Exclusion Persistence in Spatial Data

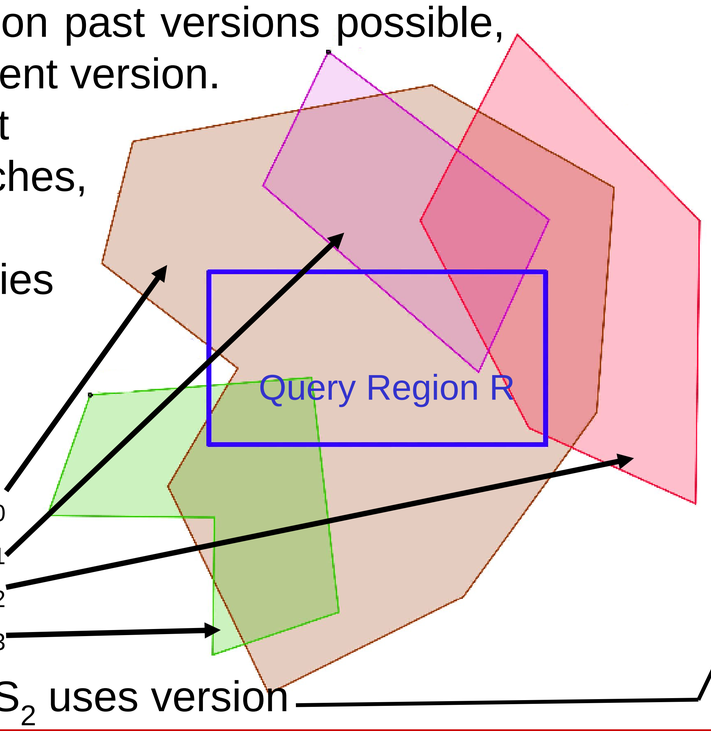
Stuart A. MacGillivray and Bradford G. Nickerson
Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada

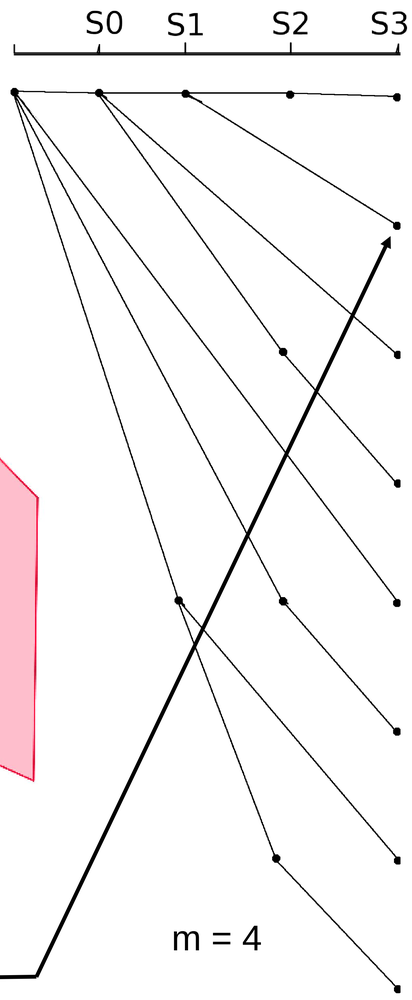
- **Motivation:** Efficient search of massive geographically referenced data surveys

Persistent Data Structures

- Persistent data structures maintain version history.
- Alterations tracked between versions of structure.
- Multiple styles of persistence for different methods of data management, e.g. source code version control.
- Partial persistence: Queries on past versions possible, can only modify the most recent version.
- Full persistence: Edits to past versions create alternate branches, forming a tree of versions.
- 'Exclusion' persistence: Queries can ignore any subset of past updates.

