Efficient Tamper-Evident Data Structures for Untrusted Servers

Dan S. Wallach
Rice University

Joint work with Scott A. Crosby

This talk vs. Preneel’s talks

- Preneel: how hash functions work (or don’t work)
- This talk: interesting things you can build with hash functions (assumption: “ideal” hash functions)

This talk isn't about…

- BitCoin and other blockchain currencies
- CA certificate revocation infrastructure
- Voting system "public bulletin boards"

All of these systems are built around similar hash-based data structure primitives.

Problem

- Lots of untrusted servers
  - Outsourced
    - Backup services
    - Publishing services
    - Outsourced databases
  - Insiders
    - Financial records
    - Forensic records
  - Hackers

Limitations and goals

- Limitation
  - Untrusted server can do anything
- Best we can do
  - Tamper evidence
- Goal:
  - Tamper-evident primitives
    - Efficient
    - Secure

Tamper-evident primitives

- Classic
  - Merkle tree [Merkle 88]
  - Digital signatures

- More interesting ones
  - Tamper-evident logs [Kelsey and Schneier 99]
  - Authenticated dictionaries [Naor and Nissim 98]
  - Graph and geometric searching [Goodrich et al 03]
  - Searching XML documents [Devanbu et al 04]

Tamper-evident logging

- Security model
  - Mostly untrusted clients
  - Untrusted log server
  - Trusted auditors
    - Detect tampering
- Useful for
  - Election results
  - Financial transactions
  - General-purpose system logging

Authenticated dictionaries

- Security model
  - Data produced by trusted authors
  - Stored on untrusted servers
  - Fetched by clients
- Key-value data store
- Useful for
  - Price lists
  - Crypto key revocation
  - DNS / other databases

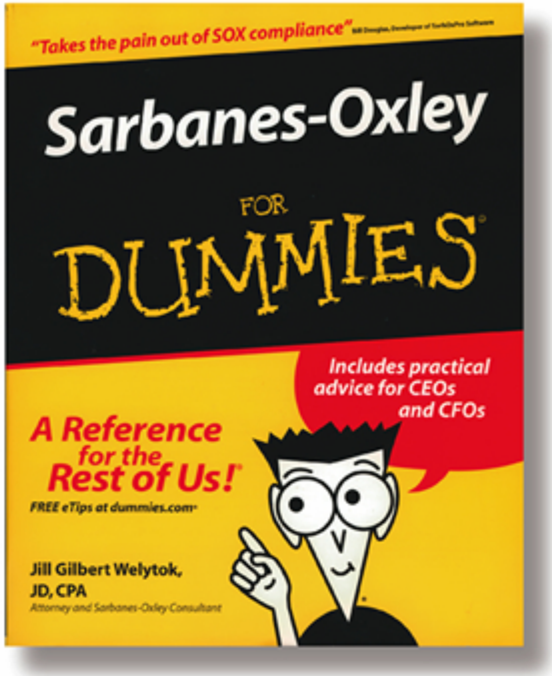
Our research

- Investigate two data structure problems
  - Persistent authenticated dictionary (PAD)
    - Efficiency improves from O(log *n*) to O(1)
  - Comprehensive PAD benchmarks
  - Tamper-evident log
    - Efficiency improves from O(*n*) to O(log *n*)
    - Newer work on fast digital signatures
- Code and papers online

  http://tamperevident.cs.rice.edu

Tamper Evident Logging

Everyone has logs



"Takes the pain out of SOX compliance"

Sarbanes-Oxley

FOR DUMMIES

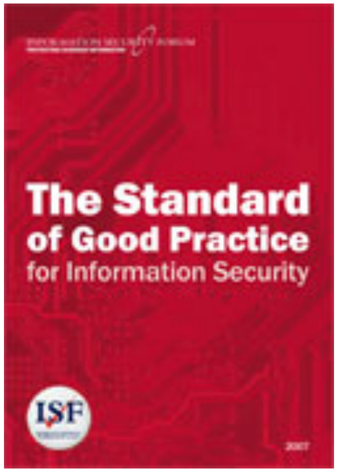
Includes practical advice for CEOs and CFOs

A Reference for the Rest of Us!

FREE eTips at dummies.com

Jill Gilbert Welytok, JD, CPA

Attorney and Sarbanes-Oxley Consultant



The Standard of Good Practice for Information Security

ISF

2007

PCI Security Standards Council

HEALTH INSURANCE PORTABILITY and ACCOUNTABILITY ACT

HIPAA

ADMINISTRATIVE SIMPLIFICATION: PRIVACY, SECURITY, TRANSACTIONS

Current solutions

- ‘Write only’ hardware appliances
- Security depends on correct operation

- Would like cryptographic techniques
  - Logger **proves** correct behavior
  - Existing approaches too slow

Our solution

- History tree
  - Logarithmic for all operations
  - Benchmarks at >1,750 events/sec
  - Benchmarks at >8,000 audits/sec

  (on 2007 hardware!)

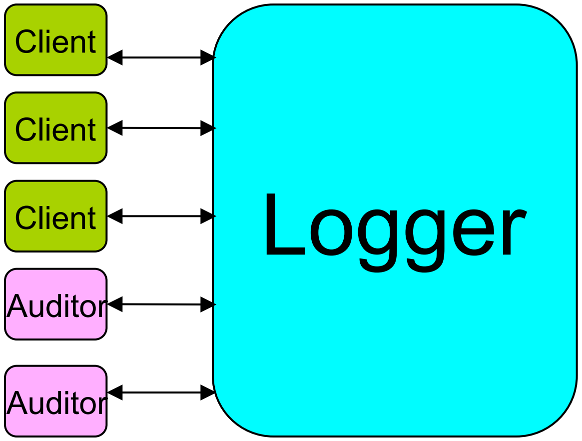
- In addition
  - Propose new threat model
  - Demonstrate the importance of auditing

Threat model

- Strong insider attacks
  - Malicious administrator
    - Evil logger
  - Users collude with administrator

- Prior threat model
  - Forward integity [Bellare et al 99]
  - Log tamper evident up to (unknown point), and untrusted thereafter

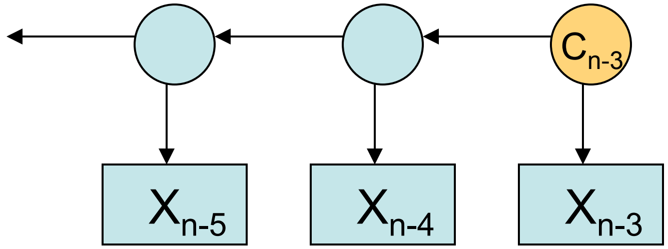
System design



- Logger
  - Stores events
  - Never trusted
- Clients
  - Little storage
  - Create events to be logged
  - Trusted only at time of event creation
  - Sends commitments to auditors
- Auditors
  - Verify correct operation
  - Little storage
  - Trusted, at least one is honest

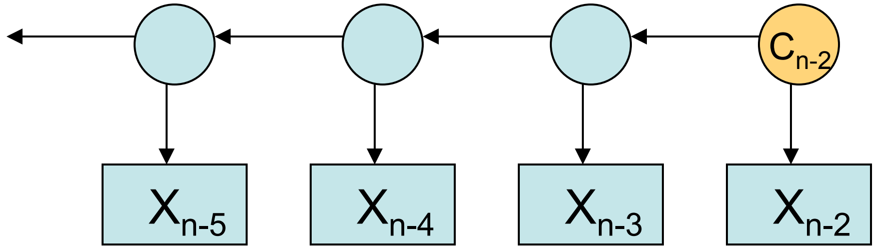
Hash chain log

- Existing approach [Kelsey and Schneier 98]
  - $C_n = H(C_{n-1} \,||\, X_n)$
  - Logger signs C_n



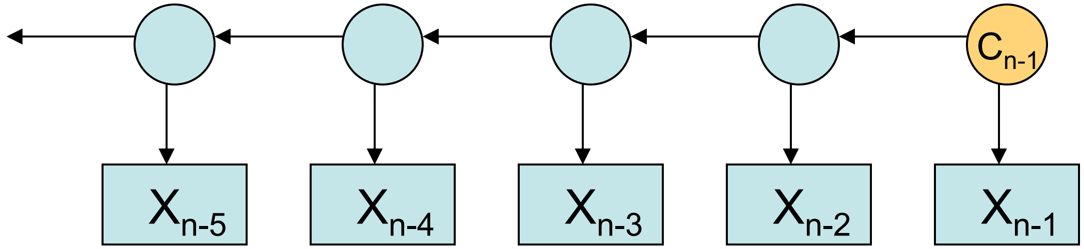
Hash chain log

- Existing approach [Kelsey,Schneier]
  - $C_n = H(C_{n-1} \| X_n)$
  - Logger signs C_n



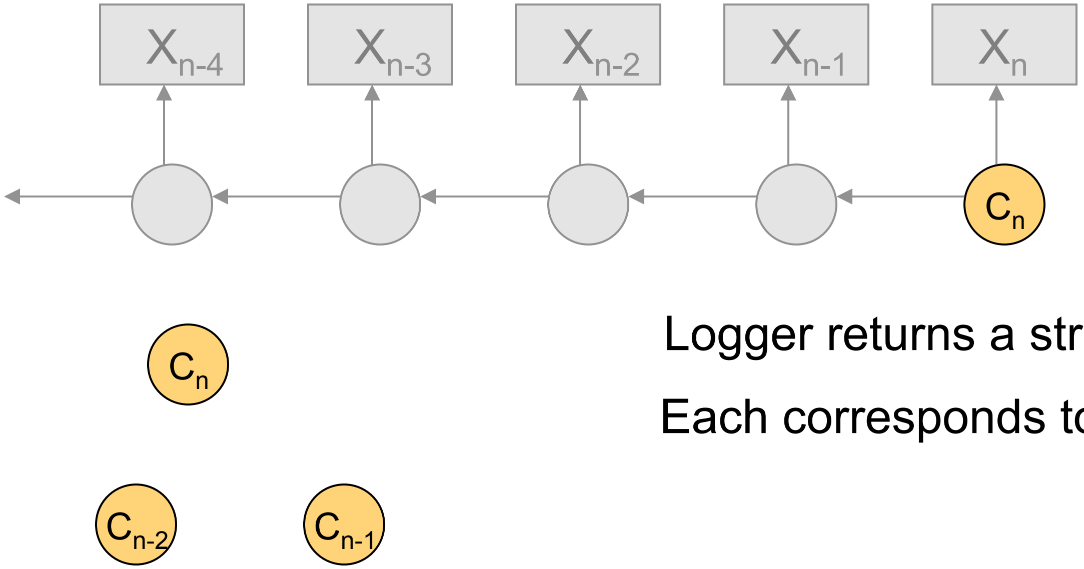
Hash chain log

- Existing approach [Kelsey,Schneier]
  - $C_n = H(C_{n-1} \| X_n)$
  - Logger signs C_n



Problem

- We don't trust the logger!

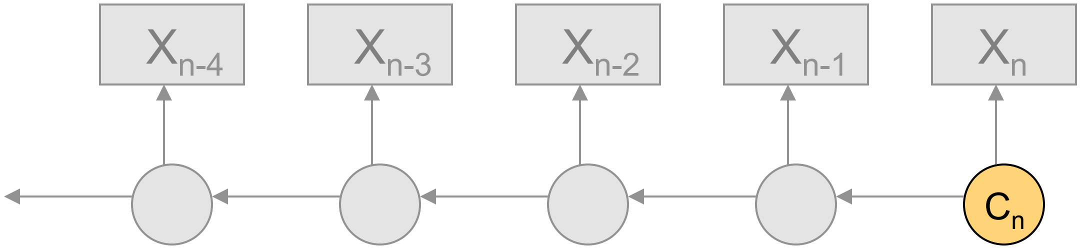


Logger returns a stream of commitments

Each corresponds to a log

Problem

- We don’t trust the logger!



Does C_n really contain the just inserted X_n ?

Do C_{n-2} and C_{n-1} really commit the same historical events?

Is the event at index *i* in log C_n really X_i ?

