



Archiving Transactions Towards Uninterruptible Web Service

Zhiwu Xie, Prashant Chandrasekar, Edward A Fox

June 4, 2015 Columbia University



Outline

- Motivation
- Technical background
- System architecture
- Demo
- Performance tuning
- Relations to web cache and backup



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



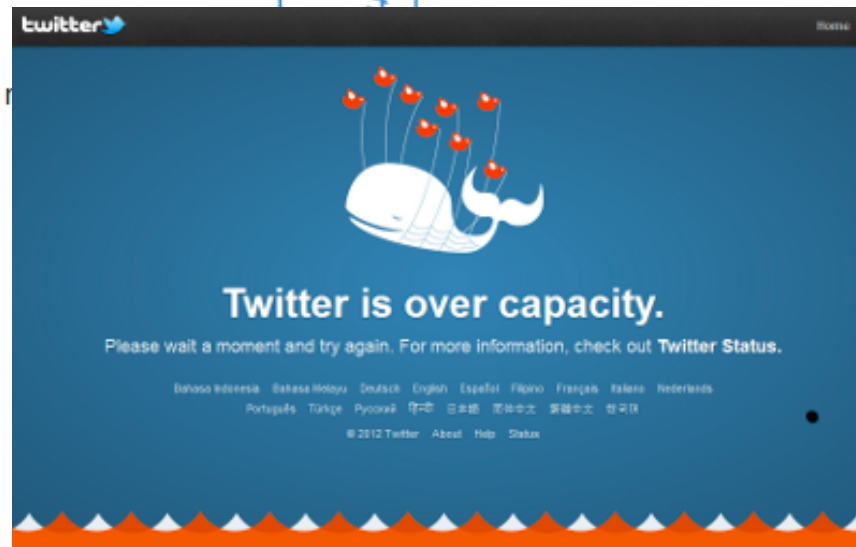
What Matters to a Website?



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?

