Unveiling the Nexus: Harnessing IoT Ecosystems for Evading Internet Censorship

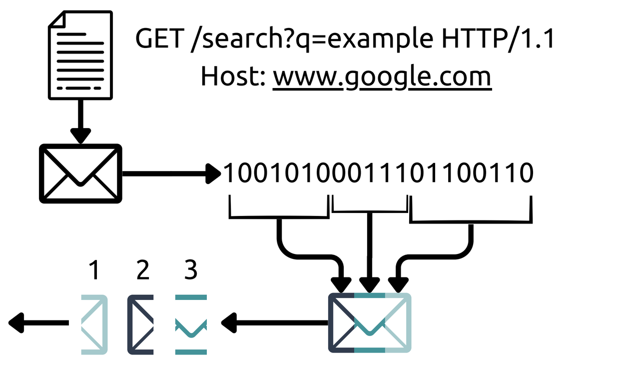
UROP
UNIVERSITY OF MICHIGAN

Wyatt Ashley Patrick Tser Jern Kon Yining Shi

ABSTRACT

- In the constantly growing realm of technology, the methods of Internet censorship by governing bodies is continually progressing. This situation poses an ongoing challenge reminiscent of what many call a "cat-and-mouse" game, where censors adjust their strategies alongside technological progress, while individuals strive to remain ahead.
- Concurrently, the Internet of Things (IoT) field within Autonomous Systems (AS) introduces a new dimension, offering vast untapped diverse computational resources.

DISECTION

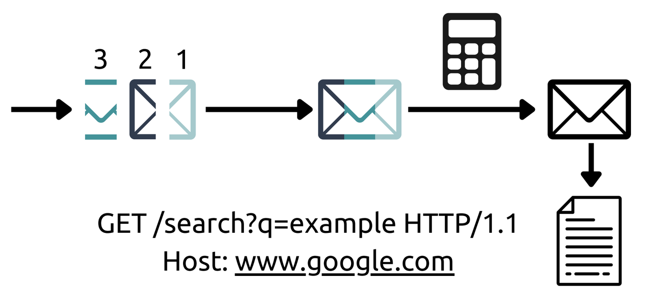


Randomness

- Each packet is split into random length chunks down to the bit level. Padding is added as needed or random values that sum to eight will be used so that the total aligns with a byte.

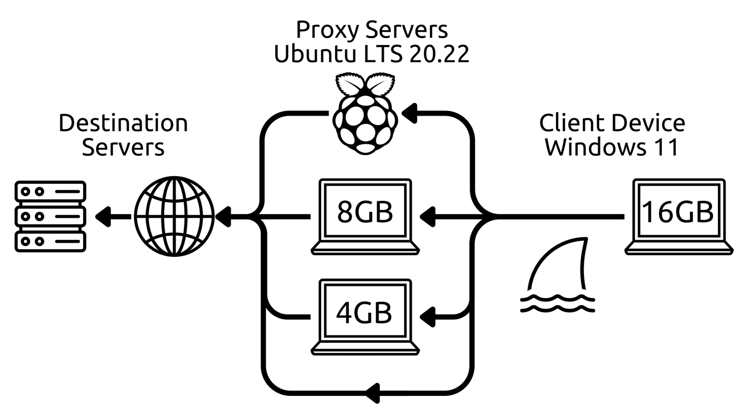
Reconstruction

- In order to reconstruct the packets IoT devices will communicate on the arrival of each packet with an algorithm to determine the correct ordering of the reconstruction.
- Furthermore, once all of the parts are recollected and in order recalculation of the packet checksum and transmission time must be completed to make a unrecognizable packet.

