# Raft-Consensus-Algorithmus

# Wissenswertes

- Diego Ongaro und John Ousterhout
- Ph.D für Diego (2014)
- Ablösung des Paxos Algorithmus

# Consensus Algorithm

- Sicher
- Voll Funktional
- Keine Zeitabhängigkeit

# Basics

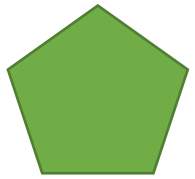
- 3 Zustände
- Terms (Zeitabschnitte)
- Wahlen
- Protokollierung
- Sicherheit

# Kommunikation

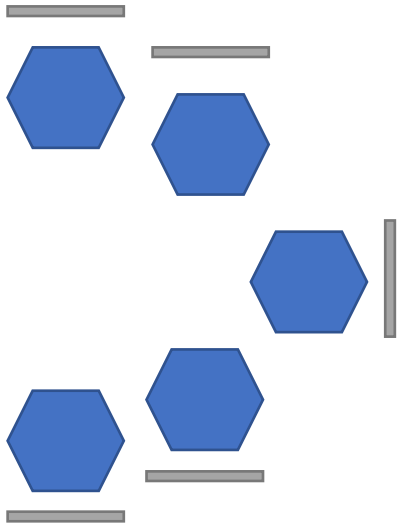
- RequestVote RPC
- AppendEntries RP
- InstallSnapshot RPC

# Beispiel

Client



Follower

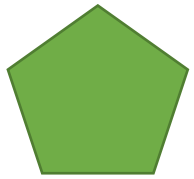


# Wahlphase

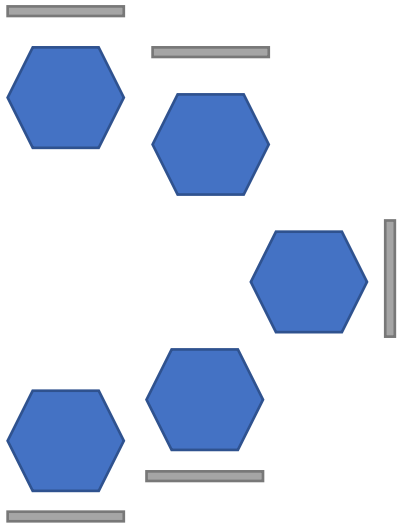
- Term beginnt mit einer Wahlphase
- Follower
- Wahl kann fehlschlagen
- Term wird incrementiert

