



Insight of File Sharing System Forum

Guillaume Jourjon, Olivier Mehani, Thierry Rakotoarivelo



Australian Government
Department of Broadband, Communications
and the Digital Economy
Australian Research Council

NICTA Funding and Supporting Members and Partners



- **What is 1-CFHS?**
 - One-Click File Hosting Systems: website which allows users to upload and download content referred to by single URL, without offering indexing services for that URL
 - Popular sites: Uploaded, Rapidgator, Filepost, Rapidshare
- **Why they matter?**
 - Increasingly being used for the exchange of copyrighted contents
 - Direct URLs to copyright infringing contents are exchanged between users
 - Use of dedicated File-Sharing Forums (FSF) to disseminate URLs via posts
- **Our Objectives in this Pilot Study**
 - Many existing studies have analysed 1-CFHS (e.g. traffic volume, content characteristics,...)
 - No study have so far focused on the dedicated forums used to disseminate URLs
 - Small scale study of some aspects of shared URLs shared and their referenced contents
 - Popularity of 1-CFHS
 - Duplication of URLs across posts and FSF
 - Birth-and-Death cycles and life expectancy of URLs and referenced contents

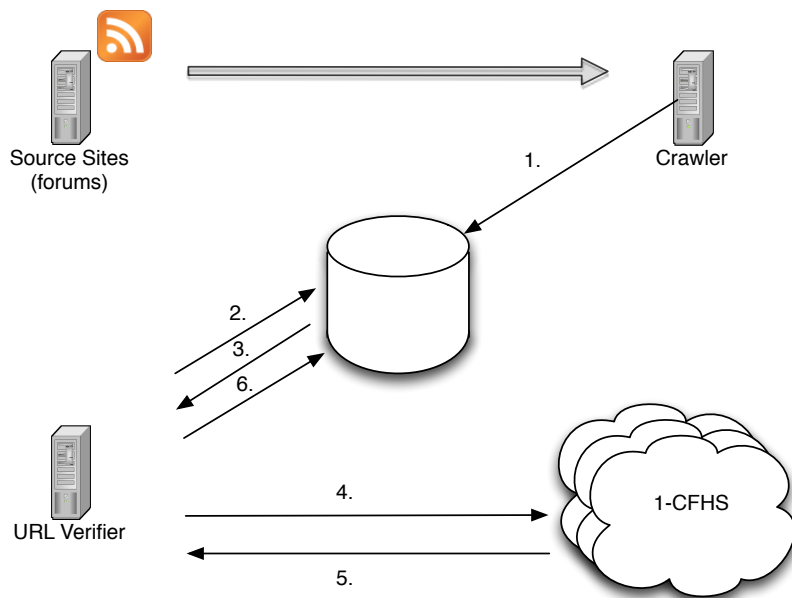
Experiment Design

- Deployment

- 4 FSF monitored for 3 months

TABLE I. SELECTED FORUMS

Name	URL	Alexa Rank
RLSLOG	http://www.rlslog.net	1,809
Avaxhome	http://avaxhome.ws	2,122
Sceper	http://sceper.ws	10,937
SceneSource	http://www.scnsrc.me	15,708



- Observed Dependent Variables

- Popularity of 1-CFHS
- URL Duplication
- Post size
- Content Lifecycle
 - **birth**: when the URL is first posted and its related content exists on the 1-CFHS
 - **death**: when the content has been removed from the 1-CFHS

