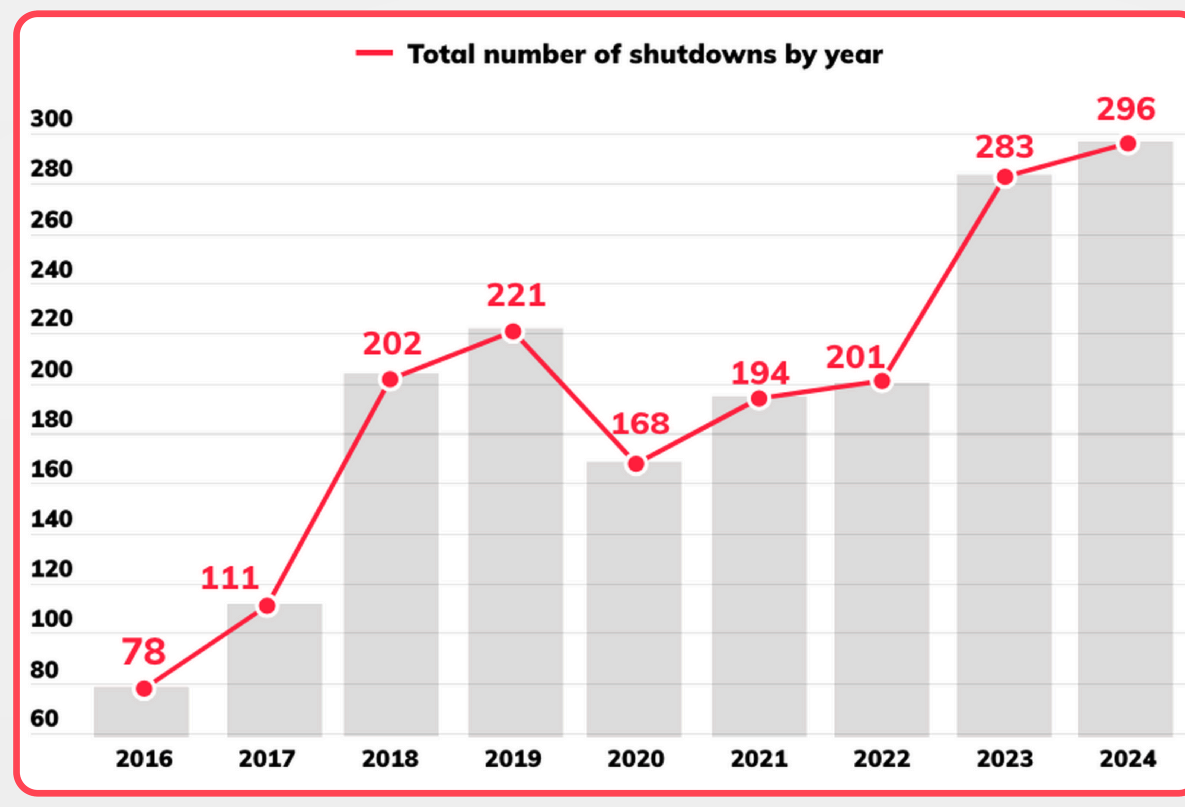
Webpage Access in a **Blackout**

Ross Evans
rpevans@uwaterloo.ca

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 all internet services.



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*

Under such extreme circumstances, what internet services can be replicated?

Cache in Hand

Prior works have primarily focused on **messaging** and **microblogging** scenarios, where nearby smartphones connect over Bluetooth and transmit messages to one another.