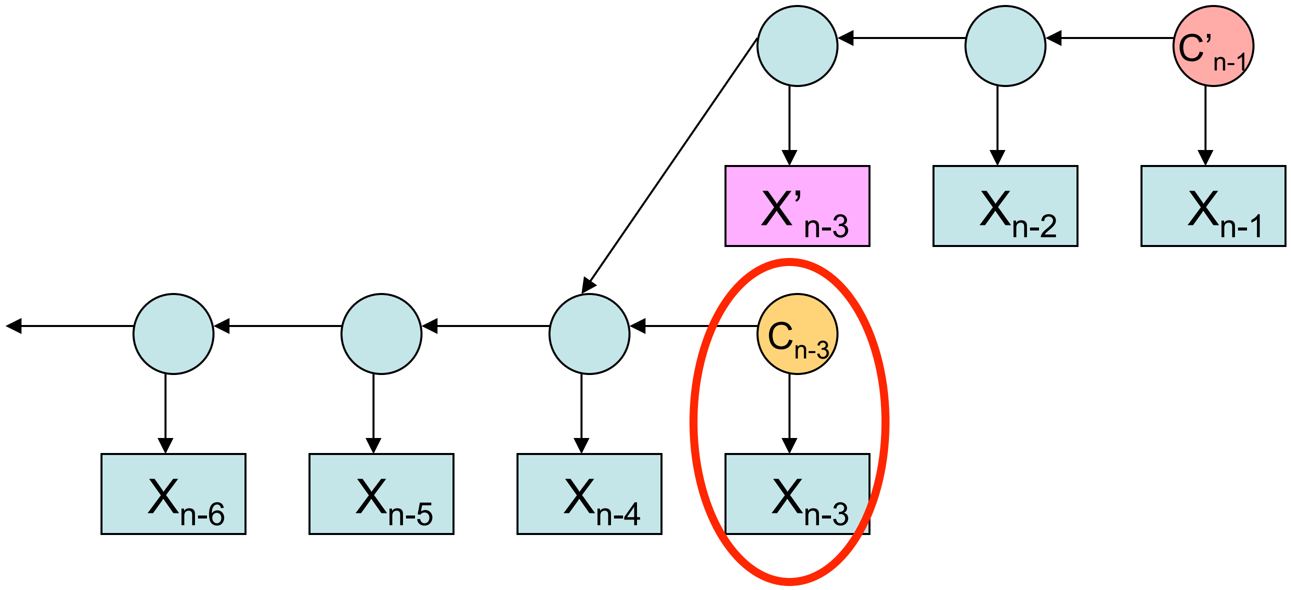
Solution

- Auditors check the returned commitments
  - For consistency $C_{n-2} \equiv C_{n-1}$
  - For correct event lookup $X_{n-3} \in C_{n-3}$

- Previously
  - Auditing = looking at historical events
    - Assumed to infrequent
    - Performance was ignored

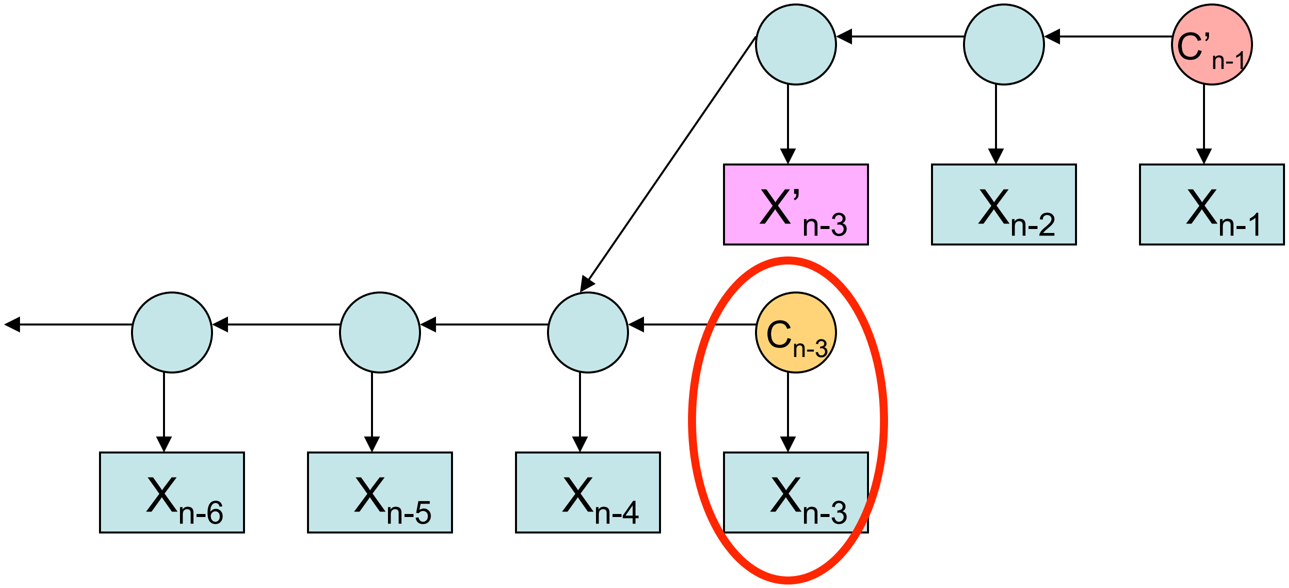
Auditing is a frequent operation

- If the logger knows this commitment will not be audited for consistency with a later commitment.



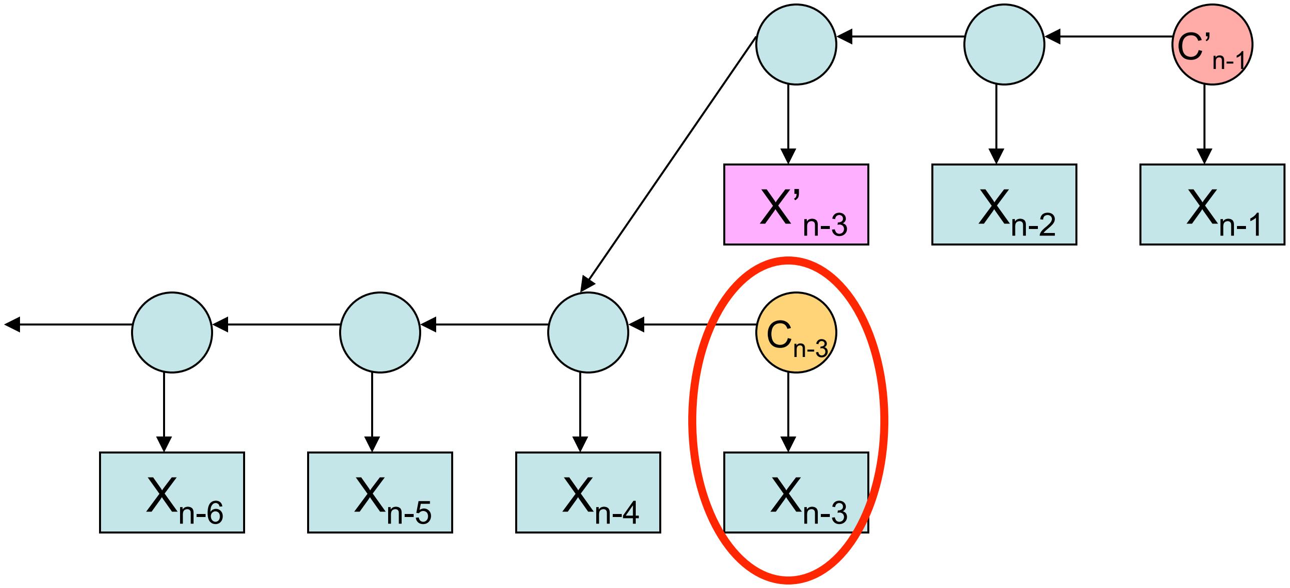
Auditing is a frequent operation

- Successfully tampered with a ‘tamper evident’ log



Auditing is a frequent operation

- Every commitment must have a non-zero chance of being audited



New paradigm

- Auditing cannot be avoided
- Audits should occur
  - On every event insertion
  - Between commitments returned by logger
- How to make inserts *and audits* cheap
  - CPU
  - Communications complexity
  - Storage

Two kinds of audits

- Membership auditing $X_i \in C_n$
  - Verify proper insertion
  - Lookup historical events
- Incremental auditing $C_i \equiv C_n$
  - Prove consistency between two commitments

Existing tamper evident log designs

- Hash chain [Kelsey and Schneier 98]
  - Auditing is linear time
  - Historical lookups
    - Very inefficient

- Skiplist history [Maniatis and Baker 02]
  - Auditing is still linear time
  - O(log *n*) historical lookups

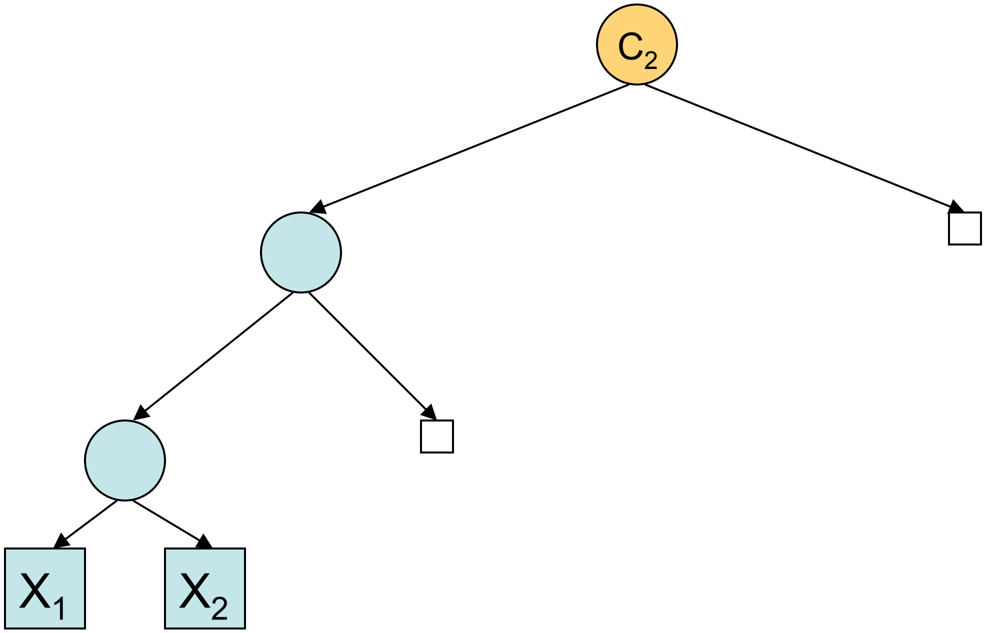
Our solution

- History tree
  - O(log *n*) instead of O(*n*) for all operations
  - Variety of useful features
    - Write-once append-only storage format
    - Predicate queries + safe deletion
    - May probabilistically detect tampering
      - Auditing random subset of events
      - Not beneficial for skip-lists or hash chains

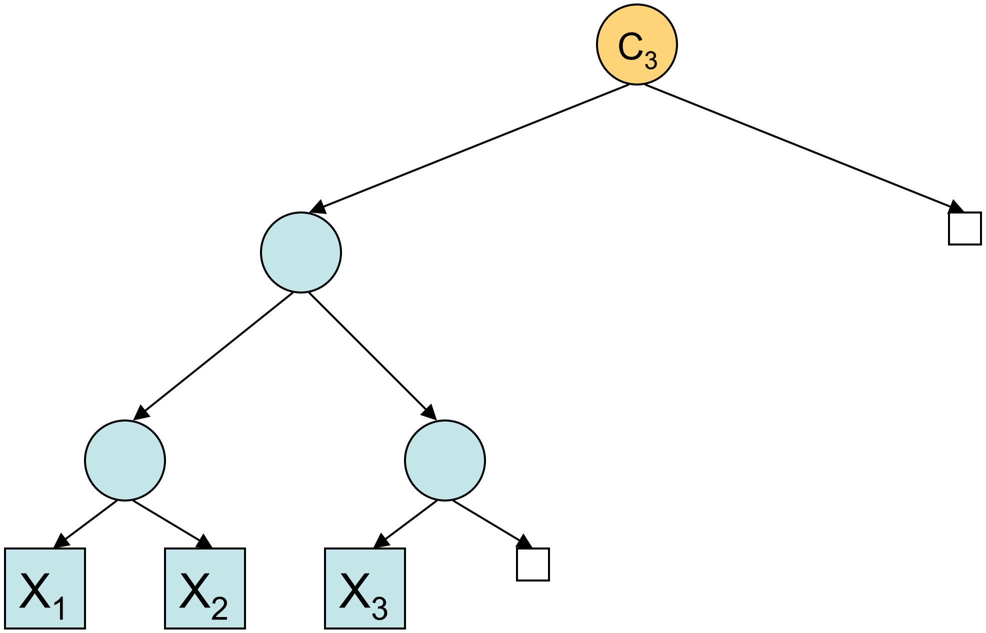
History tree

- Merkle binary tree
  - Events stored on leaves
  - Logarithmic path length
    - Random access
  - Permits reconstruction of past version and past commitments