# Beispiel

Client

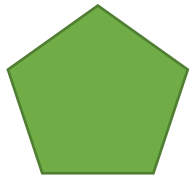


Follower

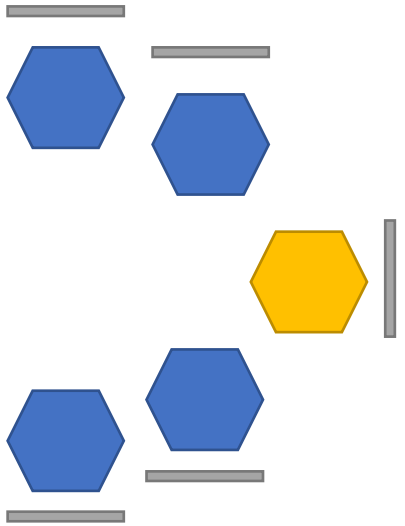


# Beispiel

Client



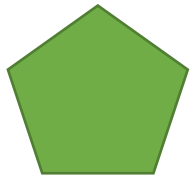
Follower



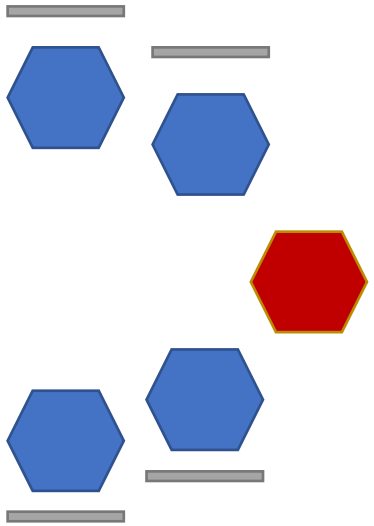
Candidate

# Beispiel

Client



Follower

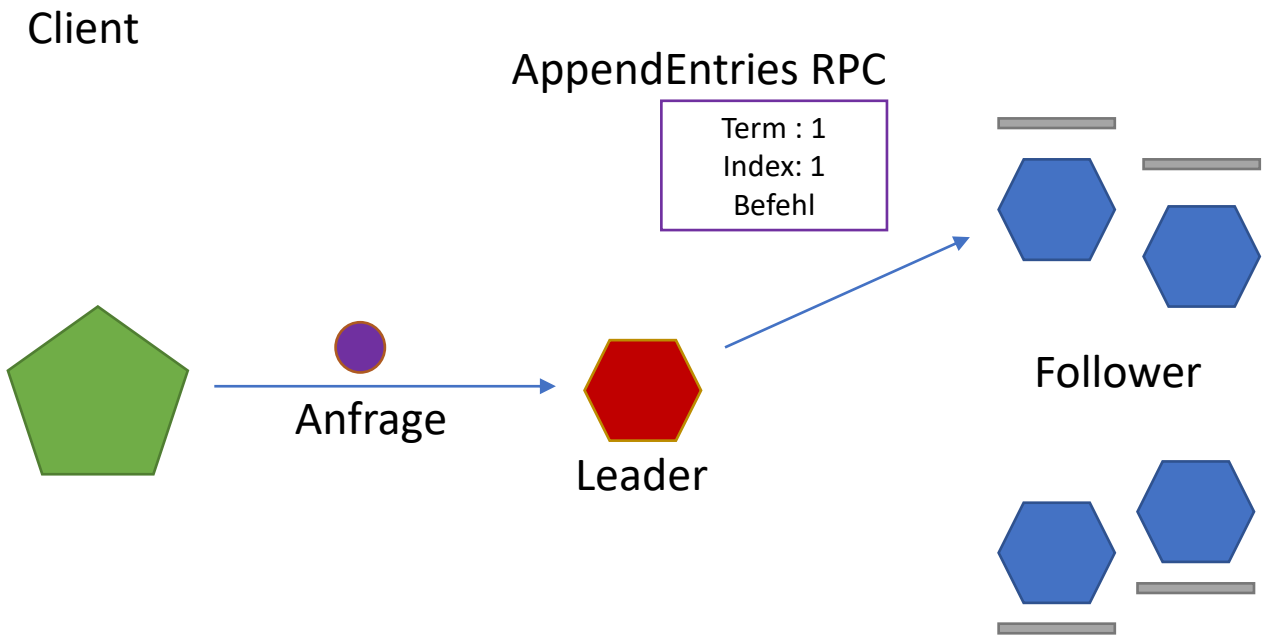


Leader

# Protokollierung

- Anfrage vom Client
- Leader merkt sich die Anfrage
- Replizierung auf die Follower
- Speichern im Log (Protokoll)

# Beispiel

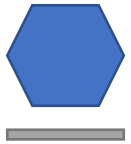

Client
AppendEntries RPC
Term : 1
Index: 1
Befehl
Anfrage
Leader
Follower

# Sicherheit

- Wahleinschränkung
- Anfragen aus vergangenen Terms
- Sicherheit
- Abstürze (Crashes)
- Timing und Verfügbarkeit

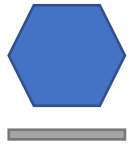
# Beispiel

Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy

Follower



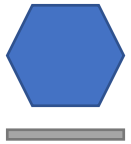
T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy
T: 3 I: 4 B:Copy

Follower

T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy

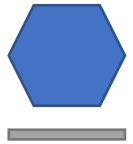
# Fall 1

Follower



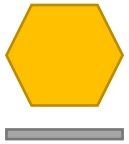
T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy

Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy
T: 3 I: 4 B:Copy

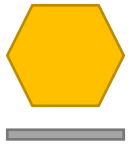
Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 3 I: 3 B:Copy

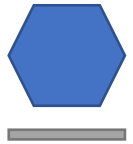
# Fall 2

Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy

Follower



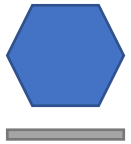
T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy
T: 3 I: 4 B:Copy

Follower

T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy

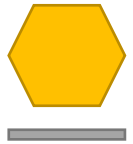
# Fall 3

Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy

Follower



T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy
T: 2 I: 3 B: Copy
T: 3 I: 4 B:Copy

Follower

T: 1 I: 1 B:Copy
T: 1 I: 2 B:Copy

# Sicherheit

- Wahleinschränkung
- Anfragen aus vergangenen Terms
- Abstürze (Crashes)
- Timing und Verfügbarkeit

# Anfragen aus vergangenen Terms

- Leader fällt aus bevor er die Anfrage commited
- Anfrage ist auf der Mehrheit der Server repliziert
- Neuer Leader wird den Eintrag im Log haben
- Sobald dieser eine neue Anfrage commited wird die alte ebenfalls commited

# Timing und Verfügbarkeit

- broadcastTime << electionTimeout << MTBF
- broadcastTime
  - Durchschnittliche Zeit die es braucht RPC´s an die Server zu senden und eine Antwort zu erhalten
- electionTimeout
  - Die Zeit, die jeder Server hat bevor er in den candidate Zustand wechselt
- MTBF
  - Durchschnittliche Zeit zwischen Fehlern eines einzelnen Servers

# Änderung in der Gruppe

- Passiert in zwei Schritten
- 1. Schritt: Joint Consensus
  - Kombination aus alter und neuer Konfiguration
- Wird über einen RPC vermittelt und als Log-Eintrag behandelt
- 3 Probleme können auftreten

# Zusammenfassung

- 3 Sever-States
- 1 Leader
  - Aktuellster Log
  - Ansprechpartner für den Client
- Follower
  - electionTimeout
  - Passiv
- Kommunikation über RPC´s
- Wahlen für einen Leader
- Logs speichern Anfragen vom Client

# Gibt es noch Fragen?