- Independent Variables

- Forum ID
- Category ID
- User ID

Results

- Dataset size

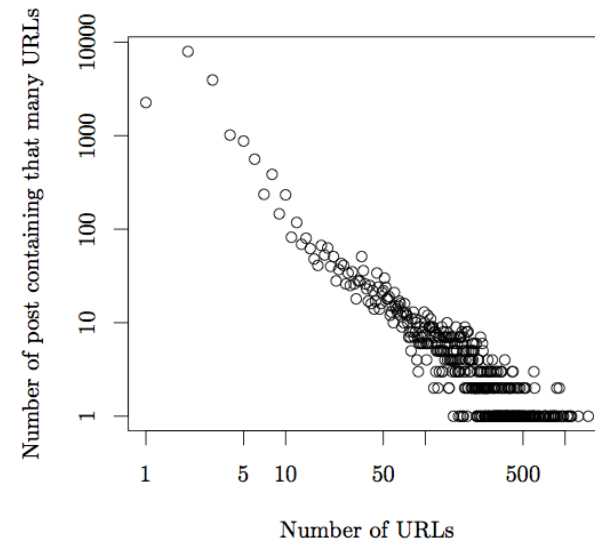
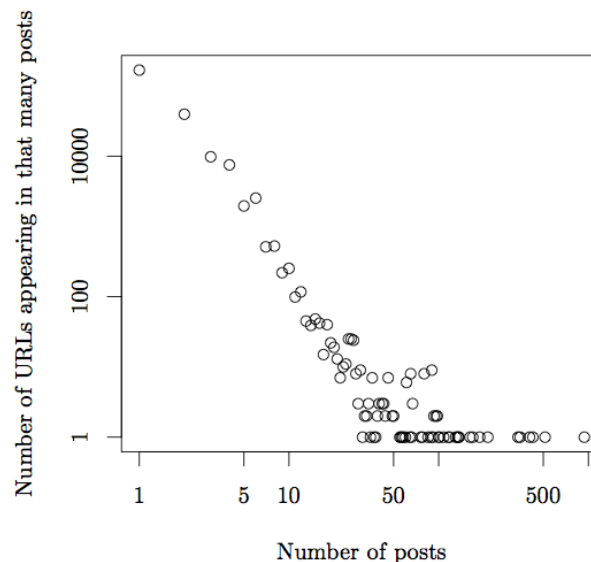
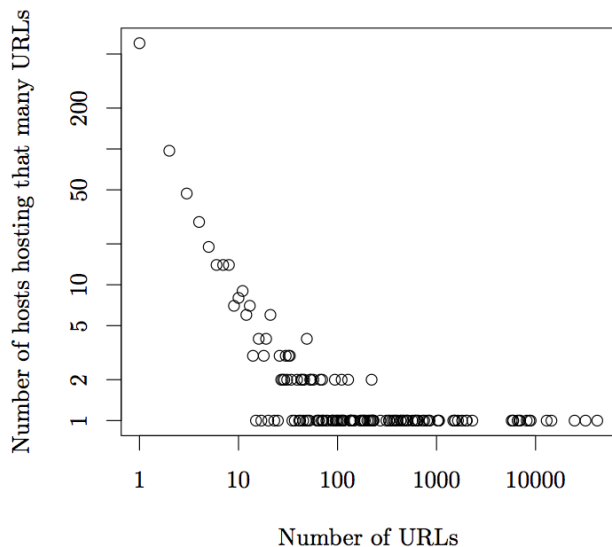
Name	Alive	Dead	No status	Total
Original Dataset	11,272	144,943	75,464	231,679
Filtered URLs	11,237	142,490	0	153,727

- 1-CFHS Popularity

- Top 10 accounts for 82% of all collected URLs
- Top 10 are different from previous CFHS studies

- URL Duplication and source posts

- 72% of URLs only present once across posts and forums
- 80% post contains less than 11 URLs