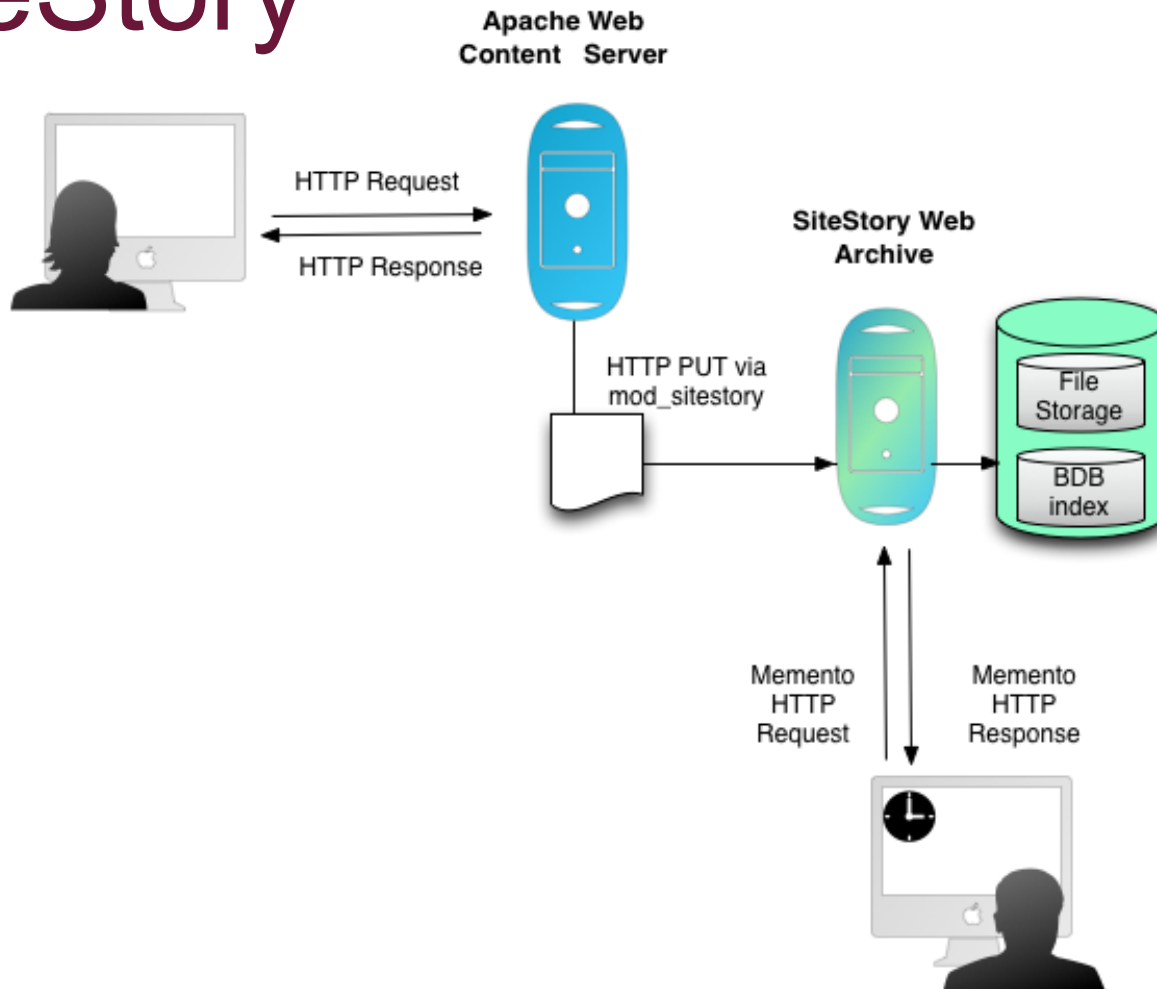




Web Archiving Approaches

- Crawler-base
- User-drive
- Transactional

SiteStory

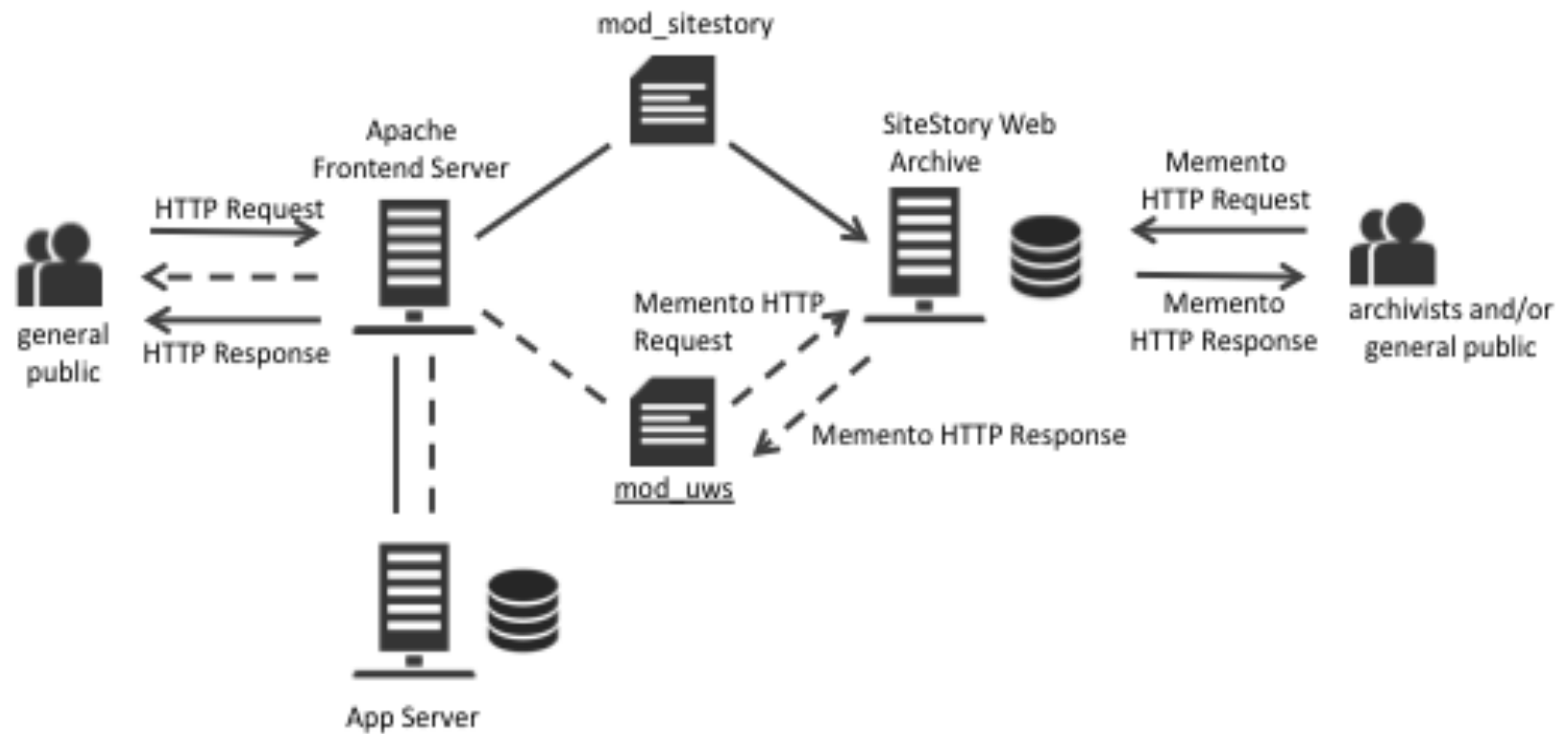




Uninterruptable Web Service (UWS)

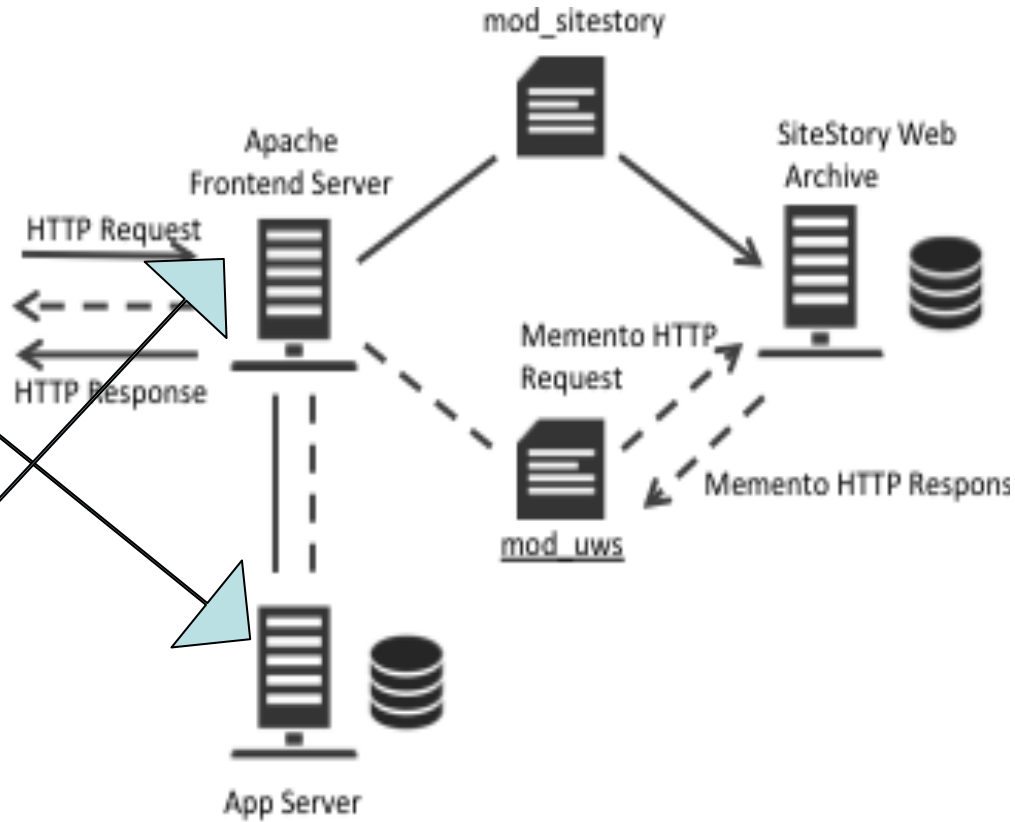
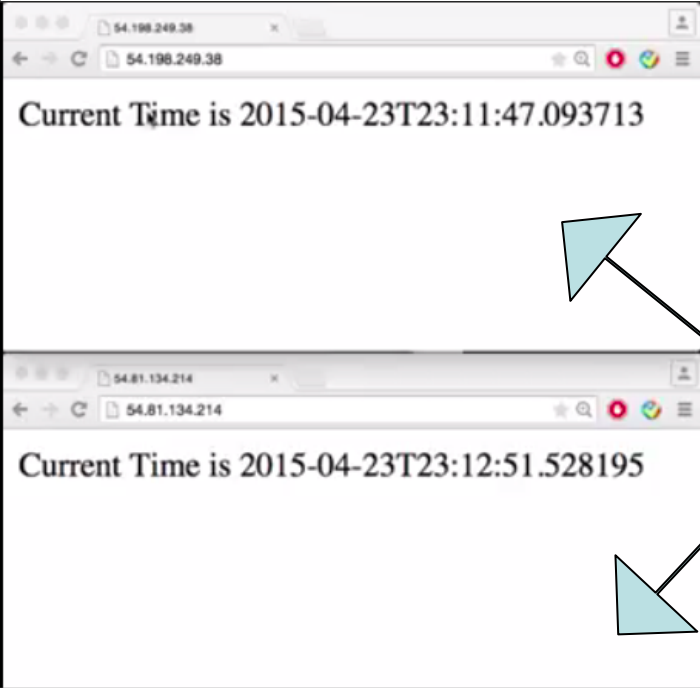
- Turn SiteStory into a web infrastructure component that improves website availability
- Mask server errors from the end user
- Gain precious time for webmasters to recover from application server failures without disrupting the majority of their users' web experience

