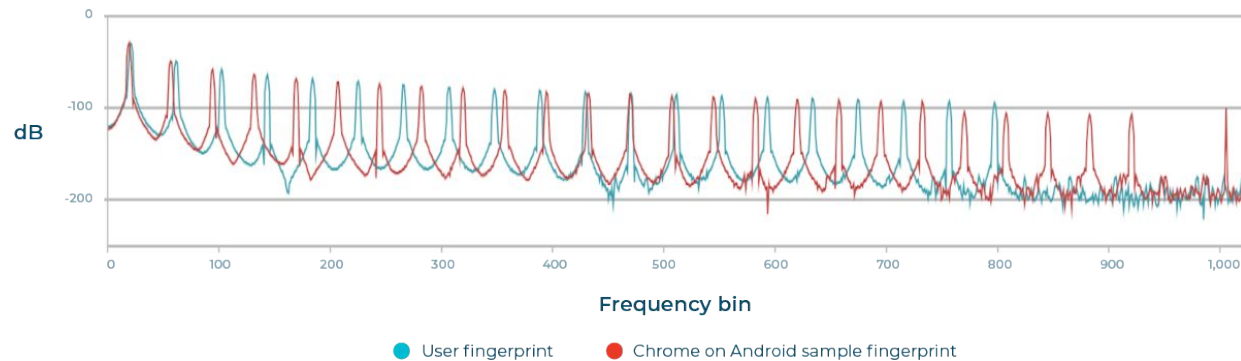
Desktop <> Mobile

Android <> iOS



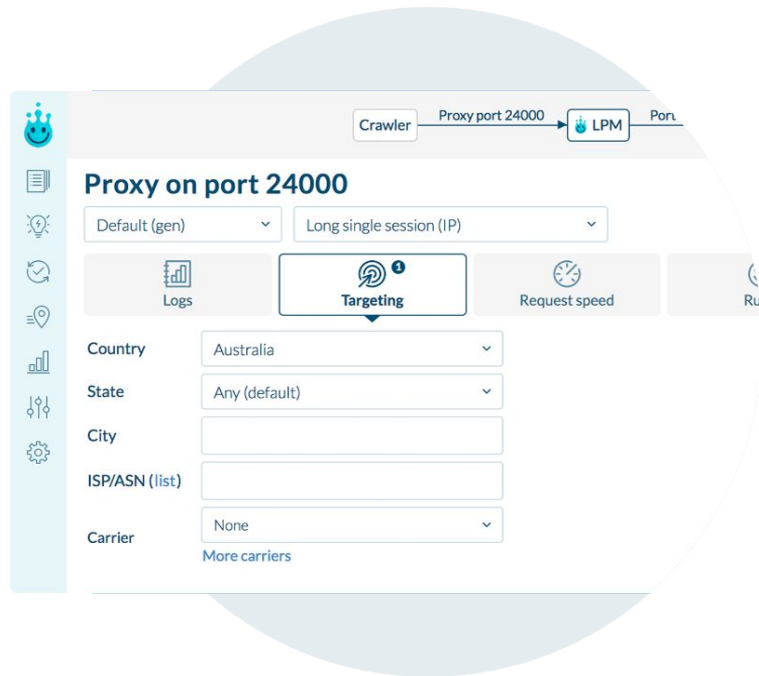
AudioContext properties:

```
{  
  "ac-baseLatency": 0.01,  
  "ac-sampleRate": 44100,  
  "ac-state": "running",  
  "ac-maxChannelCount": 2,  
  "ac-numberOfInputs": 1,  
  "ac-numberOfOutputs": 0,  
  "ac-channelCount": 2,  
  "ac-channelCountMode": "explicit",  
  "ac-channelInterpretation": "speakers",  
  "an-fftSize": 2048,  
  "an-frequencyBinCount": 1024,  
  "an-minDecibels": -100,  
  "an-maxDecibels": -30,  
  "an-smoothingTimeConstant": 0.8,  
  "an-numberOfInputs": 1,  
  "an-numberOfOutputs": 1,  
  "an-channelCount": 2,  
  "an-channelCountMode": "max",  
  "an-channelInterpretation": "speakers"  
}
```