Set S_0, time t_0
Set S_1, time t_1
Set S_2, time t_2
Set S_3, time t_3

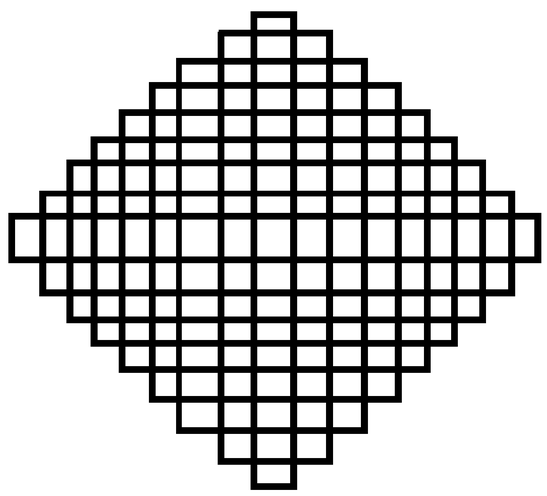




'Exclusion' query at t_3 omitting S_2 uses version

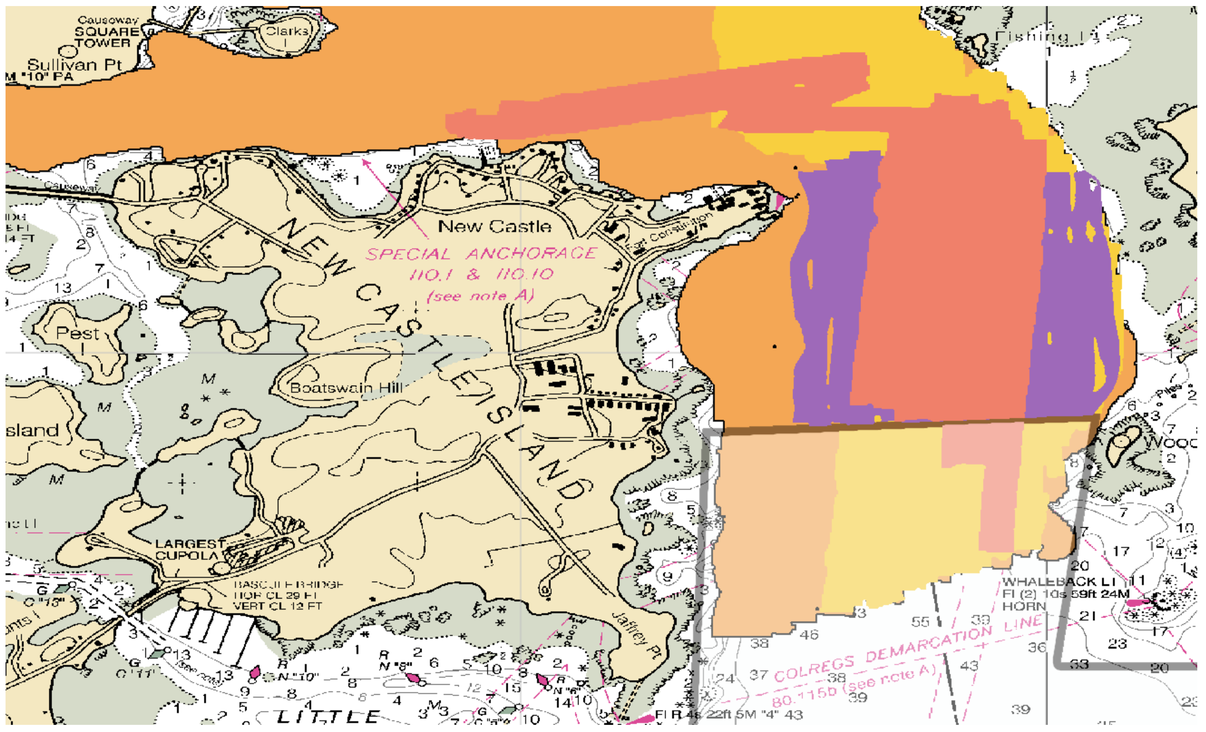
Approaches

- Given N points in m sets S_i of changes made at times t_i, exclusion persistence allows queries $Q = (R, t_q, S_i \mid t_i \leq t_q, t_i \notin T_e)$, where T_e is the set of times excluded.
- Challenges: Minimizing storage space and query I/Os simultaneously.
- Range search in $O(\log N/B + K/B)$ I/Os with $O((N2^{m-1}\log N)/(B \log\log N))$ storage space by storing all possible version combinations.
- $O(N/B)$ storage space possible with worst case searches returning K points in $O(m(N/B)^{1/2} + mK/B)$ I/Os, with each S_i stored independently, searches done on each S_i and data merged/pruned to find latest result.
- Overlapping 2-d data sets might be searchable in $O((mN/B)^{1/2}+K/B)$ I/Os with linear space, by using stack-based indexing of overlapping areas.
- With m rectangular regions, number of subregions is $(2m^2-2m+1)$ in the worst case.



Test Data

- Data sets: Shallow Survey 2008 Common Dataset (CCOM/JHC), as well as a synthetic data set.
- SSCD contains >580 GB of survey data; sets of 4-D point data range from 2 to 28 GB.



From Shallow Survey 2008. http://www.shallowsurvey2008.org/

Sponsored by: UNB Faculty of COMPUTER SCIENCE; NSERC CRSNG