UWS Architecture





Demonstration

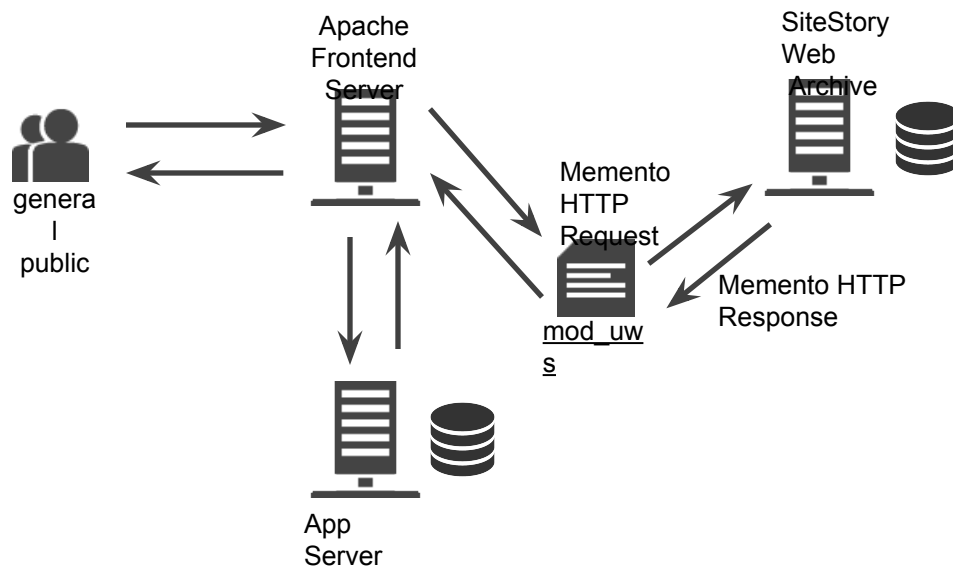




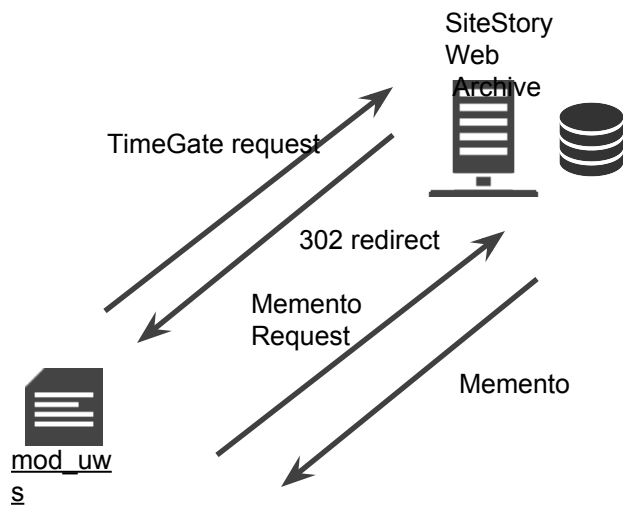
Demonstration

- [Demo link](#)