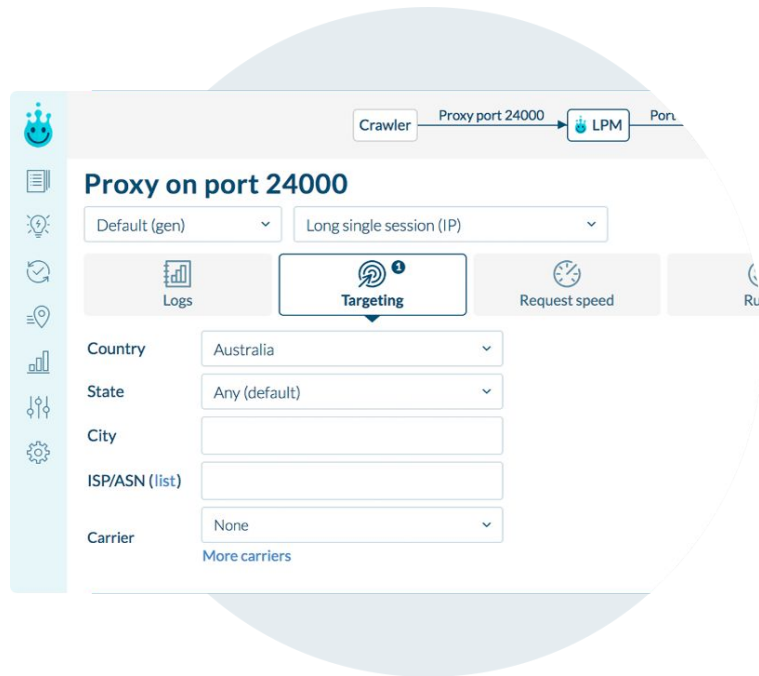


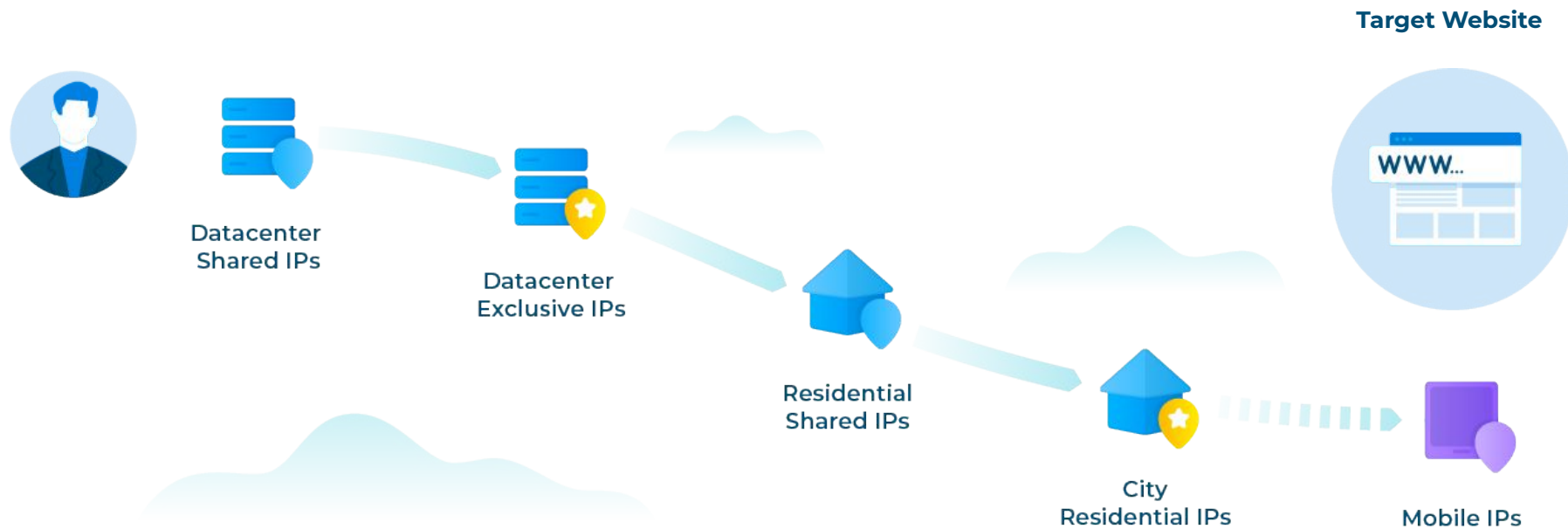


- Request rate
- Country and city discovery
- Managing headers and fingerprints
- Internet protocol version (i.e HTTP/2)
- Persistence



- By using different IPs, geo's and networks
 - Waterfall routing
- Auto retry and banning IPs
 - Optimize IP cooling period
- New IP and fingerprints
 - Error code, ReCaptcha, cloaked





Just make a simple request and let us handle the rest!

```
Curl --proxy <username>:<password>@unblock.zproxy.lum-superproxy.io:22225 https://example.com
```



Automatic Retry

Automatically retries request upon a failed response



Network Rotation

Route through multiple networks automatically (waterfall)



Manages Headers

Automatic header management based on site requirements



Manages Cookies

IP priming and cookie management based on overall request load



Country Discovery

Chooses the right country IP based on your request or target site



Detection and Matching

Ensures the response is of the right content type

Just make a simple request and let us handle the rest!

```
Curl --proxy <username>:<password>@unblock.zproxy.lum-superproxy.io:22225 https://example.com
```



Automatic Retry

Automatically retries request upon a failed response



Network Rotation

Route through multiple networks automatically (waterfall)



Manages Headers

Automatic header management based on site requirements



Manages Cookies

IP priming and cookie management based on overall request load



Country Discovery

Chooses the right country IP based on your request or target site



Detection and Matching

Ensures the response is of the right content type

Proxy Manager:

Url: <http://zagent1745.luminati.io:22999>