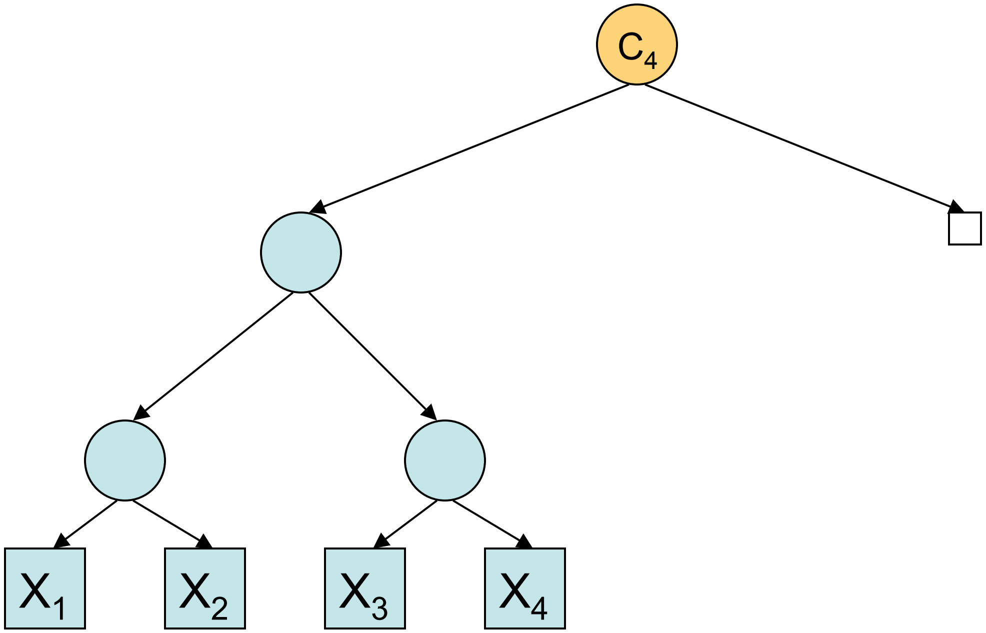
History tree



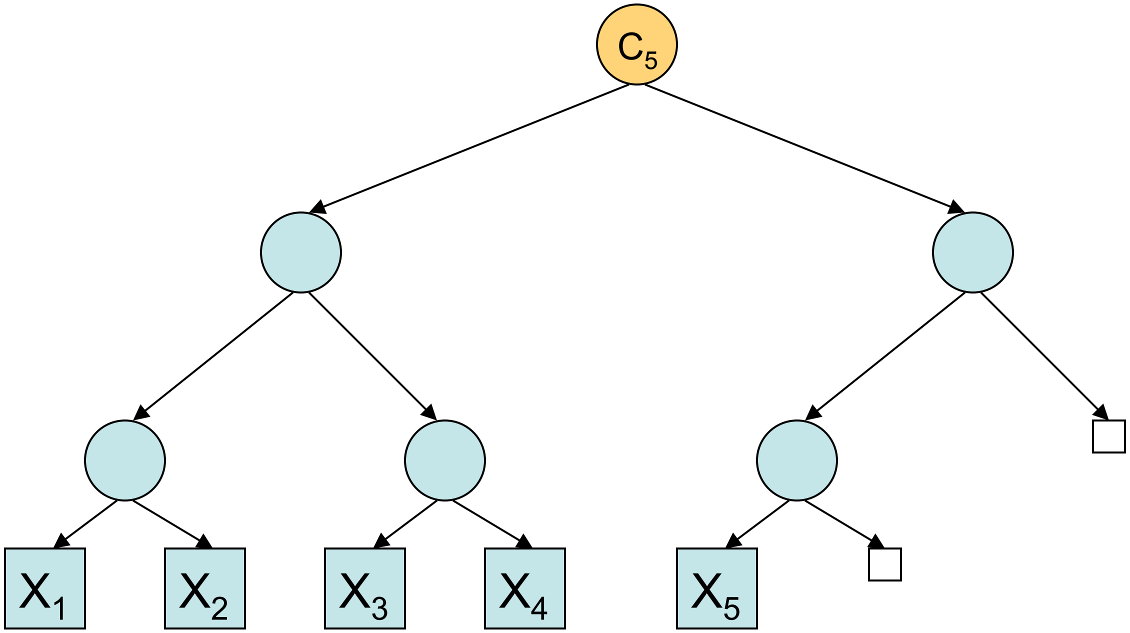
History tree



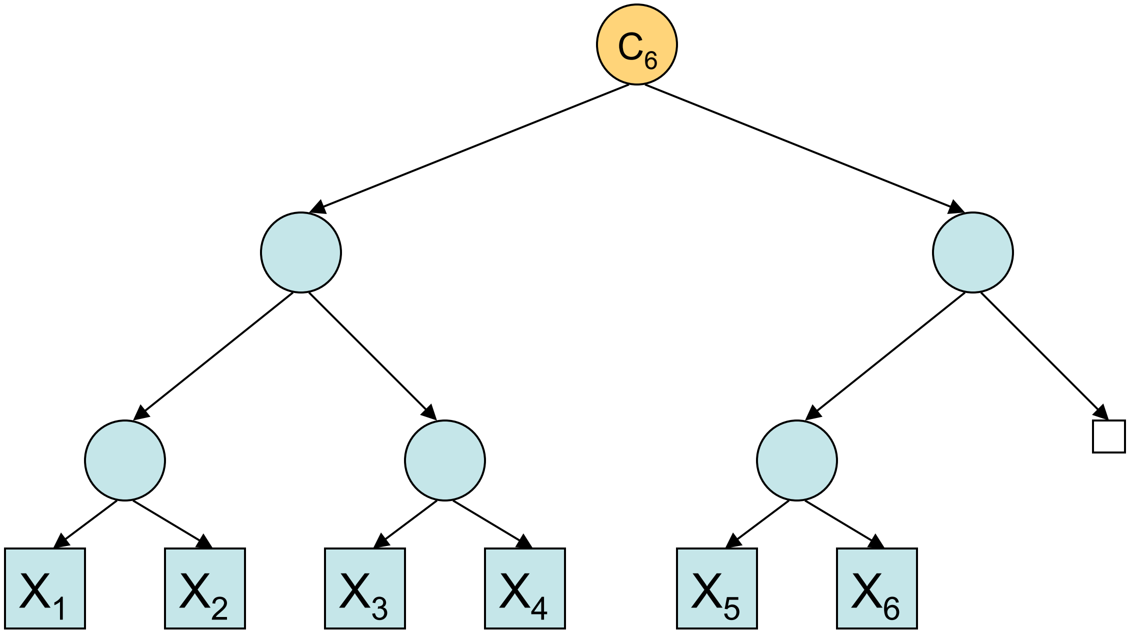
History tree



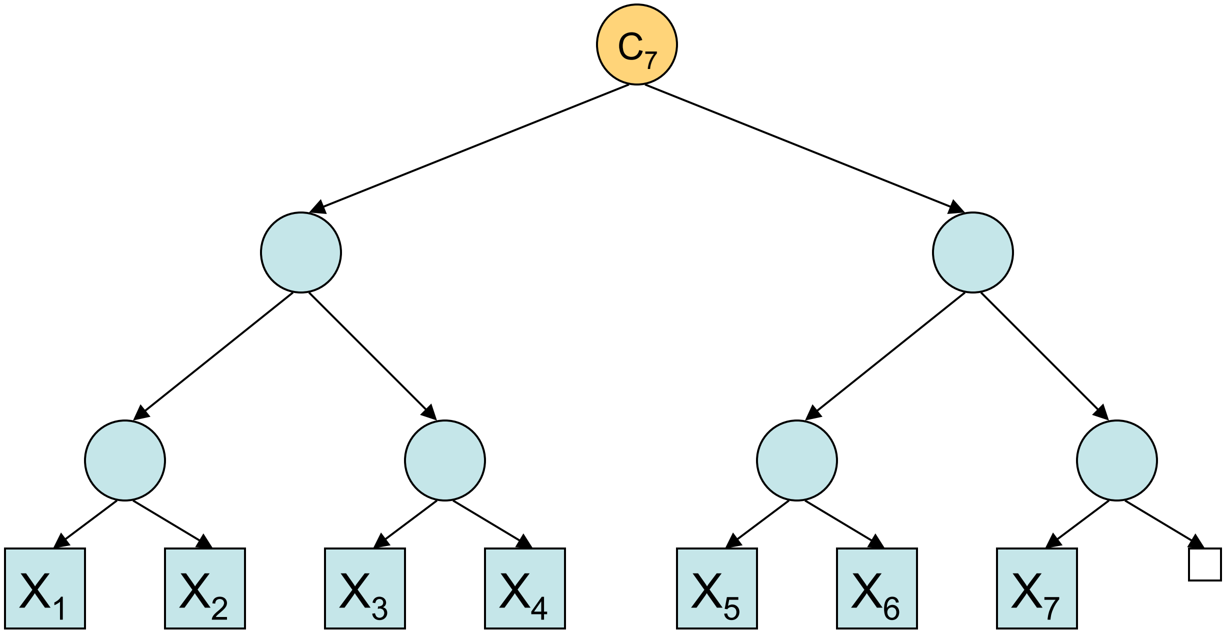
History tree



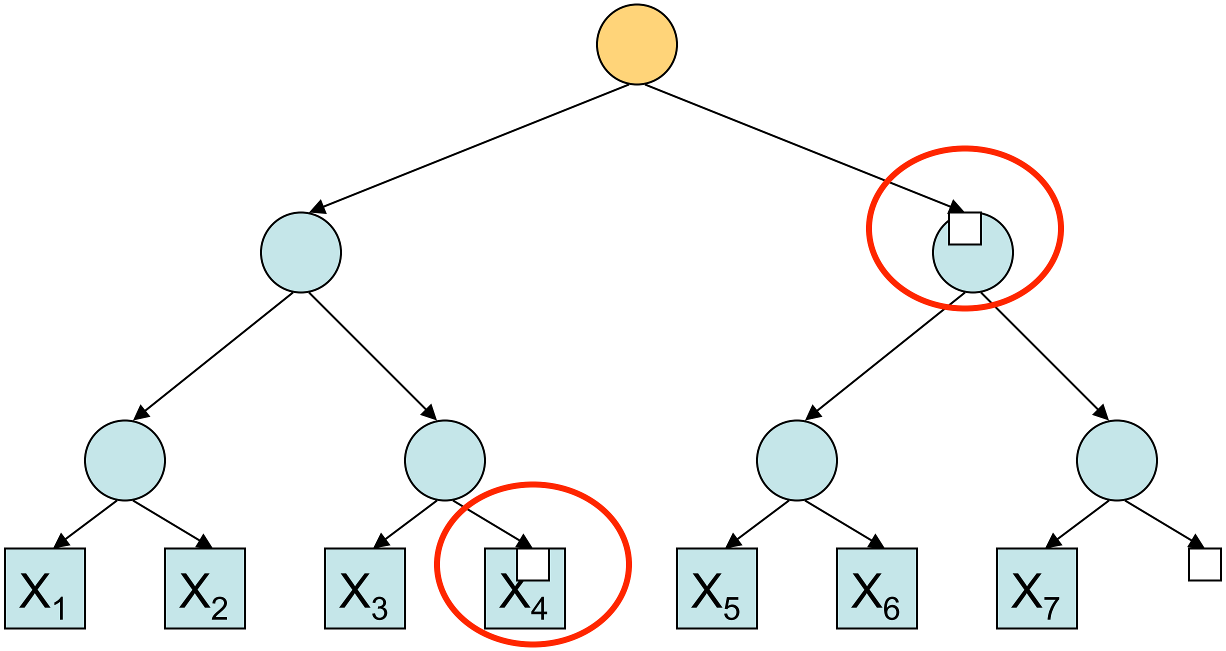
History tree



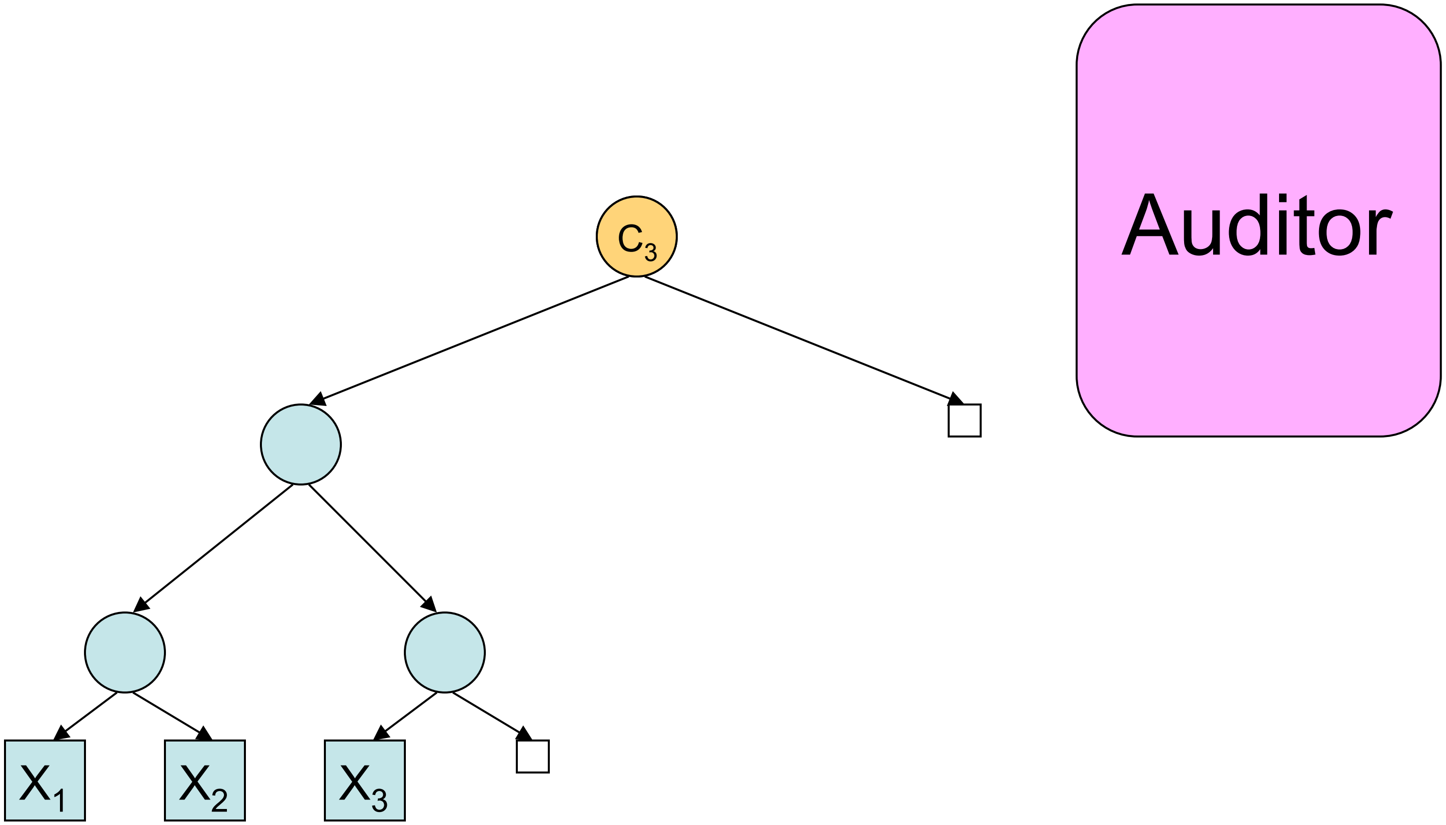
History tree