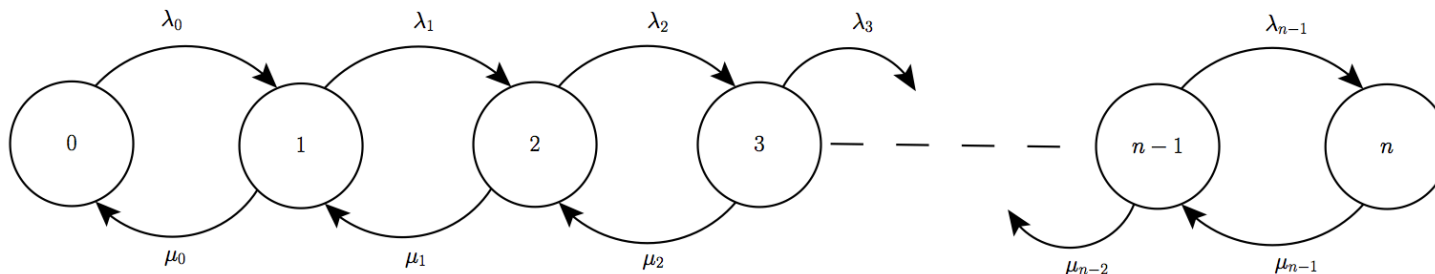
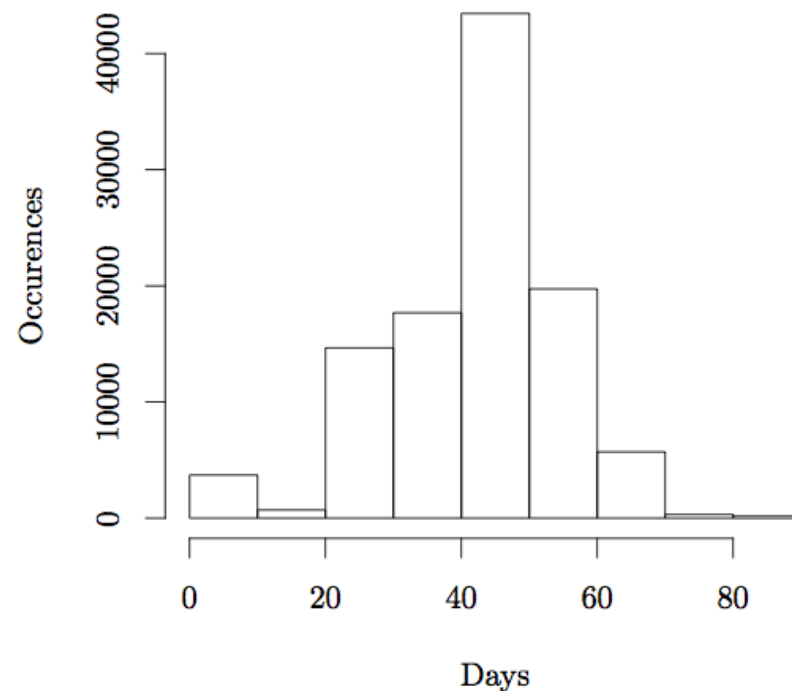


- Birth and Death rates

Variable	Mean	Standard Dev.
Inter-birth time	46.84 s	1,509.84 s
Inter-death time	5.68 s	347.87 s
Lifetime	41.85 d	12.74 d

- Life Expectancy

- On average tracked content are removed after about 40-42 days
- 90% removed at end of study
- Based on these measurements, we propose a birth-and-death model using Markov Chain



- URLs posted and/or duplicated on different forums?
 - Previous studies did not provide insights on URL disseminations, assumed duplication
 - Our results suggests:
 - URLs are rarely duplicated across posts and FSF: users are loyal to a FSF?
 - Some posts have >1000 URLs, some URLs are duplicated >100 times: spam posts/urls?
- 1-CFHS landscape
 - Top 10 services drastically changed in 1.5 years since previous studies
 - 1-CFHS market still in transient phase? No clearly defined “leader”
 - Only RapidShare maintained its presence in Top 10
 - Possibly due to change in ToS, which alter user loyalty and usage?
- Content Lifetime
 - Finite lifetime for content likely infringing copyrights, average about 40-42 days
 - Any impact of presented Independent Variables on this lifetime?

- **Pilot Study**

- Assess the interaction between 1-CFHS and their companion File Sharing Forums (FSF)
- Deliberately at a small scale
 - Validate experiment design and procedures
 - Explore some preliminary aspects of the FSF and 1-CFHS interaction

- **Preliminary Results**

- Confirm some findings from previous studies
- Low URL duplications, short URL numbers per post, presence of outliers in both cases
- Estimation of lifetime of content, 40 days average, most content removed after 3 months
- Initial design of Markov Chain model for birth-and-death of content on FSF/1-CFHS

- **Future Work**

- Full scale study underway with updated experiment procedures
- Additional data analysis: characterise spam posts, factor analysis for birth-and-death process, lag between content's legitimate publication and availability on FSF/1-CFHS