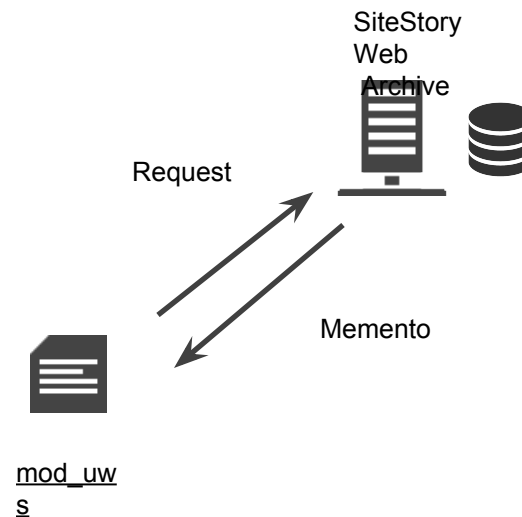
Performance Considerations



Memento Pattern 1 vs 4



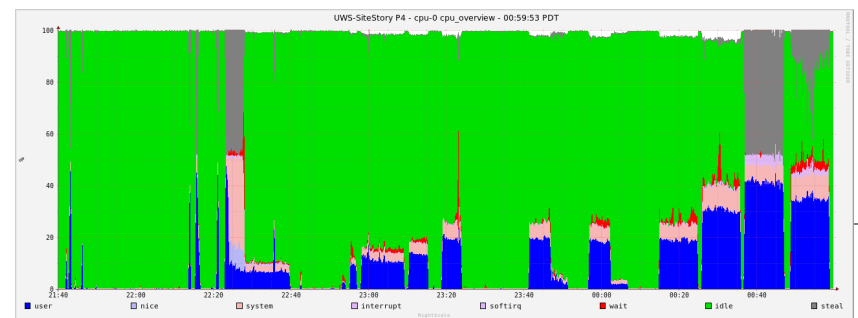
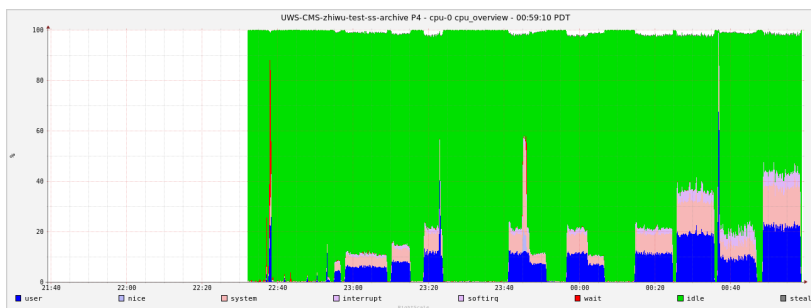
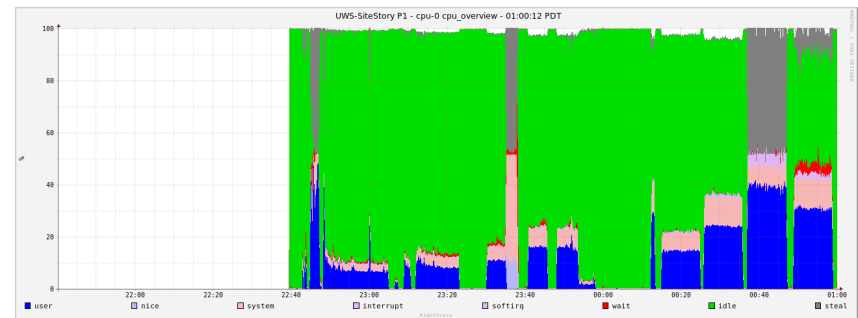
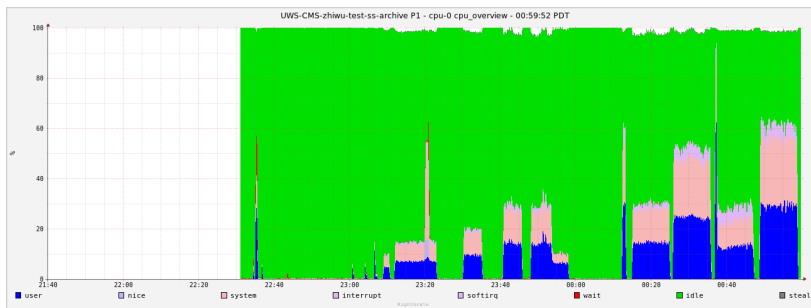
Pattern 1.1



Pattern 4

Experiment Results

- Reduces UWS load by about 1/4 to 1/3
- No significant SiteStory load increase





Other Web Components: Cache

- Efficiency tool, not a safety tool
- Not necessarily maintain a complete copy of the web site in cache
- Evict copies as necessary
- Frequently fall back to origin server to maintain semantic transparency



Other Web Components: Backup

- Granularity is too coarse
- Frequency is too low



Questions?

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