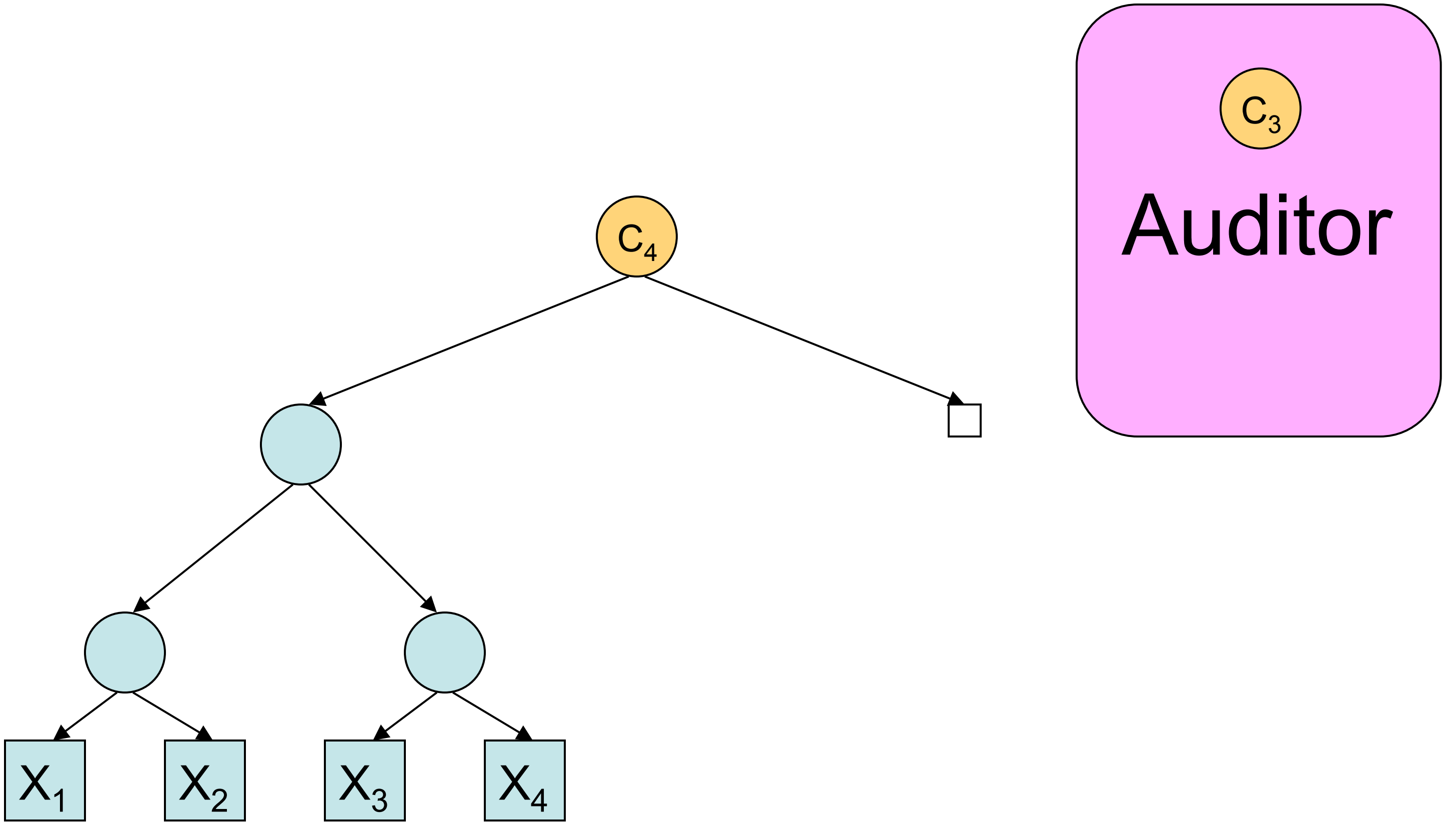


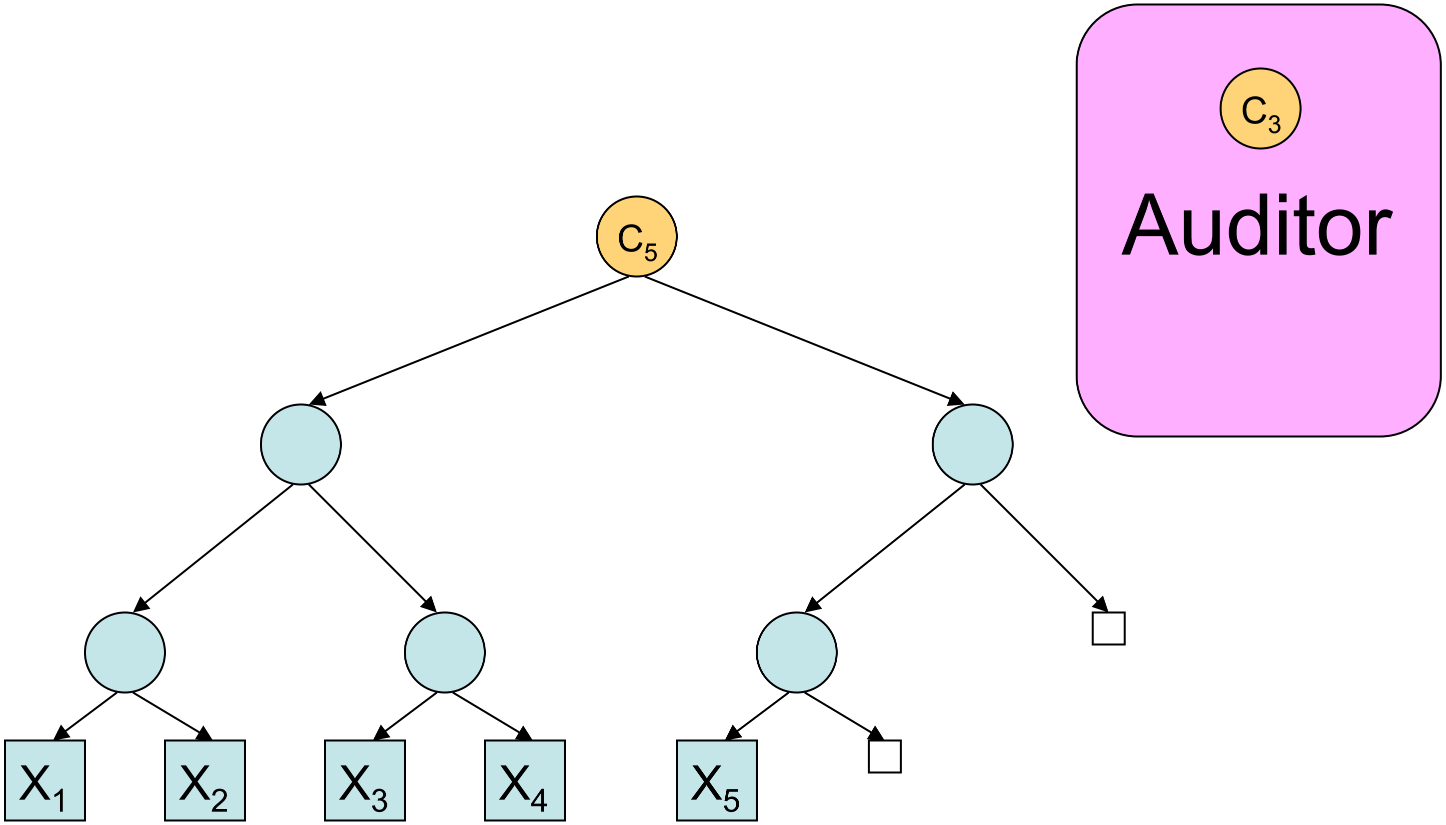
History tree

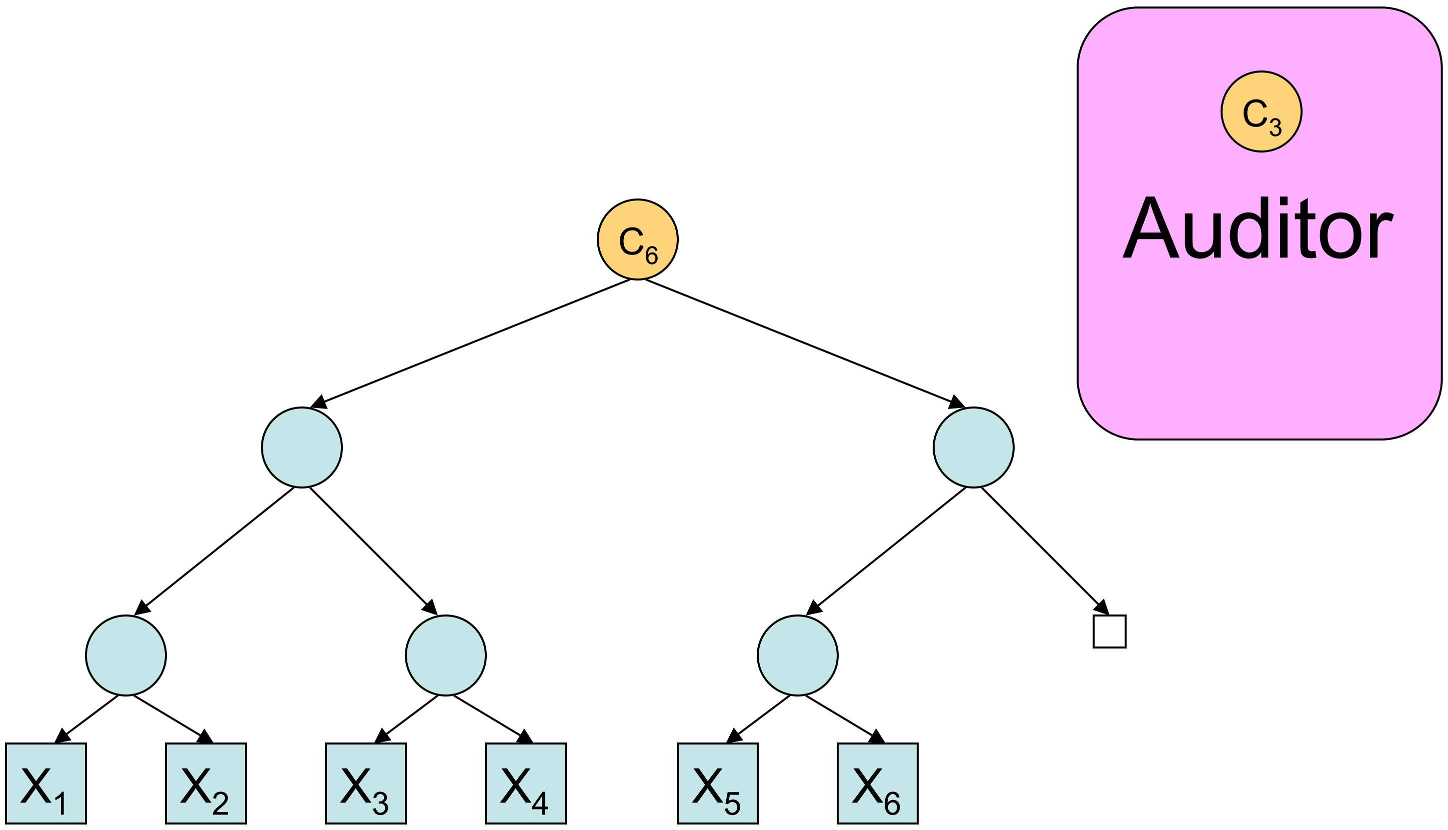


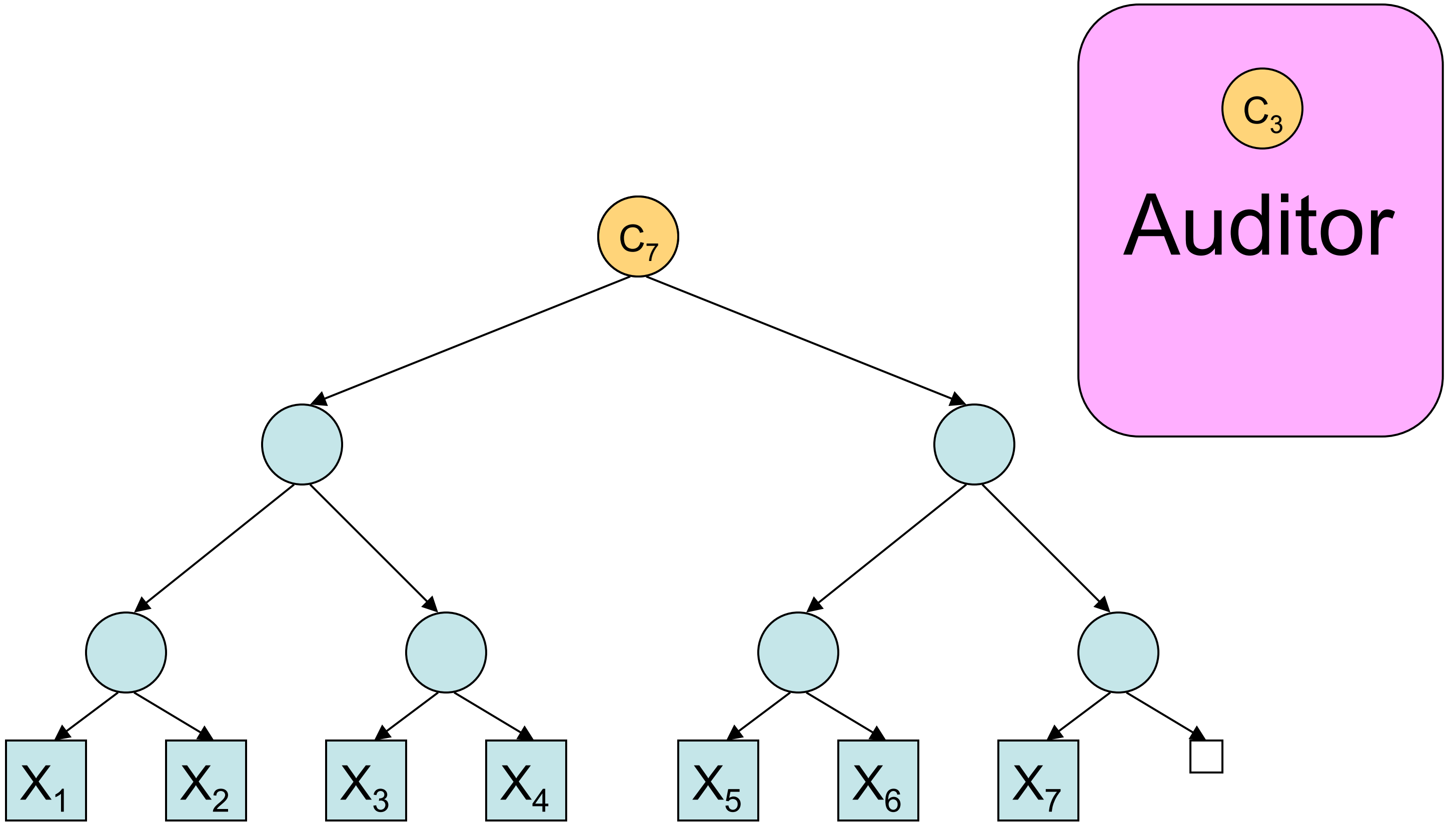
Incremental auditing


c3
X1
X2
X3
Auditor

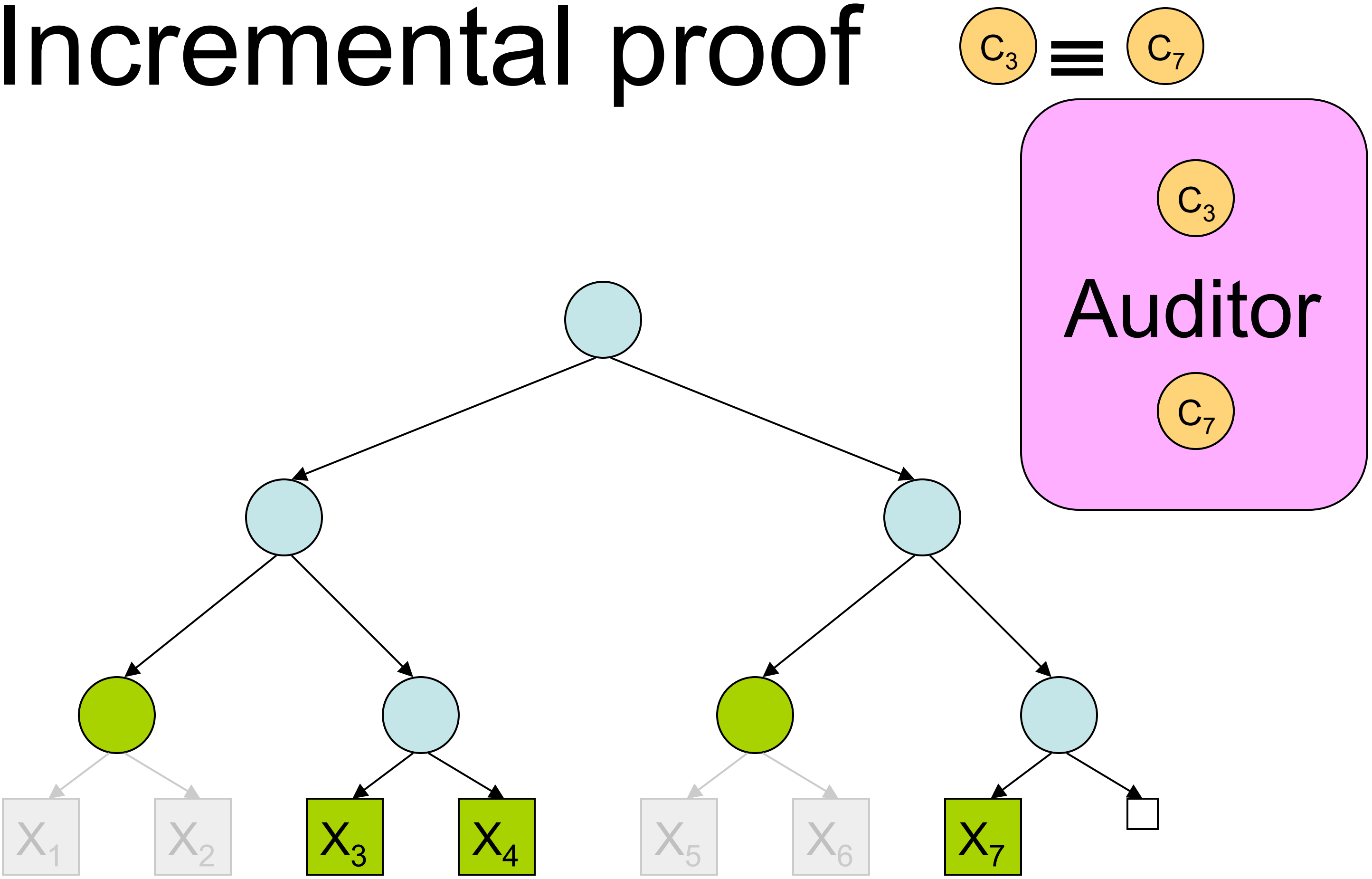

C_4
X_1
X_2
X_3
X_4
C_3
Auditor


c_3
Auditor
c_5
X_1
X_2
X_3
