Paxos Made Simple

Phil Gibbons

15-712 F15

Lecture 22

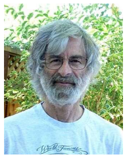
Today's Reminders

- **Interim Project Reports on Friday**
  - 2 teams have yet to sign up for a slot

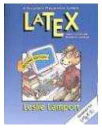
- **Office hours today, but no office hours next Wed**

Paxos Made Simple

Leslie Lamport [Sigact News 2001]



- **ACM Turing Award 2013:**
  - "For fundamental contributions to the theory and practice of distributed and concurrent systems, notably the invention of concepts such as causality and logical clocks, safety and liveness, replicated state machines, and sequential consistency."



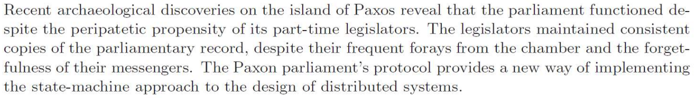
- **IEEE John von Neumann Award 2008**
  - Other winners: Brooks, Lampson

- **Dijkstra Prize (test-of-time award for distributed computing) in 2000, 2005, and 2014**

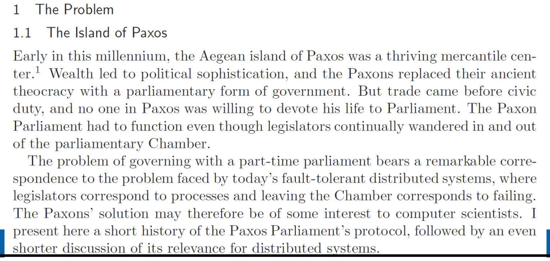
The Part-Time Parliament

[submitted 1990; published 1998]

Abstract



Recent archaeological discoveries on the island of Paxos reveal that the parliament functioned despite the peripatetic propensity of its part-time legislators. The legislators maintained consistent copies of the parliamentary record, despite their frequent forays from the chamber and the forgetfulness of their messengers. The Paxon parliament's protocol provides a new way of implementing the state-machine approach to the design of distributed systems.



1 The Problem

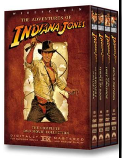
1.1 The Island of Paxos

Early in this millennium, the Aegean island of Paxos was a thriving mercantile center.[1] Wealth led to political sophistication, and the Paxons replaced their ancient theocracy with a parliamentary form of government. But trade came before civic duty, and no one in Paxos was willing to devote his life to Parliament. The Paxon Parliament had to function even though legislators continually wandered in and out of the parliamentary Chamber.

The problem of governing with a part-time parliament bears a remarkable correspondence to the problem faced by today's fault-tolerant distributed systems, where legislators correspond to processes and leaving the Chamber corresponds to failing. The Paxons' solution may therefore be of some interest to computer scientists. I present here a short history of the Paxos Parliament's protocol, followed by an even shorter discussion of its relevance for distributed systems.

"Inspired by my success at popularizing the consensus problem by describing it with Byzantine generals, I decided to cast the algorithm in terms of a parliament on an ancient Greek island. Leo Guibas suggested the name *Paxos* for the island. I gave the Greek legislators the names of computer scientists working in the field, transliterated with Guibas's help into a bogus Greek dialect.

Writing about a lost civilization allowed me to eliminate uninteresting details and indicate generalizations by saying that some details of the parliamentary protocol had been lost. To carry the image further, I gave a few lectures in the persona of an Indiana-Jones-style archaeologist, replete with Stetson hat and hip flask.



My attempt at inserting some humor into the subject was a dismal failure. People who attended my lecture remembered Indiana Jones, but not the algorithm. People reading the paper apparently got so distracted by the Greek parable that they didn't understand the algorithm." - Leslie Lamport

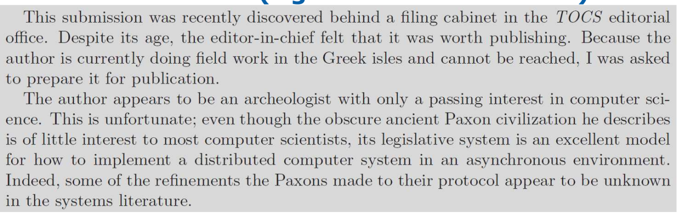
"I submitted the paper to *TOCS* in 1990. All three referees said that the paper was mildly interesting, though not very important, but that all the Paxos stuff had to be removed. I was quite annoyed at how humorless everyone working in the field seemed to be, so I did nothing with the paper."

[Real systems began using Paxos, follow-on papers were appearing, so Lamport tries to publish the same paper in 1998]

"Admittedly, the paper needed revision to take into account the work that had been published in the intervening years. As a way of both carrying on the joke and saving myself work, I suggested that instead of my writing a revision, it be published as a recently rediscovered manuscript, with annotations by Keith Marzullo."

The Part-Time Parliament [TOCS 1998]

Annotation (right after the abstract)



This submission was recently discovered behind a filing cabinet in the *TOCS* editorial office. Despite its age, the editor-in-chief felt that it was worth publishing. Because the author is currently doing field work in the Greek isles and cannot be reached, I was asked to prepare it for publication.

The author appears to be an archeologist with only a passing interest in computer science. This is unfortunate; even though the obscure ancient Paxon civilization he describes is of little interest to most computer scientists, its legislative system is an excellent model for how to implement a distributed computer system in an asynchronous environment. Indeed, some of the refinements the Paxons made to their protocol appear to be unknown in the systems literature.

"[Paxos] has become the standard for consistent, fault-tolerant state-machine replication, and is widely used in data centers to keep the state consistent despite failures and reconfiguration."
– SigOps HoF citation, 2012

Similar to: **Viewstamped Replication: A New Primary Copy Method to Support Highly-Available Distributed Systems [Brian Oki & Barbara Liskov, PODC'88]**

Implementing Replicated Logs with Paxos

https://www.youtube.com/watch?v=JEpsBg0AO6o

This is the link to the video we watched in class: a lecture on Paxos by John Ousterhout.

The lecture is part of the Raft User Study, an experiment to compare how students learn the Raft [Ongaro & Ousterhout, ATC'14] and Paxos consensus algorithms.