Cache to the Future: Webpage Access in a Blackout

Motivation

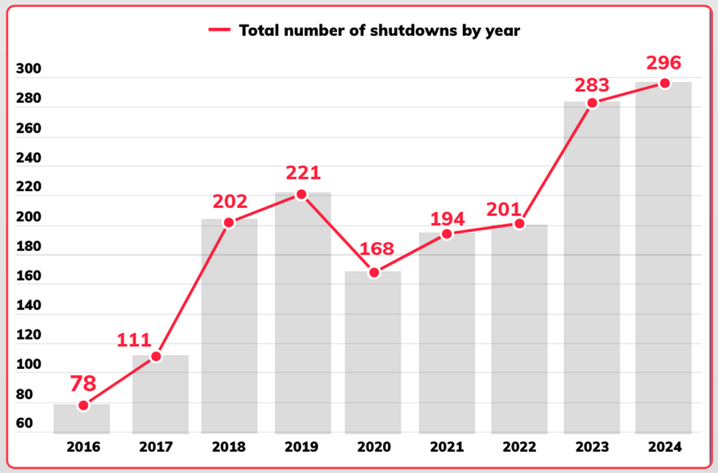
Imagine one day, you and everyone around you find yourself completely **cut off from the internet**.

Internet blackouts prevent citizens within a region from accessing the internet.

The number of blackouts is increasing over time and poses a **growing problem**.

Blackouts have become a staple of **modern war strategy**, most recently utilized by Russia in Ukraine, by Israel in Gaza, and by China in Myanmar.

Source: *Access Now*



Citizens faced with **Internet Blackouts** often resort to **blackout-resistant technologies** to re-establish communications. Most rely on Bluetooth and mobile-mesh networking. But these technologies:

- Only support *messaging* and *microblogging*
- Have severely unrealistic evaluations
- Neglect physical interference in densely-populated areas

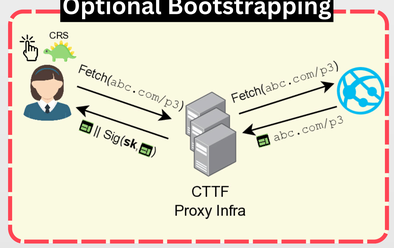
Cache to the Future

- Provides access to **authoritative, cached web content** via a **distributed internet archive**
- Is resilient against a wide range of attacks
- Presents the first evaluation based on real-world city-wide smartphone GPS data

Pre-Blackout Phase

1)

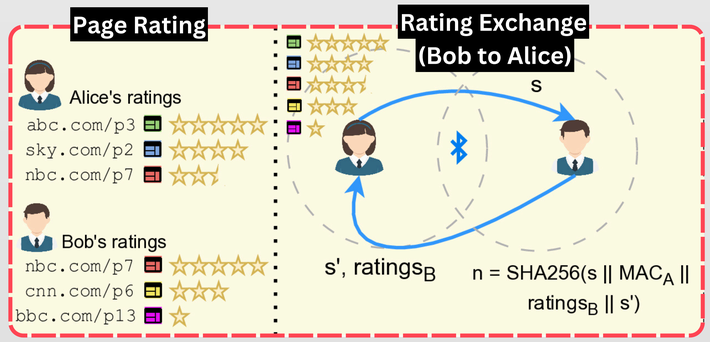
Manual Page Caching & Optional Bootstrapping



Users cache pages through a trusted proxy which provides a **digital signature.**

This signature prevents adversaries from tampering with page contents during a blackout.

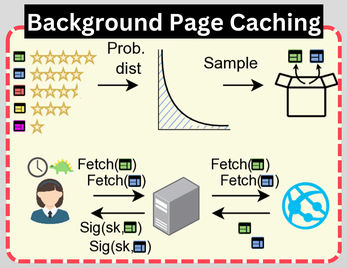
2)



- Users rate pages based on their utility during a blackout, which affects **cache replication.**
- Ratings are **automatically shared** among users when they come in close contact via Bluetooth and WiFi Direct.
- A **Proof-of-Work scheme** ensures that adversaries cannot excessively manipulate local averages.

3)

Background Page Caching

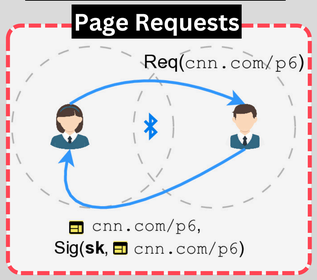


Cache replication occurs automatically by transforming ratings into a Zipf prob. dist., sampling, and caching according to leftover space

4)

Blackout Phase

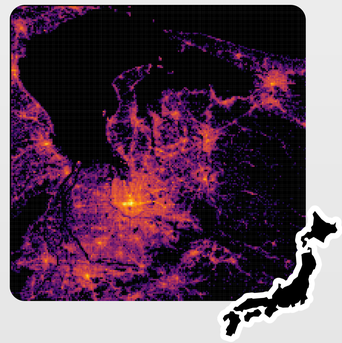
Page Requests



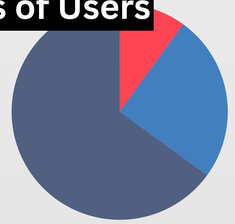
Users can request and receive pages via Bluetooth; authenticity is validated by checking the proxy's signature.

Evaluation

Evaluation was conducted on the YJMob100K dataset: real-world GPS data from **25,000 people** in a Japanese city.

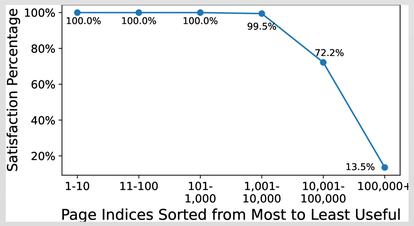


Types of Users

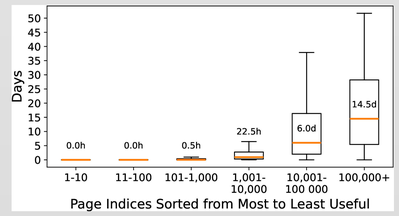


Proactive Users: Actively cache and rate pages.
Leech Users: Just cache pages in the background.
Adversaries: Manipulate ratings, perform signal jamming, respond with misinformation.

Benign Scenario

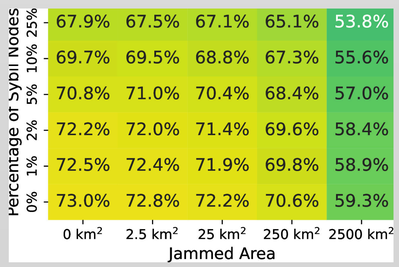


Users can fetch **top 100K** pages with **high probability.**

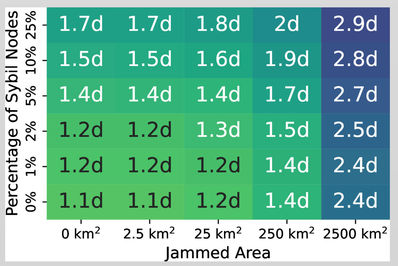


On avg., users can fetch top 10K pages within **a day,** 10K-100K within **a week.**

Adversarial Scenario



Users receive **70%** of the pages they request even under aggressive jamming and adversary participation.



Users can receive **90%** of requested pages within **1-2 days** even under aggressive jamming and adversary participation.

UNIVERSITY OF WATERLOO | DAVID R. CHERITON SCHOOL OF COMPUTER SCIENCE

Ross Evans
rpevans@uwaterloo.ca

Diogo Barradas
diogo.barradas@uwaterloo.ca