X_4
X_5


C_6
C_3
Auditor
X_1
X_2
X_3
X_4
X_5
X_6

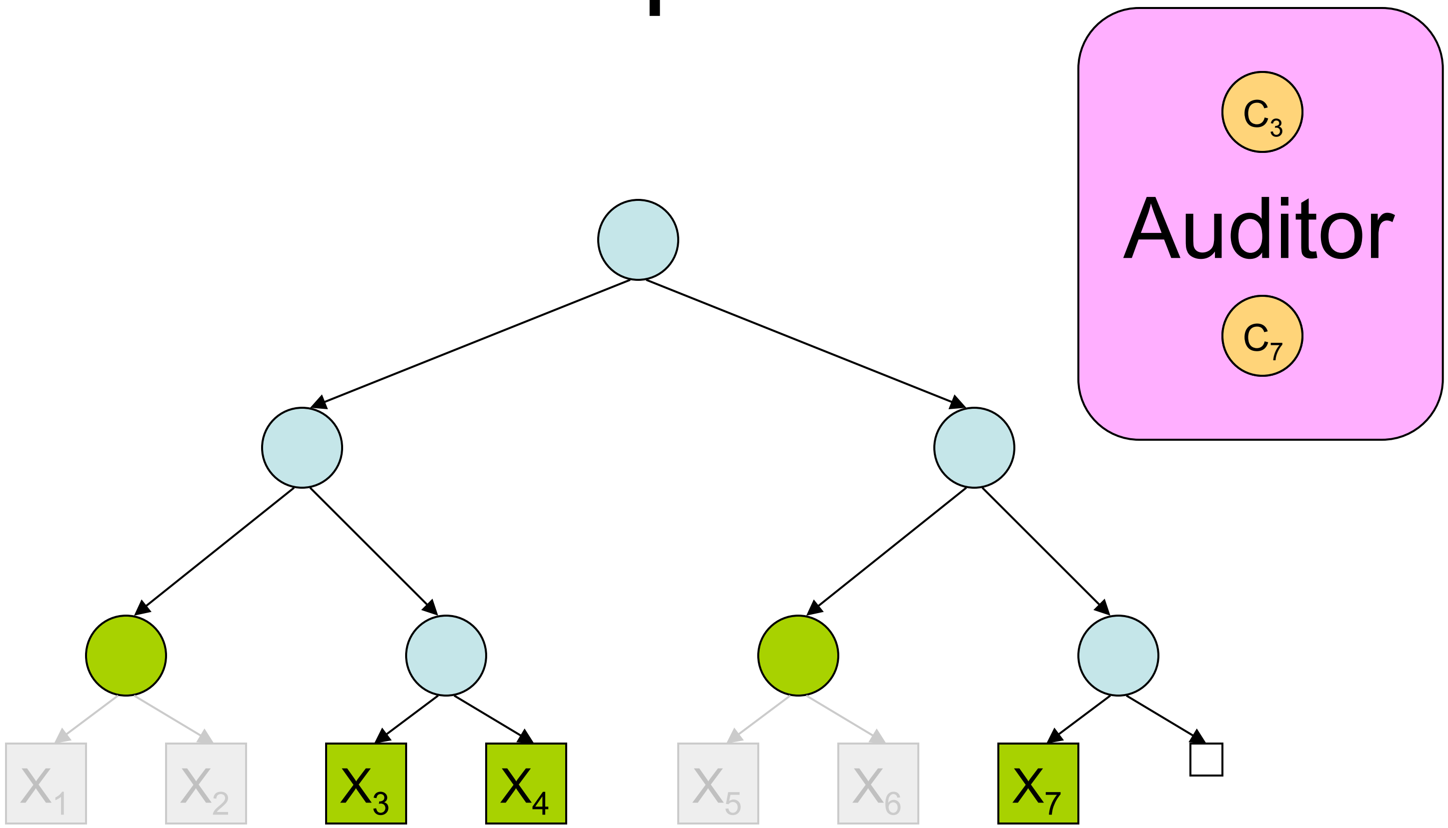

c_7
X_1
X_2
X_3
X_4
X_5
X_6
X_7
c_3
Auditor

Incremental proof



Incremental proof

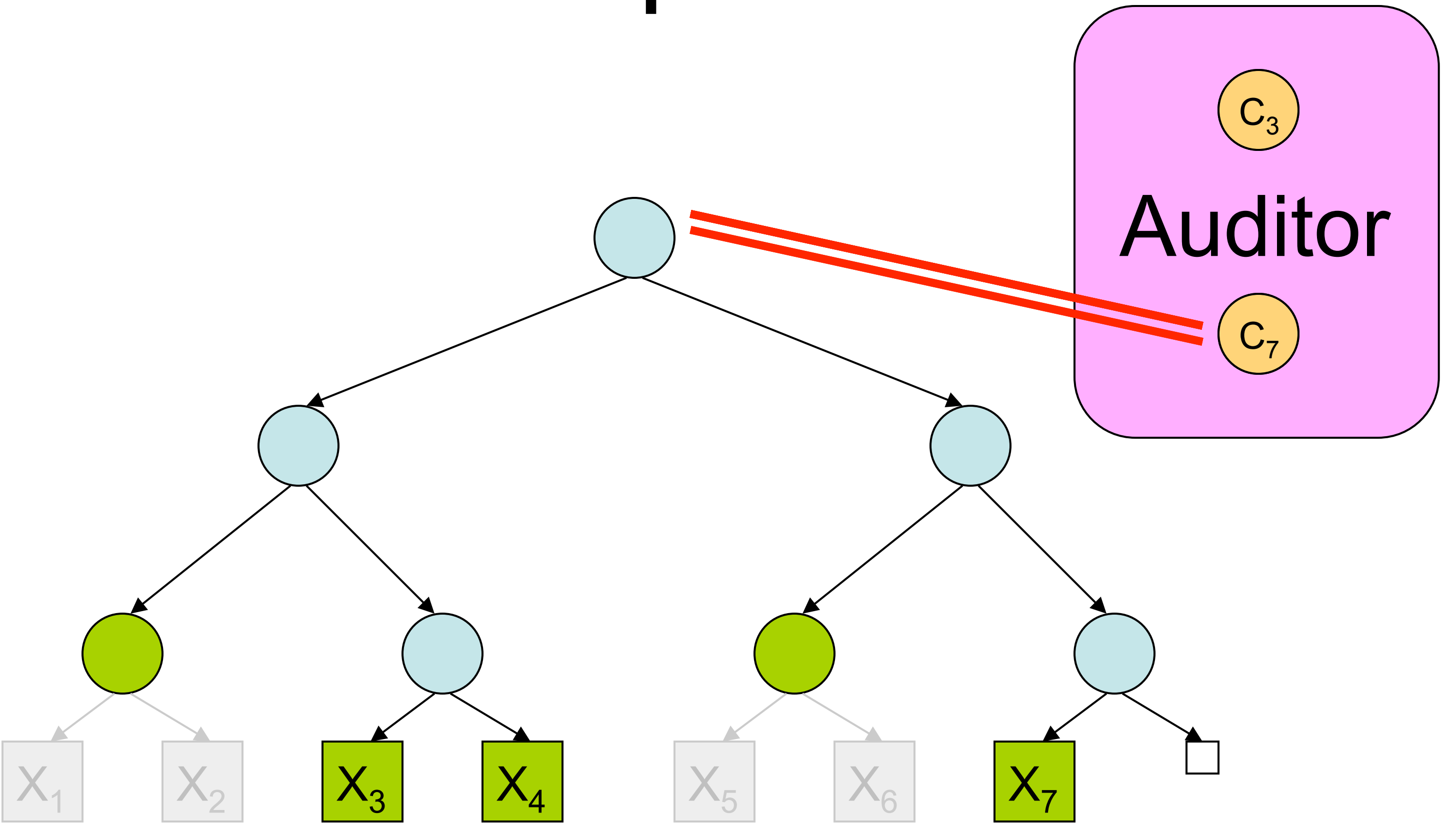
$c_3 \equiv c_7$



- P is consistent with c_7
- P is consistent with c_3
- Therefore c_7 and c_3 are consistent.

