Xiao-hong ZHANG, Kai QIAN, Jian-ji REN, Zong-pu JIA, Tian-peng JIANG, Quan ZHANG, 2019. Measurement and analysis of content diffusion characteristics in opportunity environments with Spark. *Front Inform Technol Electron Eng*, 20(10):1404-1414. <u>https://doi.org/10.1631/FITEE.1900137</u>

Measurement and analysis of content diffusion characteristics in opportunity environments with Spark

Key words: Content dissemination; Device-to-device communication; Opportunity network; Linear threshold model

Corresponding author: Jian-ji REN E-mail: renjianji@hpu.edu.cn ORCID: http://orcid.org/0000-0002-3834-340X

Motivation

- Opportunity networks provide a chance to offload the tremendous cellular traffic generated by sharing popular content on mobile networks.
- Analyzing the content spread characteristics in real opportunity environments can discover important clues for traffic offloading decision making.
- However, relevant published work is very limited since it is not easy to collect data from real environments.

Main idea

- Content transmission characteristics are analyzed with the dataset collected from a real opportunity environment formed by the users of Xender, which is one of the leading mobile applications for content sharing.
- To discover content transmission characteristics, scale, speed, and type analyses are implemented on the dataset.
- A solution is proposed to maximize file spread scales, which is very helpful for forecasting popular files.

Method

Scale analysis

The spread scale of each file is determined by the number of edges in the spread graph of that file.

Speed analysis

File spread speeds are calculated according to the edges generated at different times.

Type analysis

For each type of file, the proportions of files are analyzed at different spread scales.

- Maximization file spread scales
 - Select seeds according to the degree factor, seed factor, and PageRank algorithm.
 - Exclude nodes that will not be activated in content propagation to improve efficiency.

Scale analysis

Only a very small fraction of files spread widely.



Fig. 2 Distribution of $F_{\text{scale}}(s)$

Speed analysis

All files share a similar spread trend regardless of spread scales and spread with a strong periodicity.



Fig. 4 File spread speeds: (a) small files; (b) small-medium files; (c) medium files; (d) medium-large files; (e) large files; (f) huge files

- Type analysis
 - Image, video, and audio files are shared most.
 - App files have a much higher probability to be popular than other files.



Fig. 7 Ratio of file types

Our solution to maximization file spread scales



Fig. 8 Comparison on the Xender dataset: (a) scale with different numbers of seeds; (b) execution time

Our solution to maximization file spread scales



Fig. 9 Comparison of execution time: (a) EU institute dataset; (b) Soc-sign-epinion dataset



Fig. 10 Comparison of spread influence: (a) EU institute dataset; (b) Soc-sign-epinion dataset

Conclusions

- We analyze content spread characteristics by performing measurement and analysis on the dataset collected from Xender, which provides file-sharing services through user mobility and D2D communication.
- We take file spread procedures as graphs and implement scale, speed, and type analyses on the dataset.
- We also propose a solution to maximize file spread scales. This is very important for forecasting popular files and helping make traffic offloading strategies.