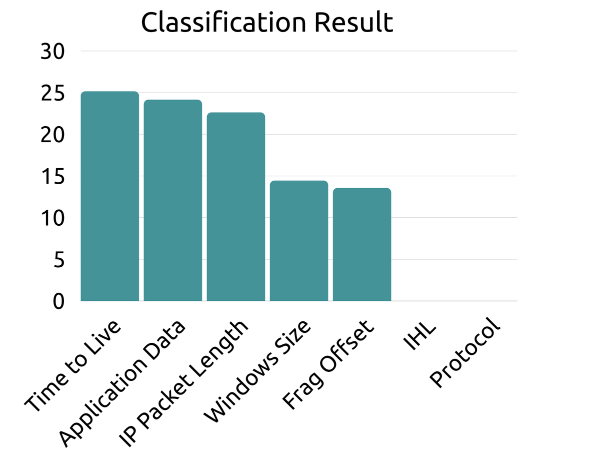


DELETING THE FINGERPRINT

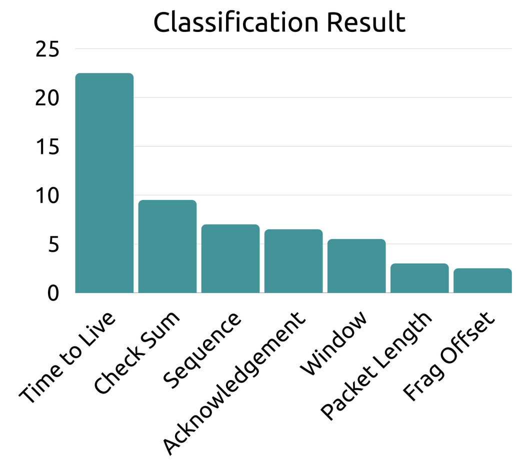
- To derive IoT device fingerprints, we deployed a Squid proxy server on Raspberry Pi 4 and an Intel PC.
- Another device recorded network traffic while accessing a list of 1000 popular websites.
- Captured packets were dissected and analyzed using a Random Forest model to identify distinguishing features from PC or IoT proxies.



Percent Impact on Stream Classification Result



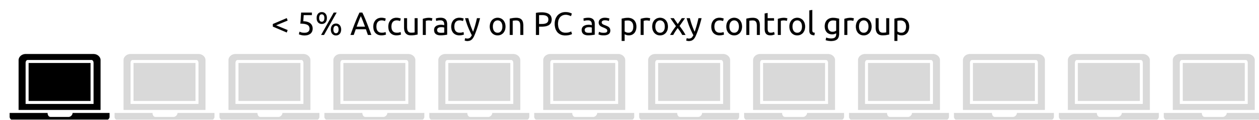
Percent Impact on Packet Classification Result



Percision	71%
Recall	64%
Accuracy	76%

Using collected data from an IoT proxy and PC proxy, we trained a Random Forest ML model in order to classify each packet and/or stream into corresponding classes. Above is the percentage weight that each feature had on the classification algorithm.

< 5% Accuracy on PC as proxy control group



PROBLEM LANDSCAPE EXPLORATION

How do Censors Block Traffic?[4]

Blocking specific IP addresses and or static fingerprints
Blocking or interfering specific protocol include DNS.
Blocking keywords in URLs and "Deep packet inspection"
Probabilistic and statistical traffic classification
Active probing and discovery of circumvention usage

Colleterial Damage

- Collateral damage is the cost incurred by the censor when it accidentally blocks something it would have preferred to allow.
- Censors must balance their desire to block certain content with the need to avoid harm to themselves, leading to the acceptance of some level of circumvention traffic.

Active Probing

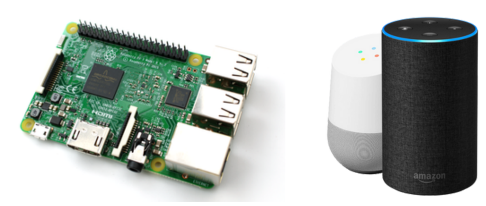
- Censors use active probing to detect proxy servers by making connections to suspected addresses.
- The goal of active probing is to increase precision in identifying proxies while minimizing false positives.
- Active probing allows censors to asynchronously run the detection process separate from other firewall responsibilities.

Content Obfuscation

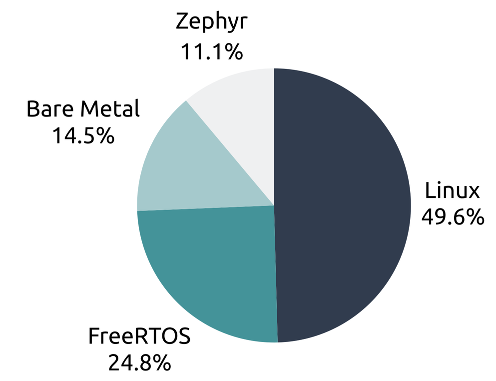
- The first strategy, steganography, involves mimicking content that the censor allows, such as HTTP or email, to evade detection.
- The second strategy, polymorphism, focuses on randomizing content to make it dissimilar to anything specifically blocked by the censor.

What are Consided IoT Devices?[2]

- Self-automated and pipelined devices which are connected to the internet run for a prolonged time without human inspections.
  - Smart Trashcan, Traffic Light, Remote Camera, Sensors, etc.