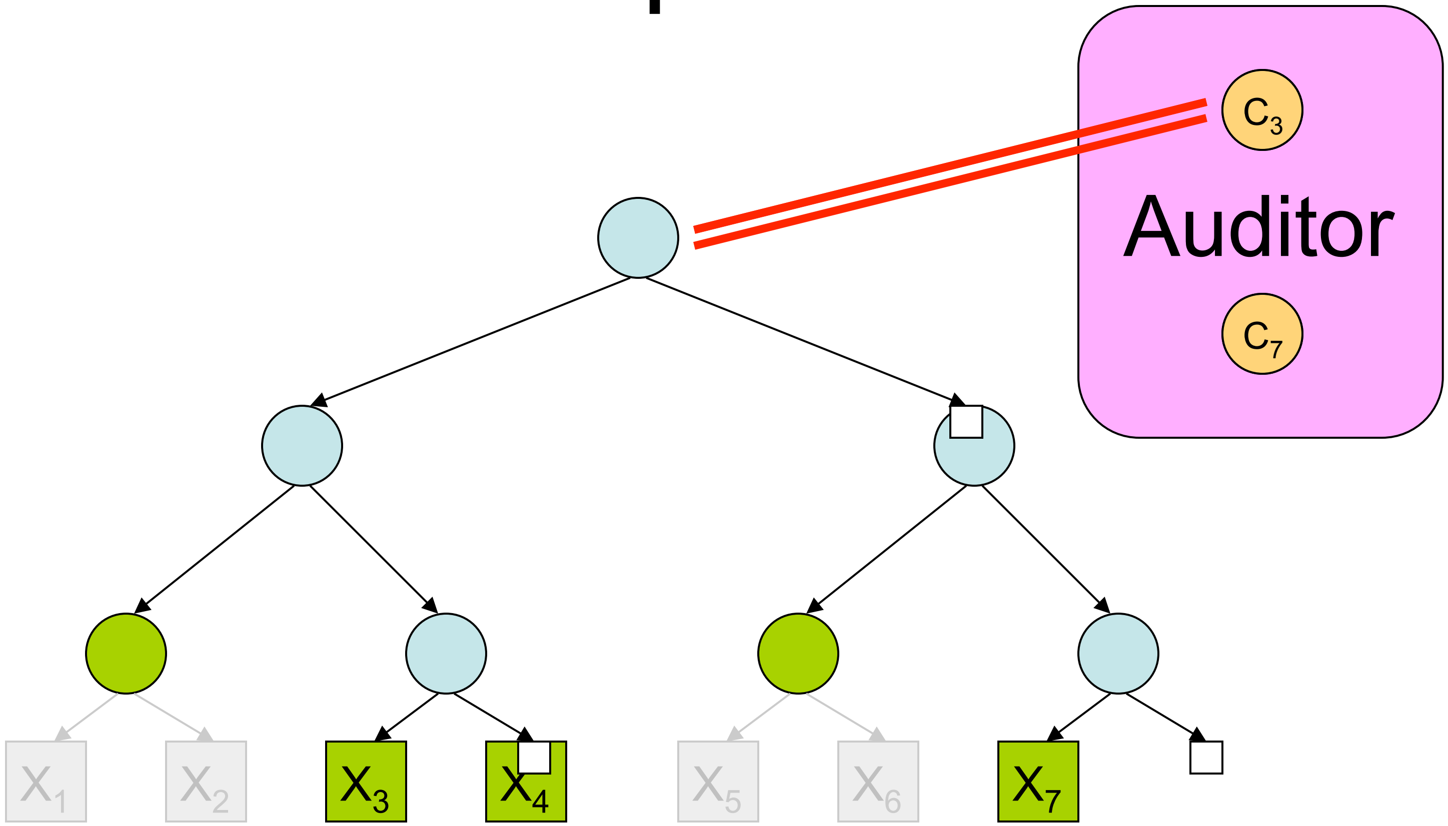
Incremental proof

$C_3 \equiv C_7$



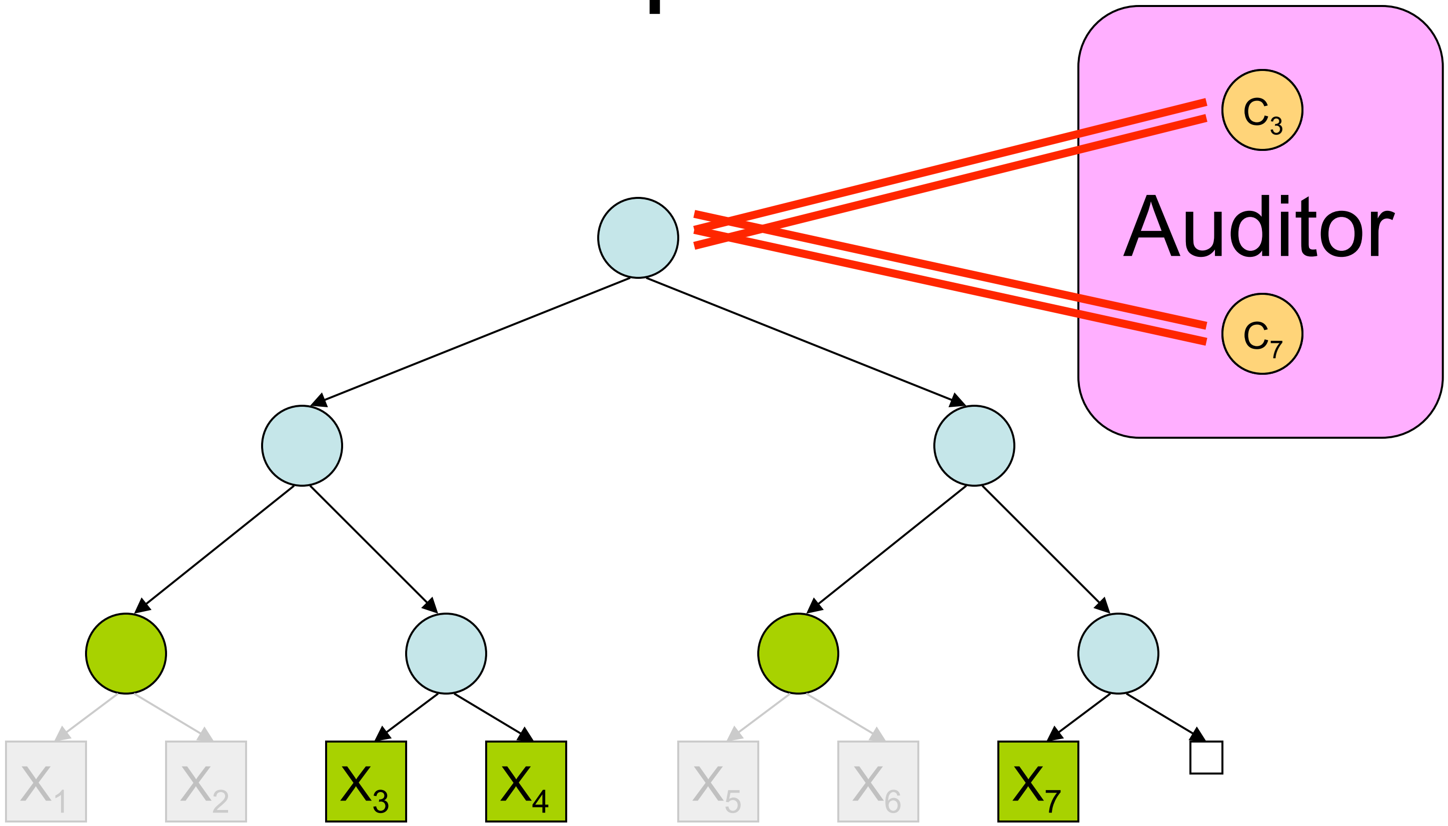
- P is consistent with C_7
- P is consistent with C_3
- Therefore C_7 and C_3 are consistent.

Incremental proof $C_3 \equiv C_7$



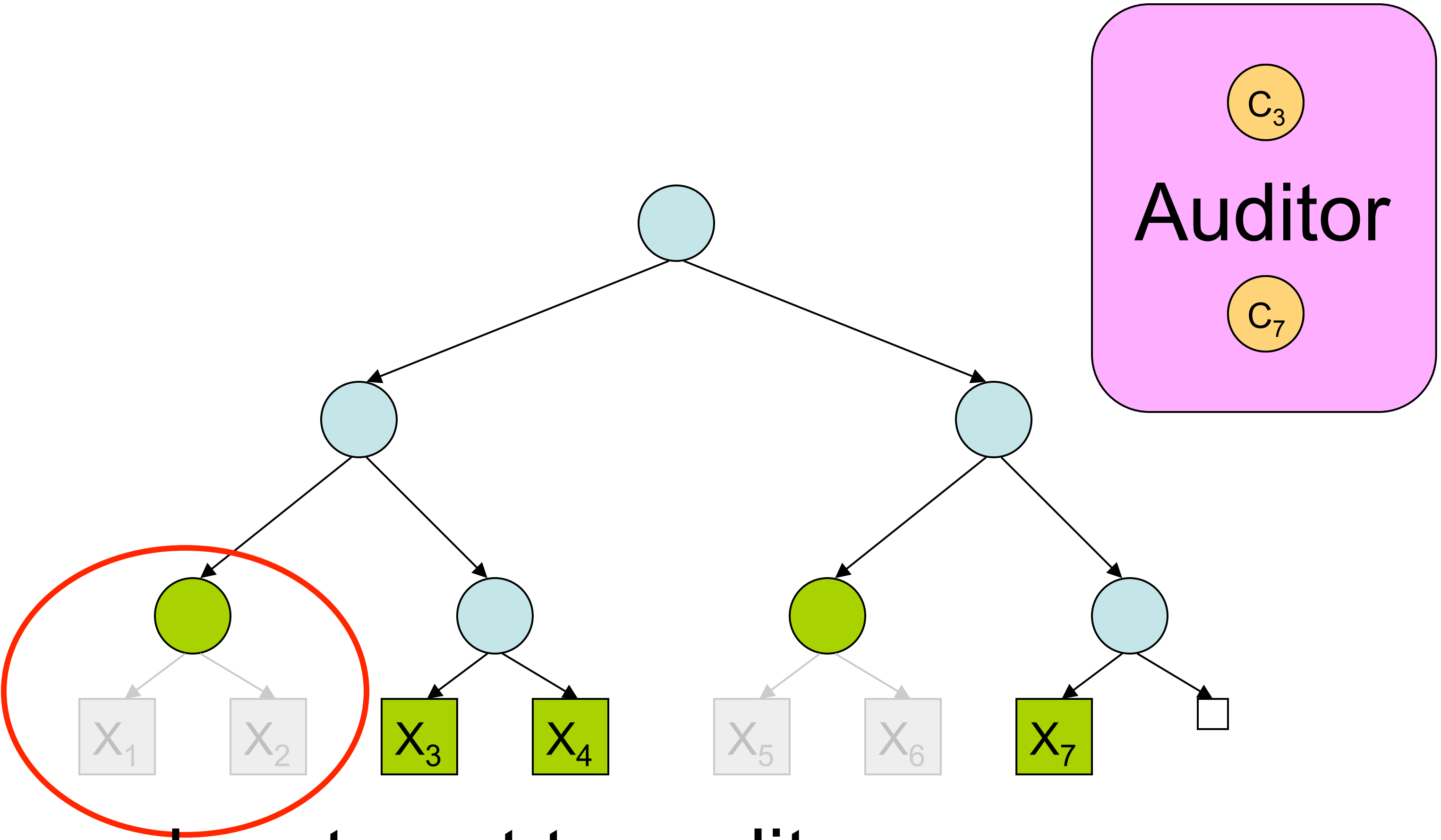
- P is consistent with C_7
- P is consistent with C_3
- Therefore C_7 and C_3 are consistent.

Incremental proof $c_3 \equiv c_7$



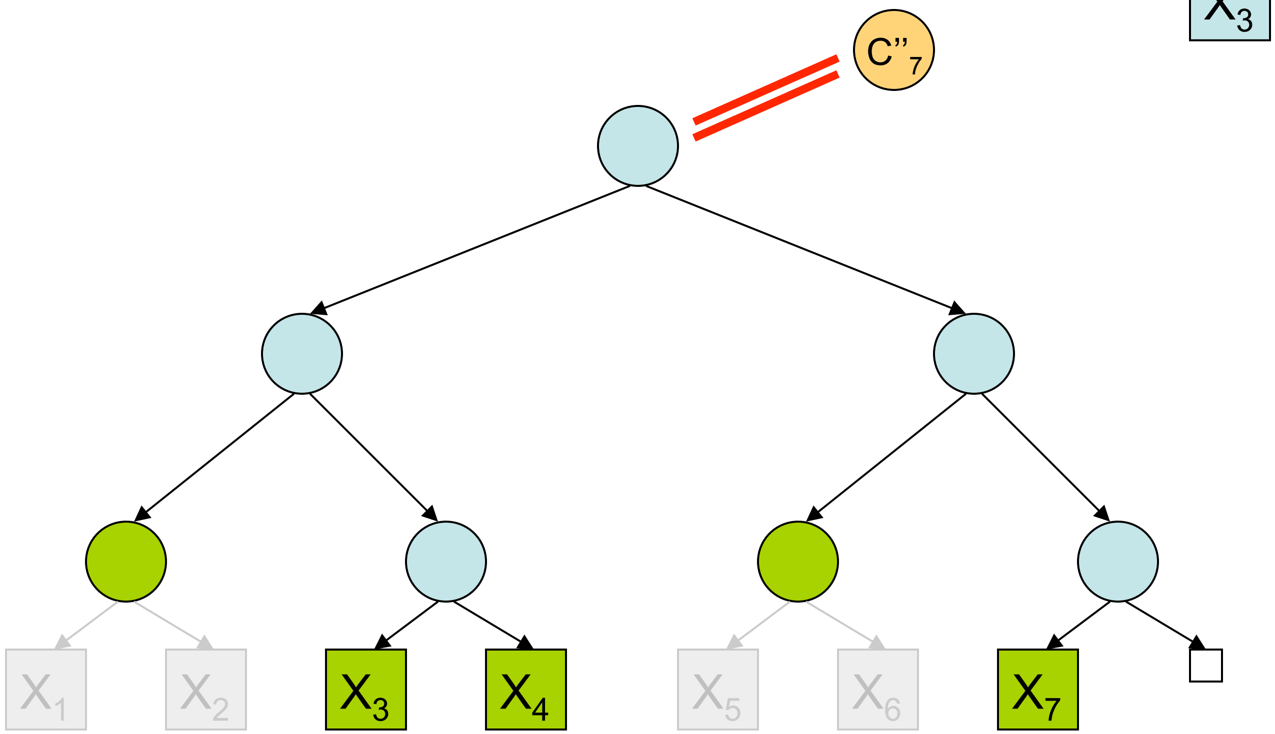
- P is consistent with c_7
- P is consistent with c_3
- Therefore c_7 and c_3 are consistent.