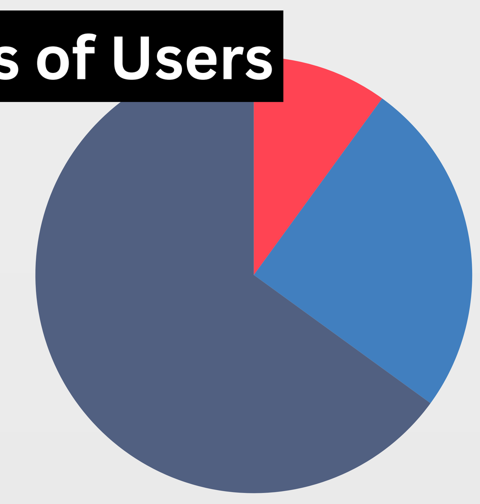
We address a different use case: caching webpages for later use during a blackout in a distributed way – a **distributed internet archive**.

Through our system, users can access **authoritative information** such as election news, corruption reports, or even encyclopedic sources that they would **otherwise be unable to access**.

Usefulness of websites will be determined by **community ratings**. **Pre-blackout,** users rate webpages according to usefulness, which will determine which pages are cached.

During a blackout, users can request pages and receive them from nearby devices over Bluetooth.

Types of Users



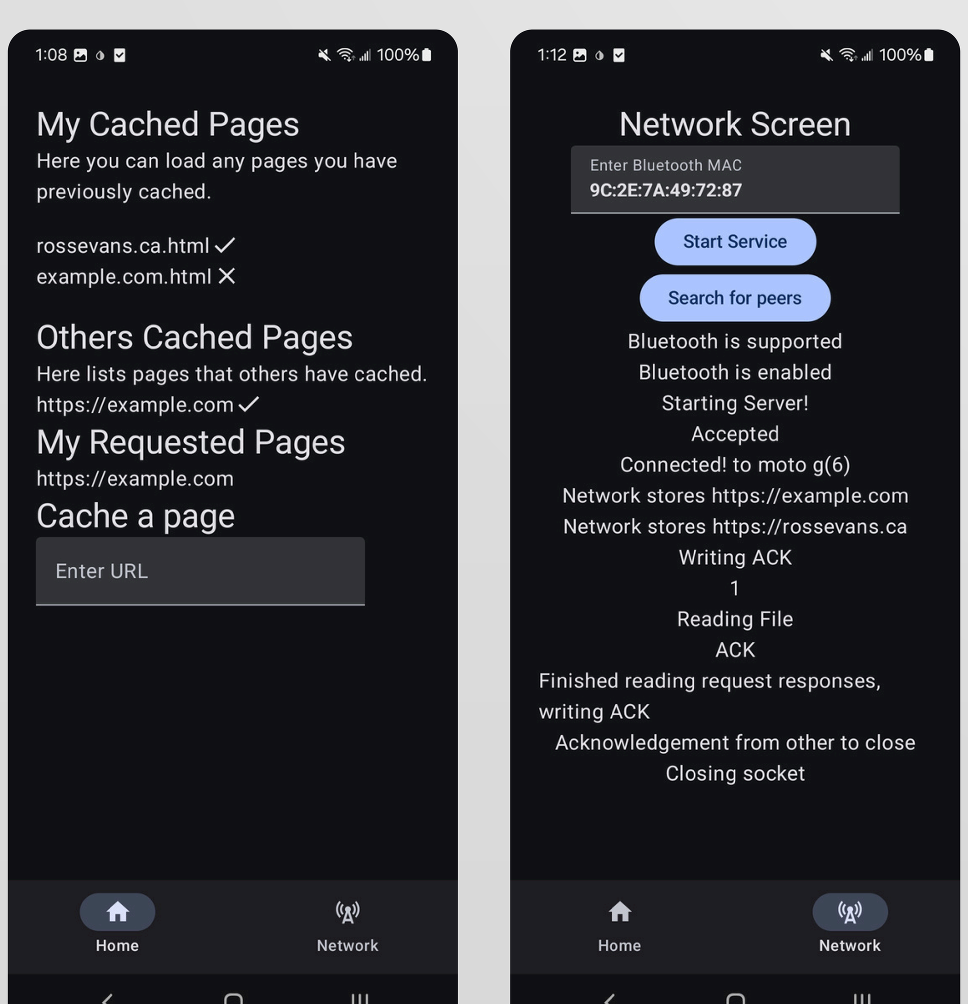
- **Proactive Users:** Actively cache and rate pages according to usefulness.
- **Leech Users:** Cache pages in the background, but do not actively rate pages.
- **Adversaries:** Act to disrupt the system wherever possible.

