



Archiving Transactions Towards Uninterruptible Web Service

Zhiwu Xie, Prashant Chandrasekar, Edward A Fox

April 28, 2015 Stanford IIPC 2015 GA



Outline

- Motivation
- Technical background
- System architecture
- Demo
- Future work



Web Archiving Current State

- Stakeholders
 - Cultural heritage organizations
 - Academia
- Coverage
 - Wide but shallow



Web Archiving Current State

- Approach
 - Crawler-based archiving
- Value proposition
 - Comprehensive public good: preserve the history, collect research data, ...



Wider Participation?

- “Free rider”
- “What’s in it for me?”



Expand the Value Proposition

- Ask what the web archiving can do for the potential stakeholders and participants



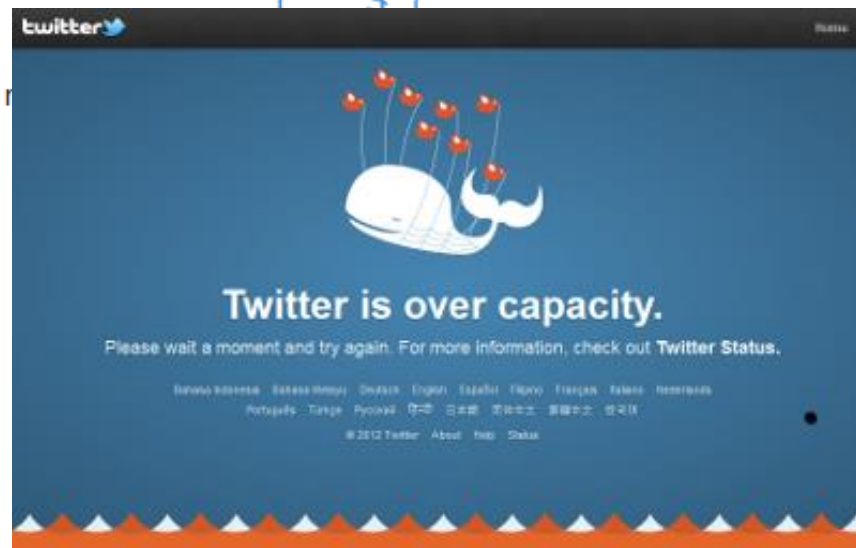
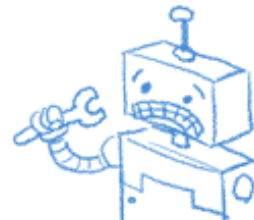
What Matters to a Website?



502. That's an error.

The server encountered a temporary error and could not complete your request.

Please try again in 30 seconds. That's all we know.



How Can Web Archiving Help?