Pruned subtrees



- Although not sent to auditor
  - Fixed by hashes above them
  - C_3 , C_7 fix the same (unknown) events

Membership proof that $X_3 \in C''_7$



- Verify that C''_7 has the same contents as P
- Read out event X_3

Evaluating the history tree

- Big-O performance
- Syslog implementation

Big-O performance

	$c_j \equiv c_i$	$X_i \in c_j$	Insert
History tree	O(log *n*)	O(log *n*)	O(log *n*)
Hash chain (e.g., BitCoin)	O(*j*-*i*)	O(*j*-*i*)	O(1)
Skip-list history [Maniatis and Baker]	O(*j*-*i*) or O(*n*)	O(log *n*) or O(*n*)	O(1)

Syslog implementation

- Syslog
  - – Trace from Rice CS departmental servers
  - – 4M events, 11 hosts over 4 days, 5 attributes per event
    - Repeated 20 times to create 80M event trace

Syslog implementation

- Implementation
  - Hybrid C++ and Python
  - Single threaded
  - `mmap()`-based append-only write-once storage
  - 1024-bit DSA signatures and 160-bit SHA-1 hashes
- Test platform
  - 2.4 GHz Core 2 Duo (circa 2007) desktop machine
  - 4GB RAM

Performance

- Insert performance: 1,750 events/sec
  - 83.3% : Sign commitment
- Auditing performance
  - With locality (last 5M events)
    - 10,000-18,000 incremental proofs/sec
    - 8,600 membership proofs/sec
  - Without locality
    - 30 membership proofs/sec
  - < 4,000 byte self-contained proof size

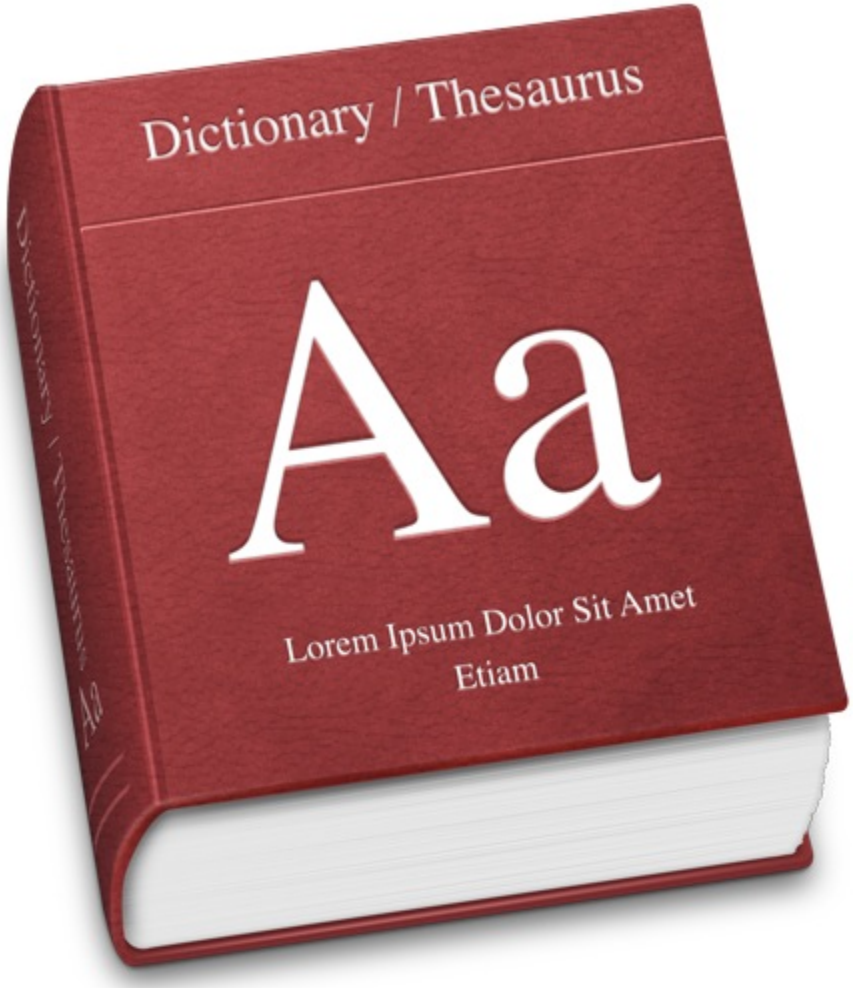
Tamper-evident logging

- New paradigm
  - Importance of frequent auditing
- History tree
  - Efficient auditing
  - Scalable
  - Offers other features

  - Proofs and more in the papers

Persistent authenticated dictionaries (PADs)

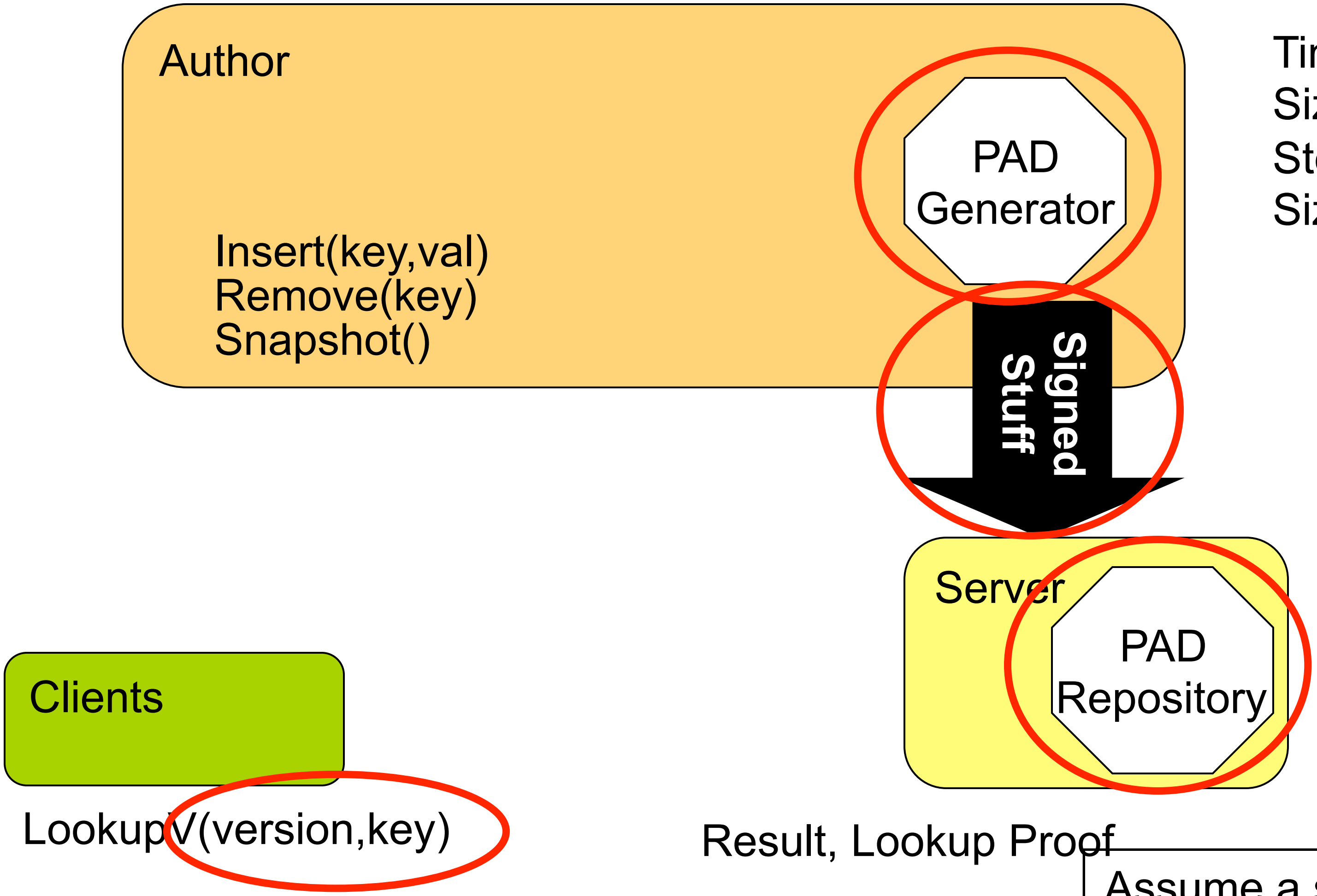
What is a PAD?


Dictionary / Thesaurus
Aa
Lorem Ipsum Dolor Sit Amet
Etiam
Dictionary / Thesaurus
Aa

What is a PAD?

- What is an authenticated dictionary?
  - Tamper-evident key/value data store
  - Invented for storing CRLs [Naor and Nissim 98]
- Security model
  - Created by trusted author
  - Stored on untrusted server
  - Accessed by clients
    - Responses authenticated by author’s signature
- **PAD adds the ability to access old versions**
  - [Anagnostopoulos et al 01]

PAD design



Author

PAD Generator

Insert(key,val)
Remove(key)
Snapshot()

Signed Stuff

Server

PAD Repository

Clients

LookupV(version,key)

Result, Lookup Proof

Time per update
Size of update
Storage per update
Size of a proof

Assume a single author

Assume snapshot after every update

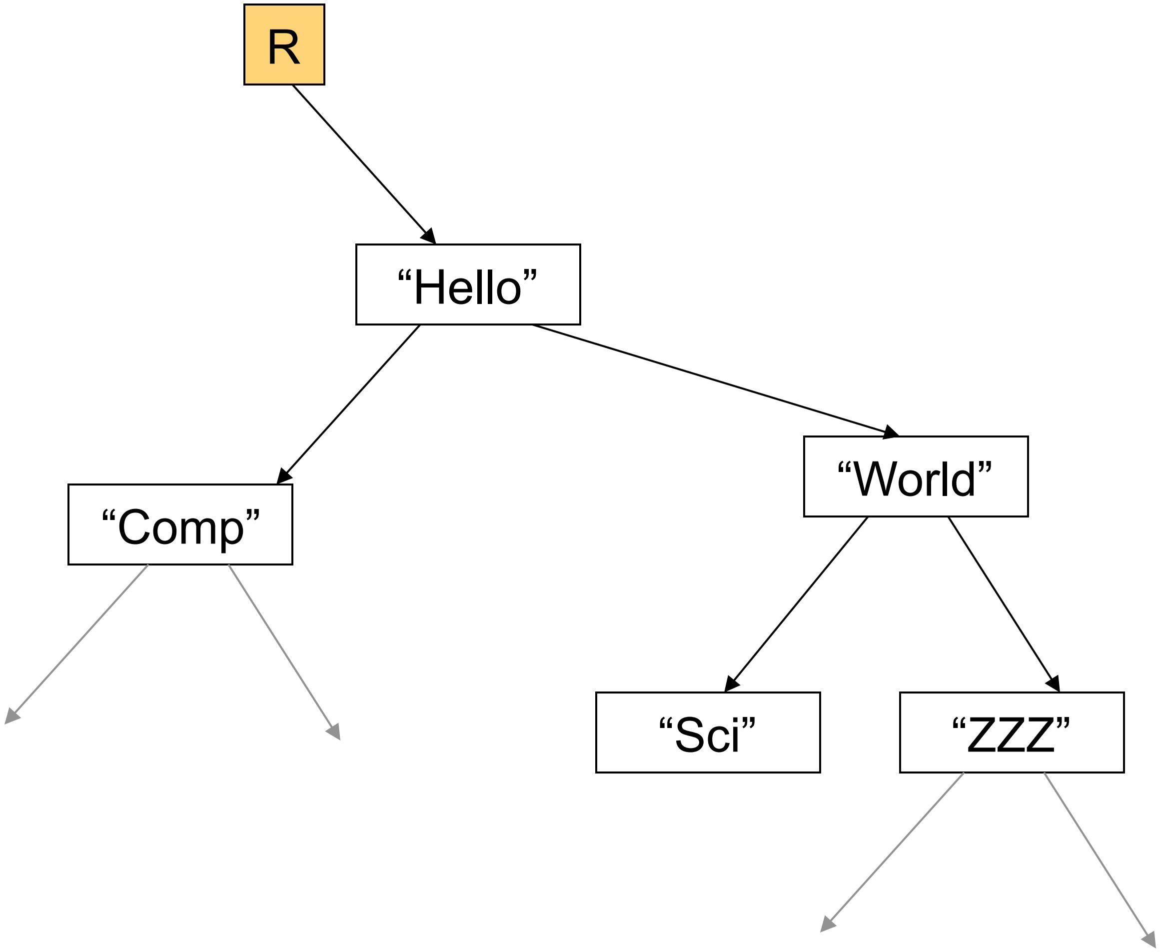
Applications of PADs

- Outsource storage and publishing
  - CRL
  - Cloud computing
  - Remote backups
  - Subversion repository
  - Stock ticker
  - Software updates
  - Smart cards
- Want to look up historical data

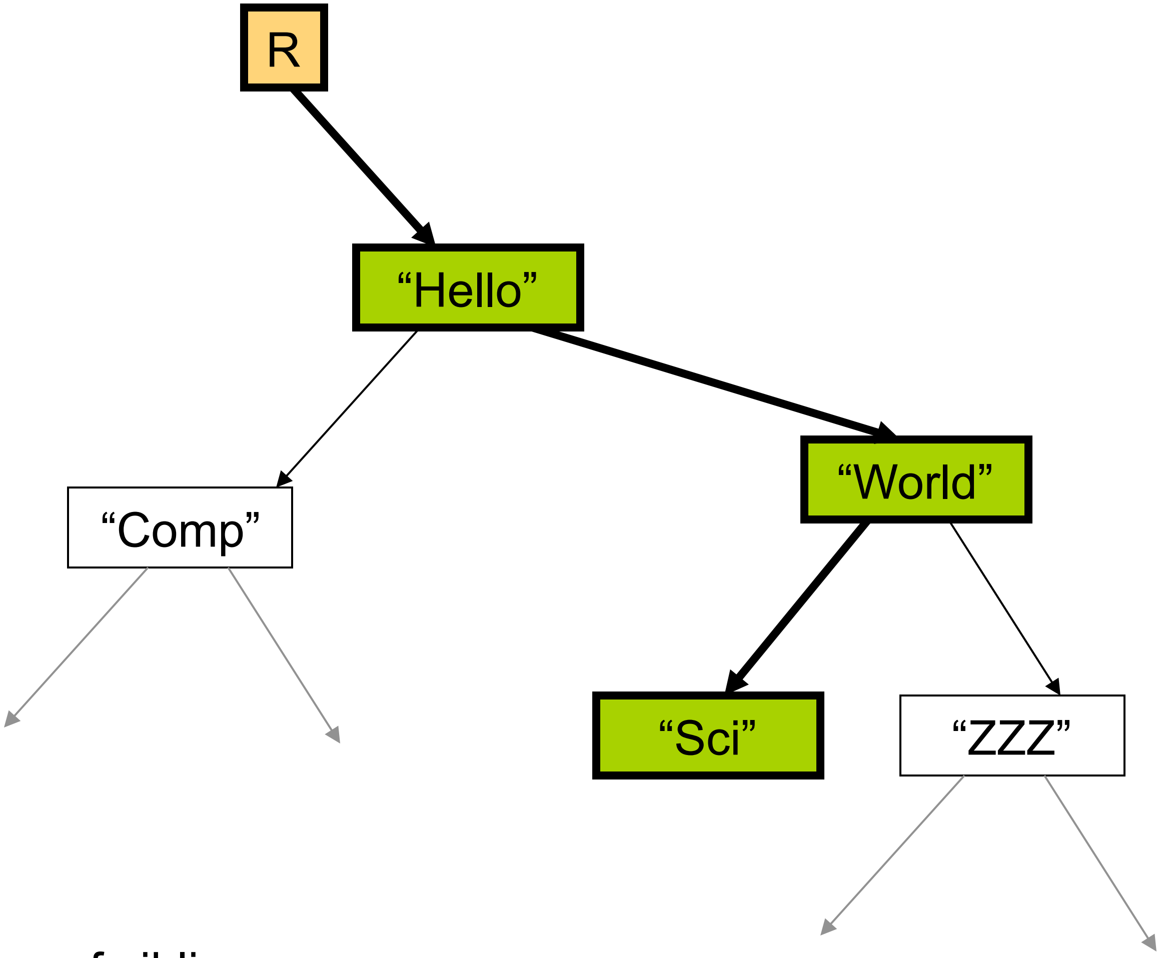
PAD Designs

- Tree-based PADs [Anagnostopoulos et al., Crosby and Wallach]
  - O(log *n*) storage per update
  - O(log *n*) lookup proof size
- Tuple PADS [Crosby and Wallach]
  - O(1) storage per update
  - O(1) proof size