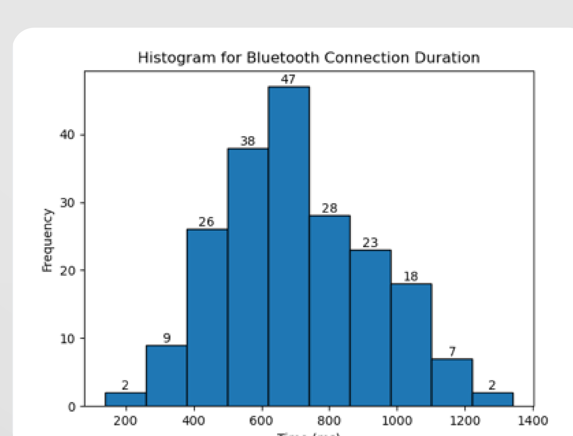
To disrupt the system, **adversaries** may **attempt to maliciously alter community ratings**, so that **leeches** end up caching pages which then end up infrequently accessed.

Adversaries may employ **jamming** in highly populated areas during blackouts, so that Bluetooth cannot be used to transfer pages between individuals.

Adversaries may also respond to requests from **legitimate users** with pages that have been **altered to contain misinformation.**

Initial Prototype





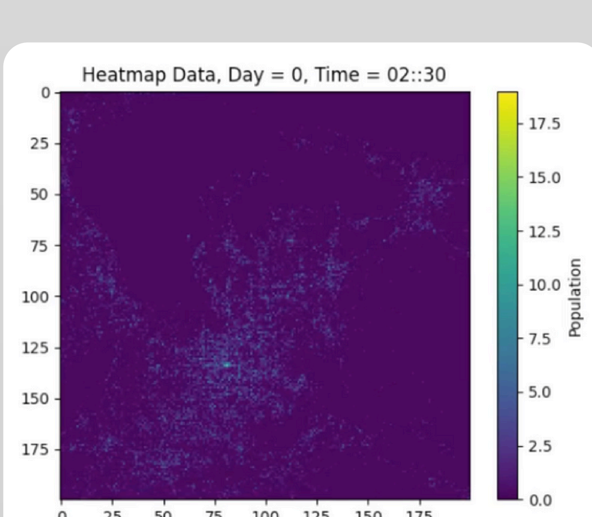
We have an **initial prototype** developed as an Android app.

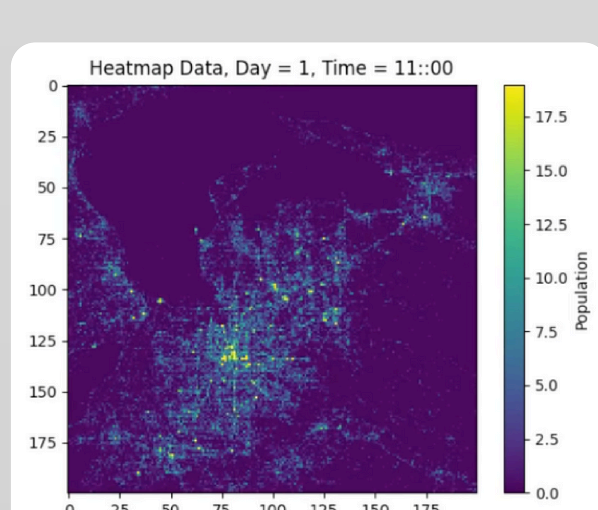
Users can transfer pages via **Bluetooth** and **Wifi-Direct**, ratings will be implemented soon.