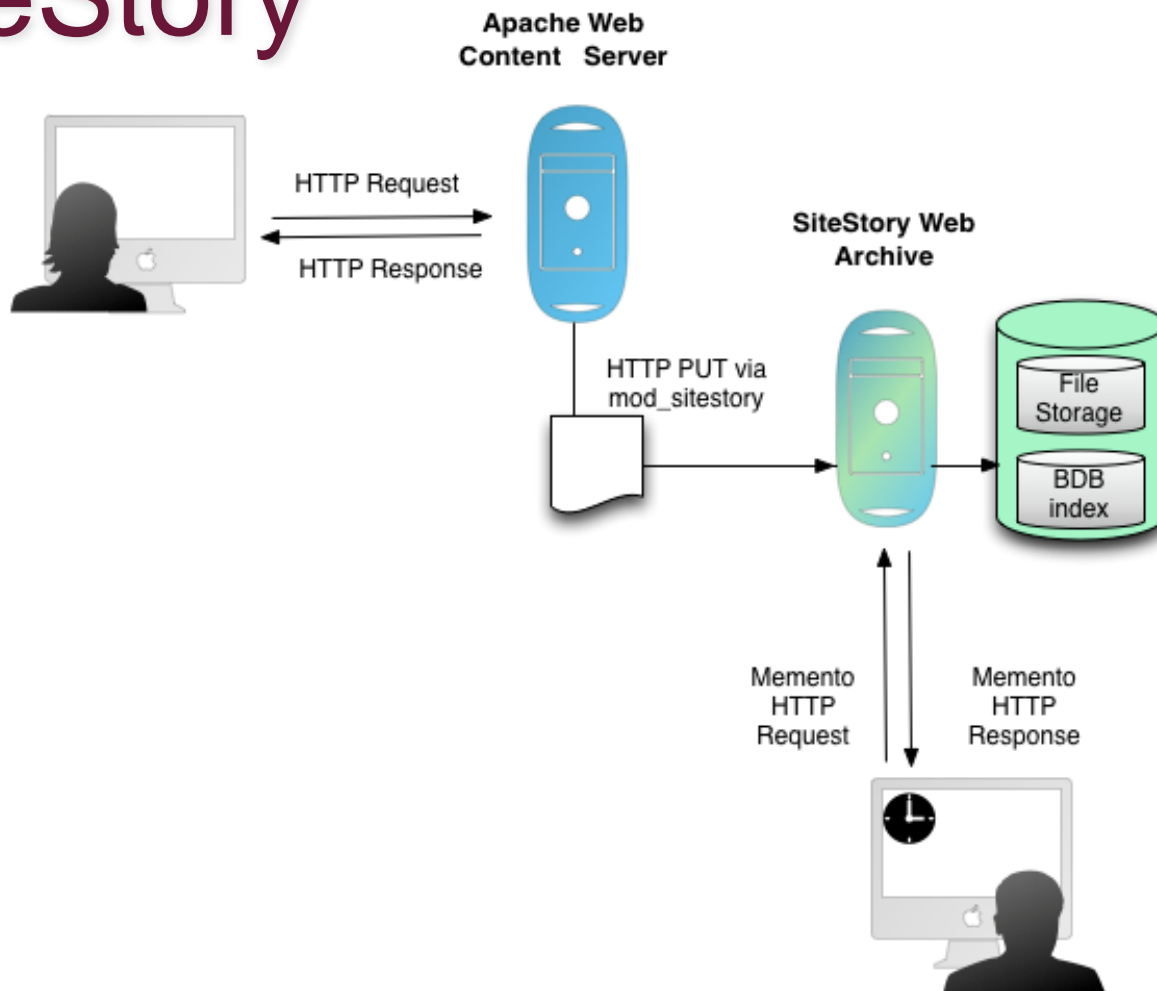




Web Archiving Approaches

- Crawler-base
- User-drive
- Transactional

SiteStory

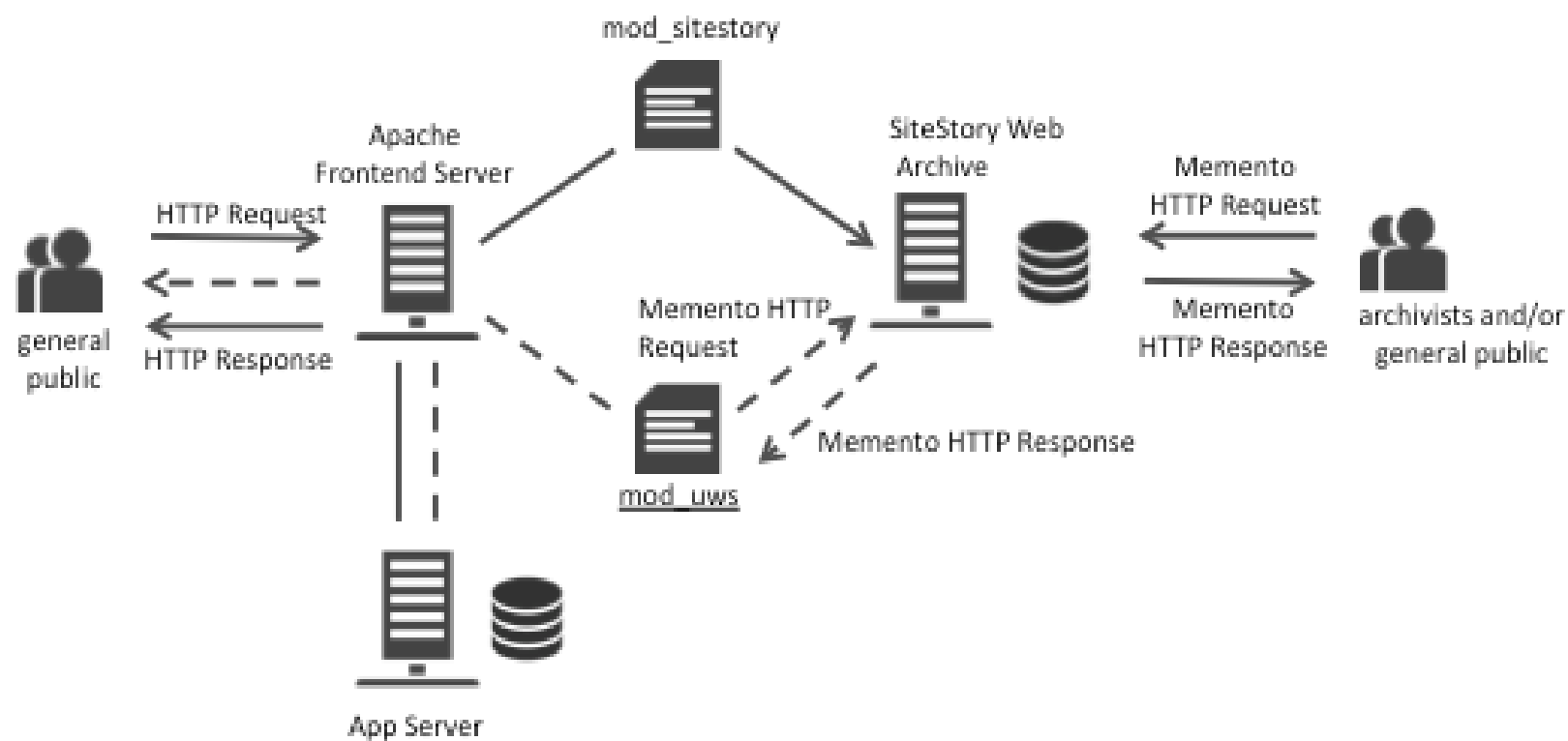




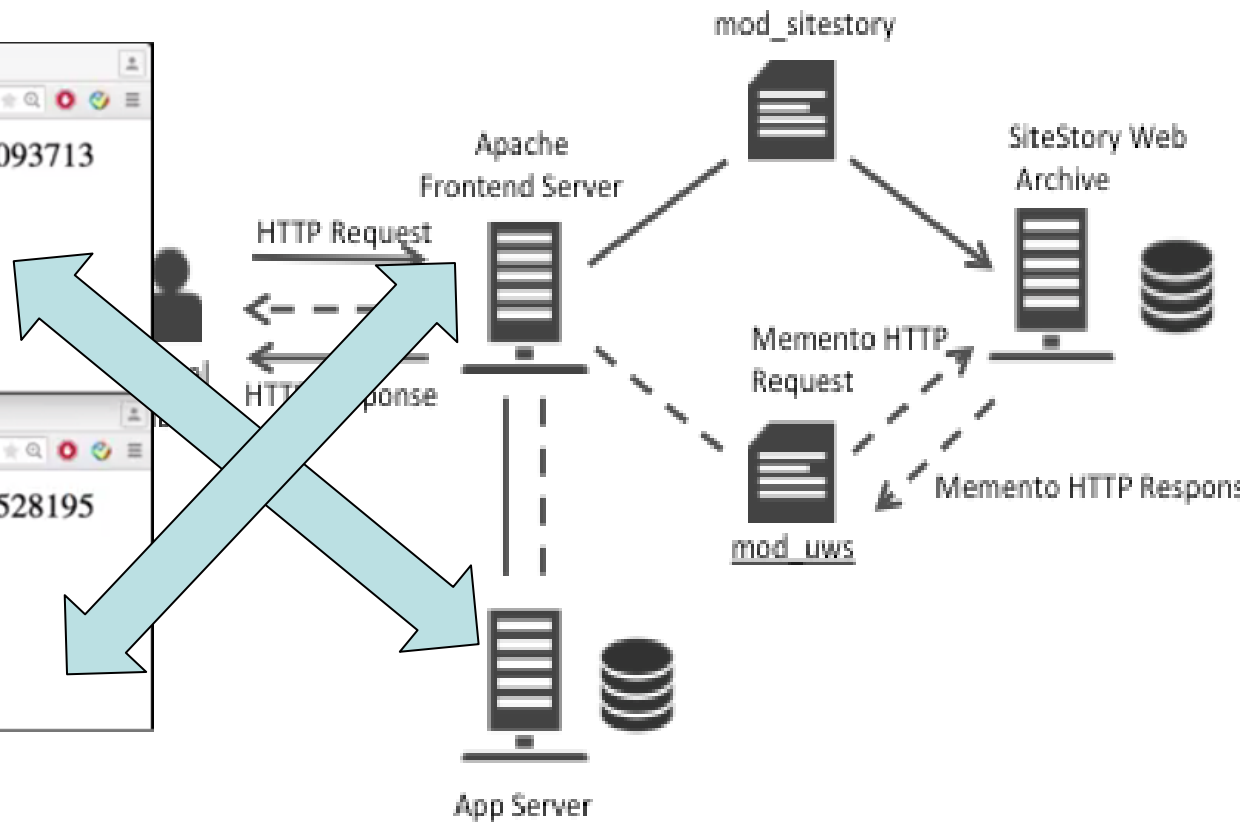
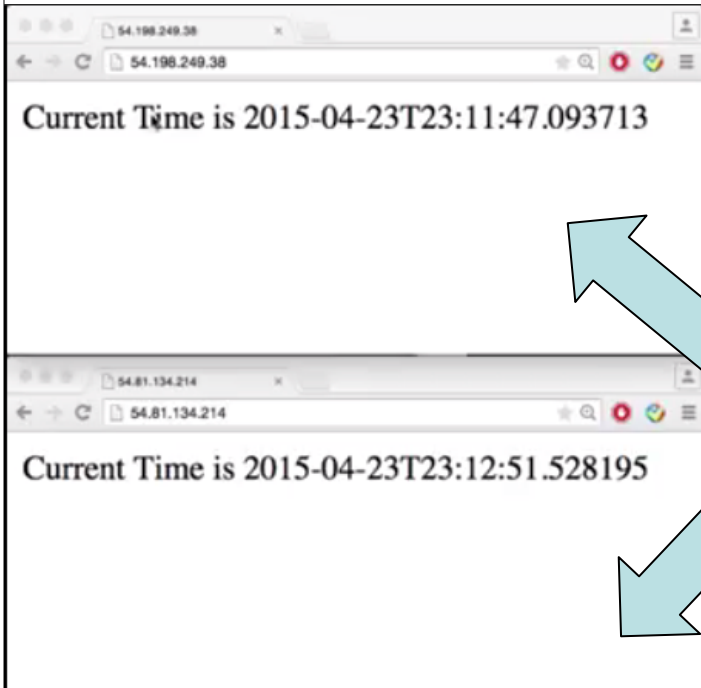
Uninterruptable Web Service (UWS)

- Turn SiteStory into a web infrastructure component that improves website availability
- Mask server errors from the end user
- Gain precious time for webmasters to recover from application server failures without disrupting the majority of their users' web experience

UWS Architecture



Demonstration





Future Work

- Moving UWS functions to application servers
- Improving performance



Acknowledgements

- Partially funded by the Web Archiving Incentive Award by Mellon Foundation via Columbia University



Questions?

- Zhiwu Xie zhiwuxie@vt.edu
- Prashant Chandrasekar peecee@vt.edu
- Edward A Fox fox@vt.edu