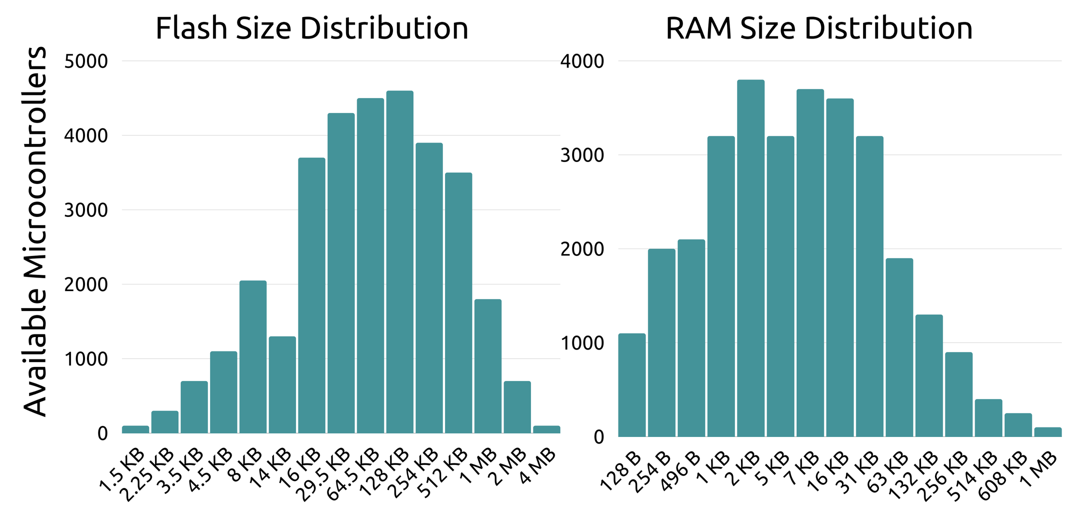
IoT Operating Systems (OS)[3]



- The most prevalent OS determined by a developer survey are as follows: Linux, FreeRTOS, Zephyr.
- Knowing the prevalence of Linux, targeting the UNIX based systems provides a solution for the deployment of a potential solution.
- Furthermore, the devices that the UNIX shells are running have higher computational power able to support tradition encryption and proxy software.

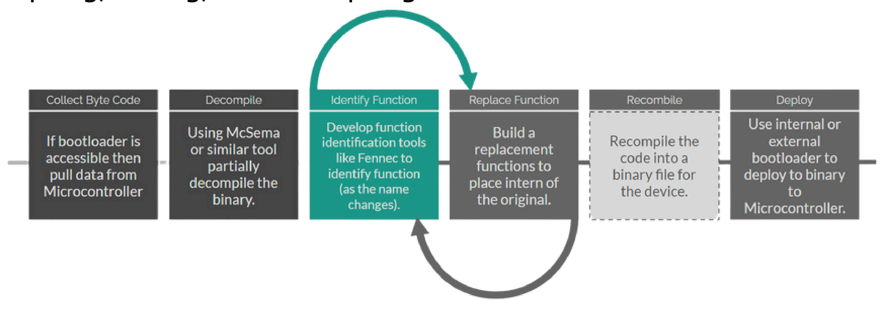
Resource Constraint IoT

Due to the size and diversity of the IoT part of the preliminary landscape was to decide the feasibility of targeting resource constraint devices. The data below is the distribution of devices from DigiKey a prominent IC supplier.



Using the data collected and testing using Cooja it was determined that resource constraint IoT devices in the classes 1,2, and 3 defined by RFC-7288 are not able to be targeted effectively for the following two reasons...