We currently mitigate adversaries serving pages with misinformation during a blackout via a checksum stored pre-blackout.

We also intend to develop more complex checksum procedures utilizing perceptual hashing schemes for the cases where a page may have received minor updates since a checksum calculation was performed.

Simulations

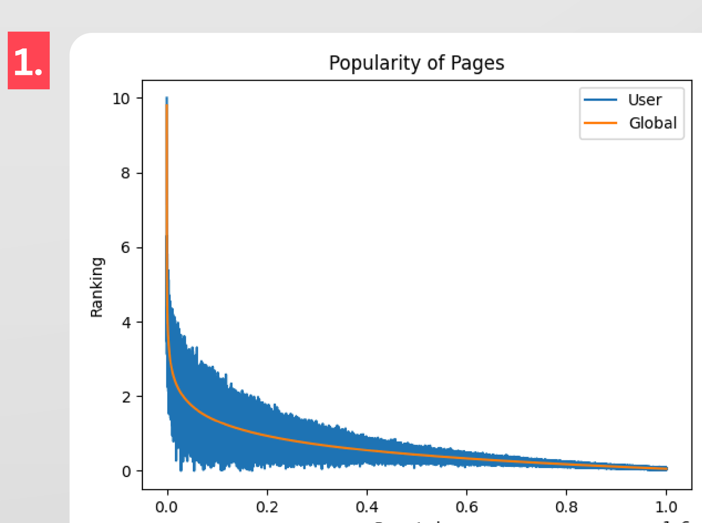


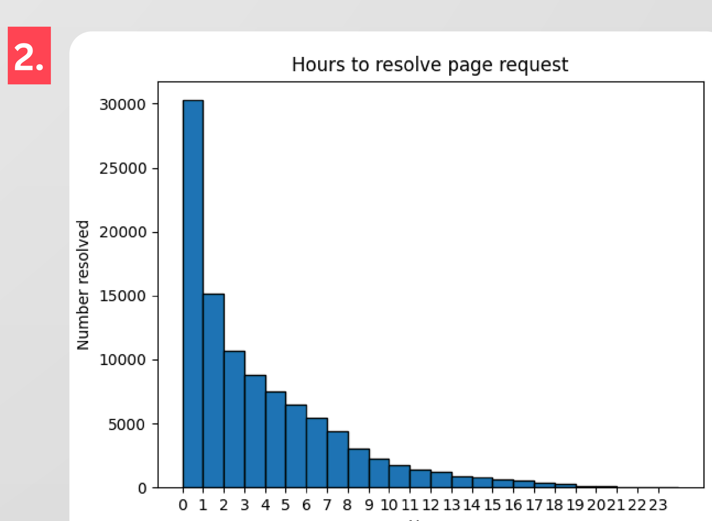


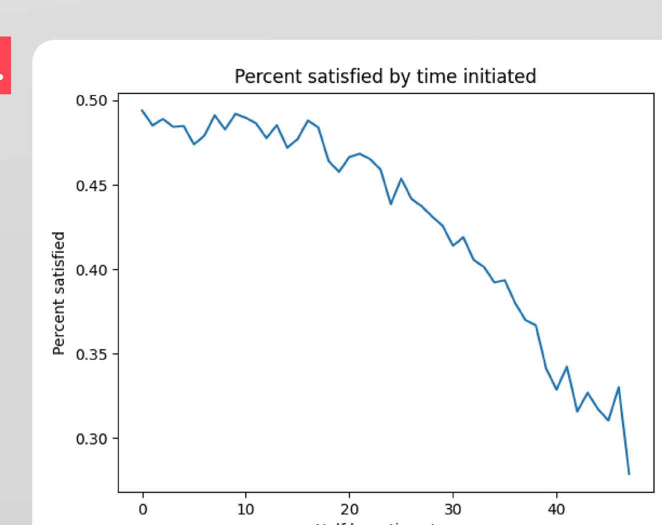
Simulations run on an idealized grid model and the YJMob100K dataset

