Method

- Data Source: VAST 2011 Mini Challenge 2 Firewall Log
  - ~12M events over 3 days
  - contains multiple attacks (denial of service, port scan)
- Preprocessing (reduces number of unique log entries)
  - Port: replaced with # of digits;
  - IP address: replaced with network role
  - Service[:port]: port omitted
  - Message code, source & destination host names omitted
- Each row in the firewall log is treated as a word
- Each 5 minute interval is a document (bag of words abstraction)
- Latent Dirichlet Allocation [1,4] used to learn topics
  - Documents are distributions over topics
  - Topics are distributions over words
  - Words are simplified firewall log entries
- Topics over time visualized (below) as a stacked graph [2,3]

Results

- Normal traffic pattern appears as topic 7
- Medium to long duration attacks also apparent
- Additional network events present

Future Work

- Fuse multiple data sources (i.e., Firewall + IDS + PCAP)
- Train topics on normal traffic and/or known attacks
- Interactive D3 version

References