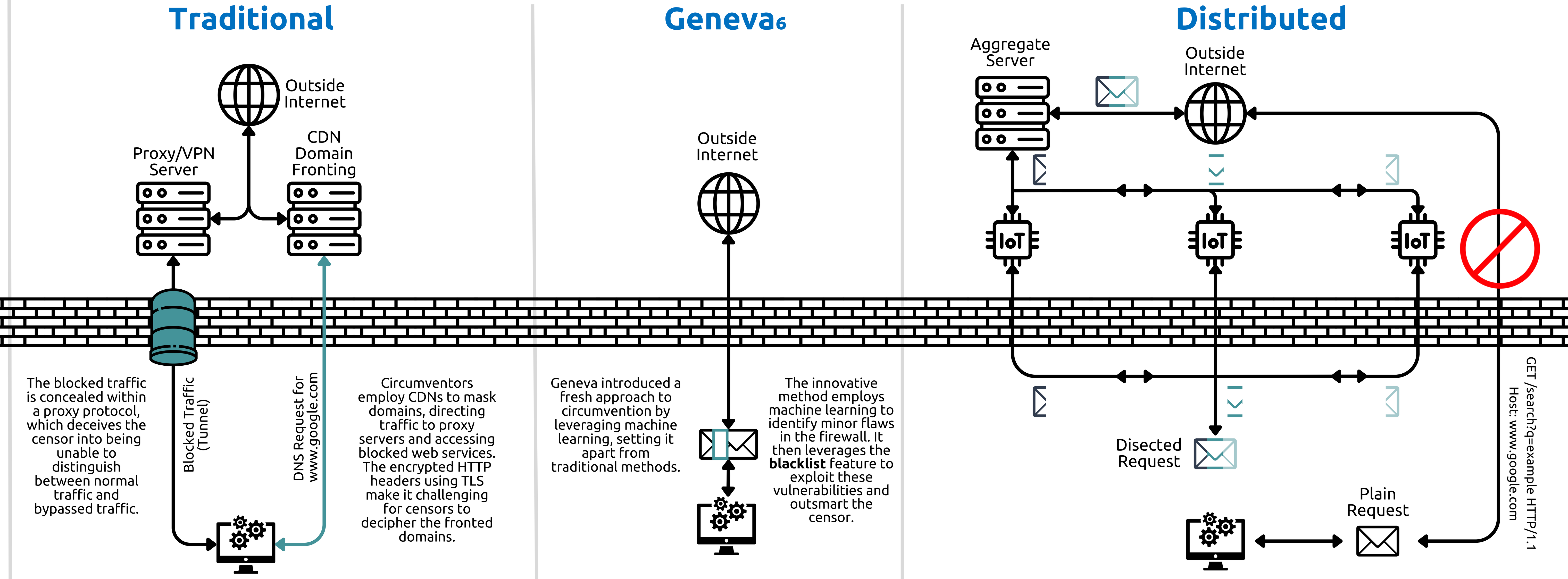
1. Resource constraint IoT devices struggle to run full TLS and or DTLS libraries let alone routing and other supporting structure.
2. Installation on resource constraint devices with Microcontroller involves decompiling, editing, and recompiling raw binaries from the device flash.



DESIGN

NUETRAL / CENSORED

Traditional | Geneva[6] | Distributed



Traditional

Advantages:

- Well Researched: thorough research ensures that the project is based on solid foundations and reliable data, increasing its credibility.
- Traffic Obfuscation: data is kept private, safeguarding users from potential physical threats.
- Existing Infrastructure (Snowflake/Tor/etc Volunteers): a robust network of resources and expertise for sustainable success.

Drawbacks:

- Static fingerprints: traffic source is identifiable.
- Static solution: perpetual "Cat & Mouse" game for developers.
- Typically Reliant on CDNs (Cloud Delivery Network).

Geneva

Advantages:

- Dynamic solution: creates an end to the "Cat & Mouse" game between researchers and developers.
- Typically faster: no middle-man proxy servers to forward requests.
- Dynamic fingerprints: traffic source is still however identifiable.

Drawbacks:

- No Traffic Obfuscation: data is not kept private.
- Reliance on "holes" in the Firewall.
- IP blocking: censor can black-list client IP that is using Geneva.

Distributed

Advantages:

- Static solution: partially defeats "Cat & Mouse" game.
- Usage of existing infrastructure: IoT devices are already present and have spare computing resources.
- Traffic Obfuscation: data is kept reasonably private.

Drawbacks:

- Typically slower: many proxy servers.
- Reliance on volunteer infrastructure.
- Static fingerprints: traffic is identifiable.