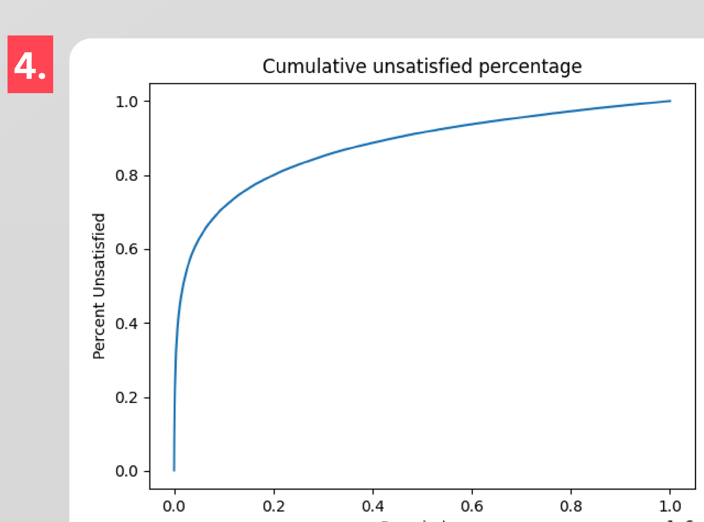
This dataset provides location data from 100,000 people living in a Japanese city over 75 days: 15 days of movement correspond to an "emergency event"

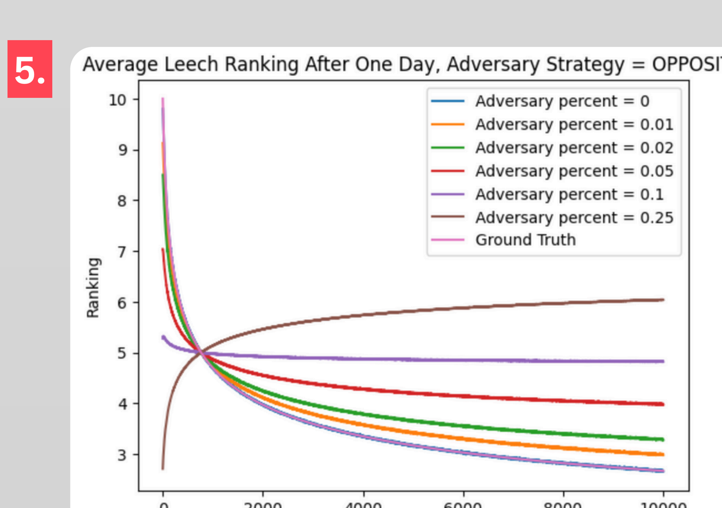
Preliminary Results

1.


2.


3.


4.


5.


1. We model the popularity of pages with a **Zipf** distribution. This distribution is modelled from a paper out of **Stanford** and **Google** which provides data comparing website popularity with traffic concentration. We apply a **scaling transformation** to map popularity to a rating from 0-10. Each user follows the same **general distribution** with some normal additive decaying noise.

2. After simulating ranking dispersal and leech caching, we simulated a **1-day page access scenario**. In the **vast majority** of cases, webpages were accessed from others in **fewer than 8 hours**.

3. As the simulation proceeds, it becomes less likely for users to get a response to their page request in the limited amount of time remaining. Nevertheless, **even in the last 4 hours** of the simulation, where user mobility is relatively low, **we can still satisfy 1/3 of all new requests.**

4. Over 50% of page requests which do not receive a response from other users belong to the **long tail** of the distribution, corresponding to pages with an **average rating of less than 2**.

5. We modelled how page ratings are affected when adversaries are introduced that attempt to **negatively influence rankings**. As the percentage of adversaries approaches the percentage of proactive users, **leeches struggle to effectively differentiate useful pages**. Nearly all pages get rated a 5, eventually flipping such that useful pages are rated very low scores.