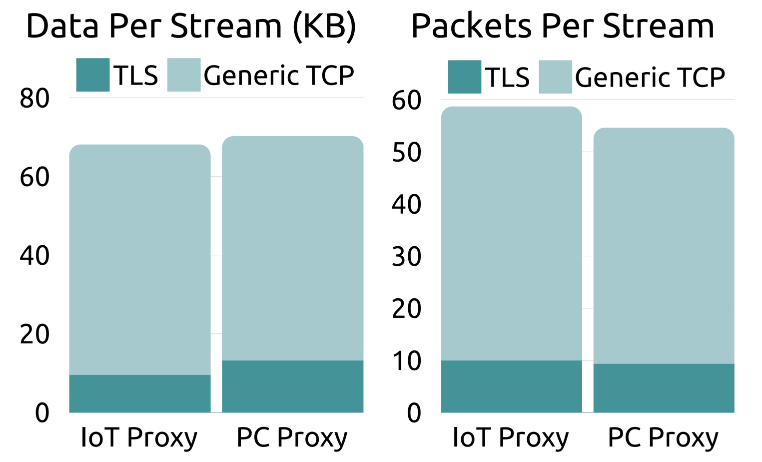
The Concept:

The fundamental principle of a distributed system lies in harnessing the autonomy of its components to enable seamless communication. Operating across international boundaries, volunteers in this system benefit from unimpeded communication, allowing messages originating domestically to reach them without hindrance. Upon reaching the volunteers, these communications remain immune to censorship, thus preserving the unobstructed dissemination of information.

THE RATIONAL

69 KB

On average, there is 69 KB of data to store for each basic web request. To reassemble the stream, a censor would need to cache all the data. However, on a large scale, across a country with ongoing stream traffic, this cost becomes unmanageable.



n!

On average, there are 55 packets for every basic web request, resulting in 55! combinations for a sensor to reconstruct the stream. Although the size decreases considerably with heuristics, it remains computationally impractical.

This model could potentially lead to notable ethical dilemmas, as its implementation may pose challenges for the property owner.

Lower risk to participants contributing as IoT proxy servers, as the deployment of IoT devices is spatially (locations) and financially more convenient than PCs.

This approach capitalizes on the complexity of computing the Wagner-Fischer Algorithm or a similar method to reconstruct captured packets.[1]

The time factor necessitates storing packets in a buffer for a certain period for reassembly. This process is challenging on a large scale due to computational complexities.

Blacklisting serves as a technique employed by censors for control. Our system circumvents this by using unclassifiable traffic, enabling its passage as the default setting.

The circumvention model has a lower profile in the eyes of censors, as it doesn't rely on TLS, which is the primary target for censors like the GFW.[5]

When utilizing multiple interconnected devices through the Internet, processing such information centrally will be considerably faster than when done decentralized.

SUMMARY

WHAT WE BELIEVE

- We believe that our model could be a promising solution to the censorship problem at hand because it exhibits **lower risk, lower profile, computational complexity,** and **unclassifiable traffic.**

WHAT WE ACKNOWLEDGE

- In the same breath, we also recognize weaknesses in our approach including **ethical concerns** and **slower connectivity.**
- Furthermore, we recognize issues such as an IoT device having a **static fingerprint,** and **low cost** to a censor.

WHAT WE STILL NEED TO DO

- To transform this model into a comprehensive solution, there are several essential tasks that must be completed.
- Firstly, **active probes** require attention and neutralization.
- Secondly, resolving challenges related to sending and receiving **DNS requests** is crucial.
- Thirdly, gathering and analyzing **real-life data** from the implementation of this model is necessary.

REFERENCES

1. Backurs, A., & Indyk, P. (2015, June). Edit distance cannot be computed in strongly subquadratic time (unless SETH is false). In Proceedings of the forty-seventh annual ACM symposium on Theory of computing (pp. 51-58).
2. Garcia-Morchon, O., et al. "Internet of Things (IOT) Security: State ofthe Art and Challenges." RFC Editor, Apr. 2019, www.rfc-editor.org/rfc/rfc8576#section-1.
3. 2020 IOT Developer Survey - Eclipse IOT, iot.eclipse.org/community/resources/iot-surveys/assets/iot-developer-survey-2020.pdf. Accessed 17 Apr. 2024.
4. Threat Modeling and Circumvention of Internet Censorship, www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-225.pdf. Accessed 18 Apr. 2024.
5. Fingerprinting Obfuscated Proxy Traffic with Encapsulated ..., www.usenix.org/system/files/sec24summer-prepub-465-xue.pdf. Accessed 18 Apr. 2024.
6. "Geneva: Evolving Censorship Evasion Strategies." Censorship.Ai, geneva.cs.umd.edu/papers/. Accessed 18 Apr. 2024.
7. K. Kostas, M. Just and M. A. Lones, "IoTDevID: A Behavior-Based Device Identification Method for the IoT." Dec.1, 2022. Accessed 20 Apr. 2024.