Tree-based authenticated dictionary

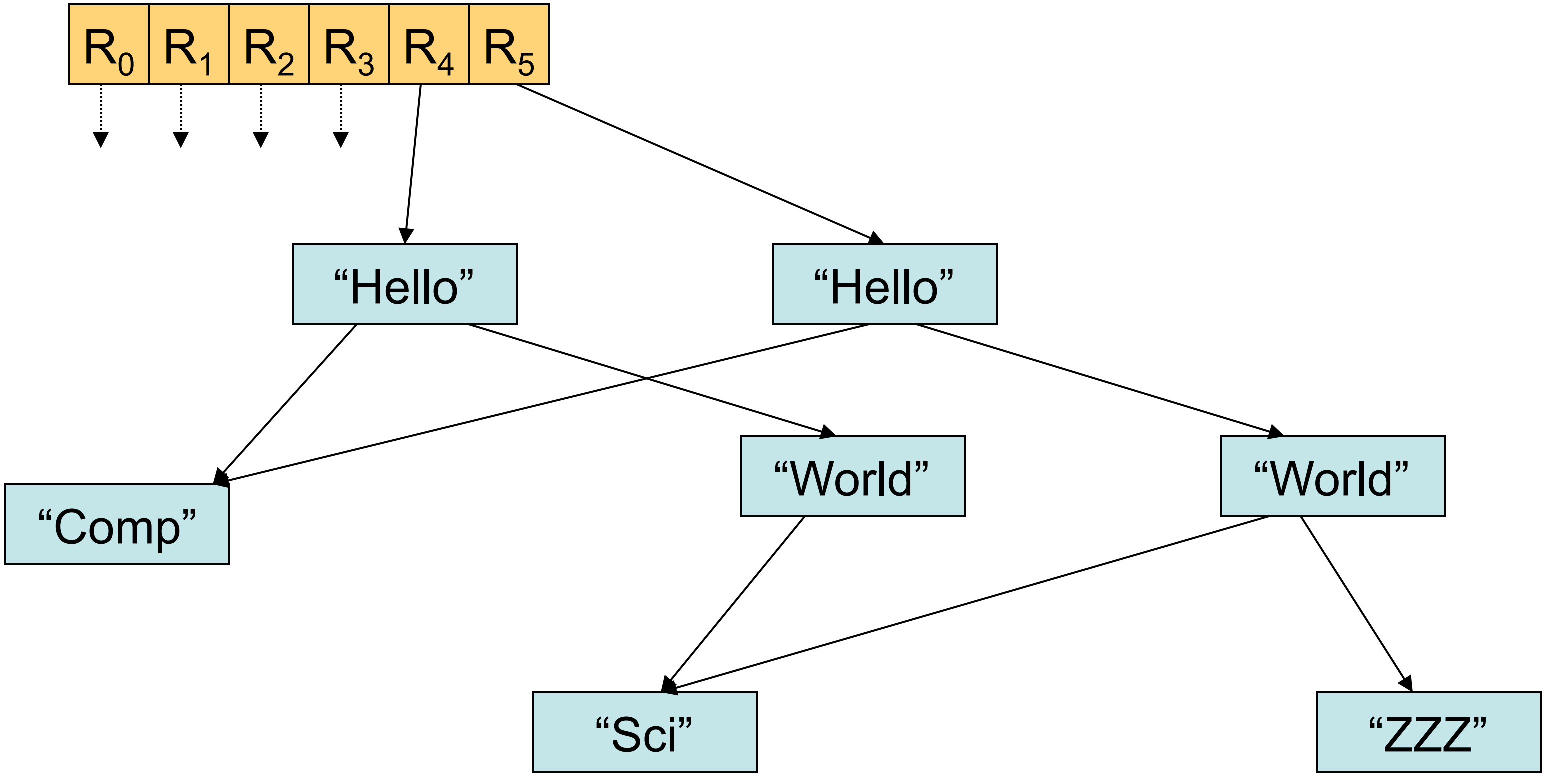


Proofs in a tree-based authenticated dictionary



Proof: Hashes of sibling nodes on path to lookup key

Path copying

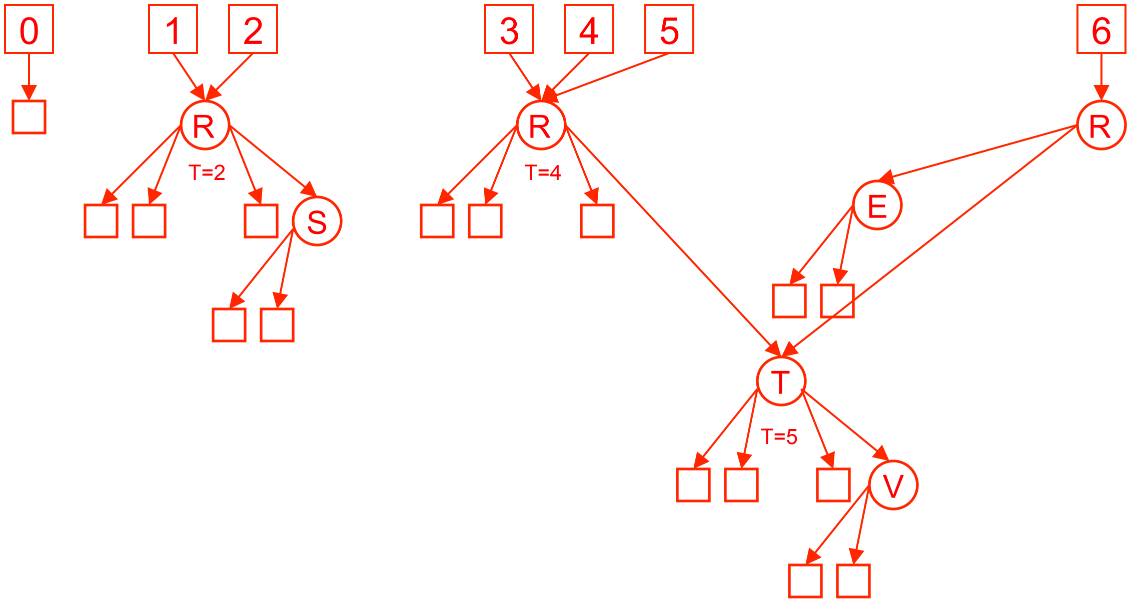


Storage: O(log n) per update

Building a PAD

- Other ways to make trees persistent
  - Versioned nodes [Sarnak and Tarjan 86]
    - O(1) amortized storage per update.
  - Our contribution:
    - Combining versioned nodes with authenticated dictionaries
    - Reduce memory consumption on the server

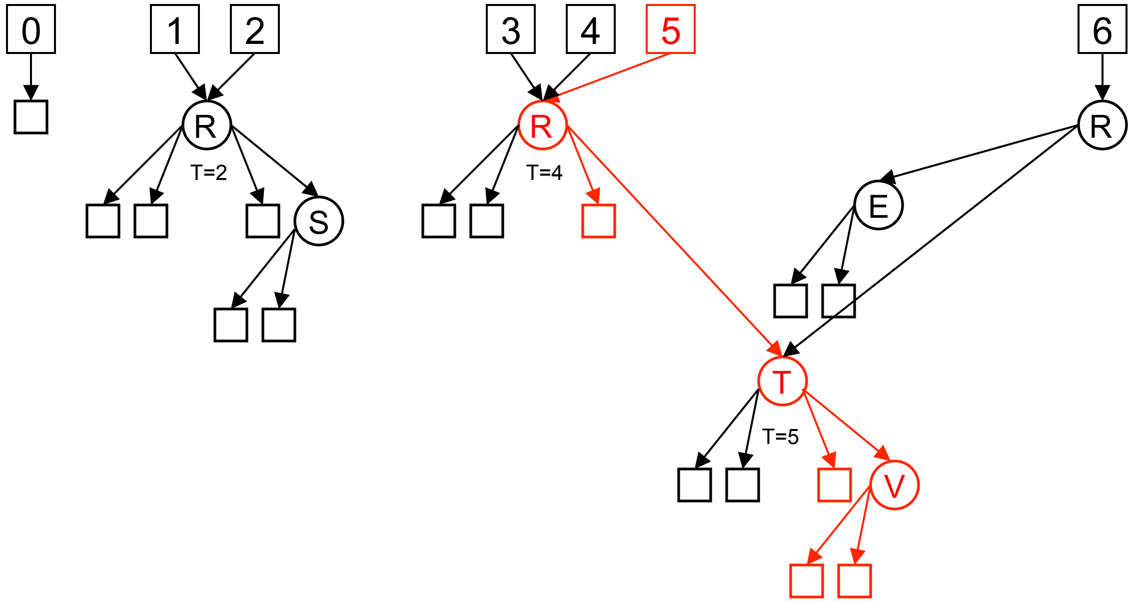
Sarnak-Tarjan tree



Add R
Add S
Del S
Add T
Add V
Add E

Note: 7 snapshots represented with 7 nodes.

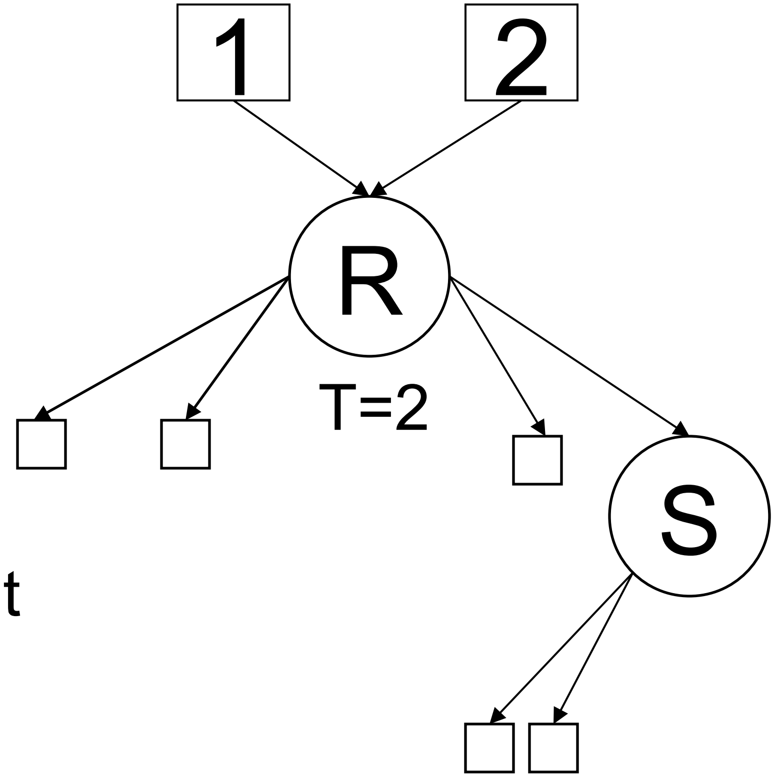
Accessing snapshot 5



Add R
Add S
Del S
Add T
Add V
Add E

Sarnak-Tarjan node

- Each node has two sets of children pointers and a "time"
- Hash is not constant
  - Can be recomputed from tree at any "time"
- Storing vs. recomputing
  - Same semantics, different performance



Comparing caching strategies

	Storage (Server)	Lookup Proof Generation (Server)
Cache nowhere	O(1)	O(n)
Cache everywhere	O(log n)	O((log n) *(log v))
Cache median layer	O(2)	O(√n * (log v))

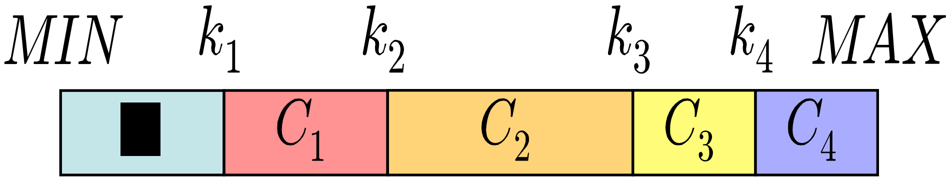
- Logarithmic
  - Update time
  - Lookup size
  - Verification time
- Constant
  - Update size

Tuple PADs

- Our new PAD design
  - **Constant lookup proof size**
  - **Constant storage per update**

Tuple PADs

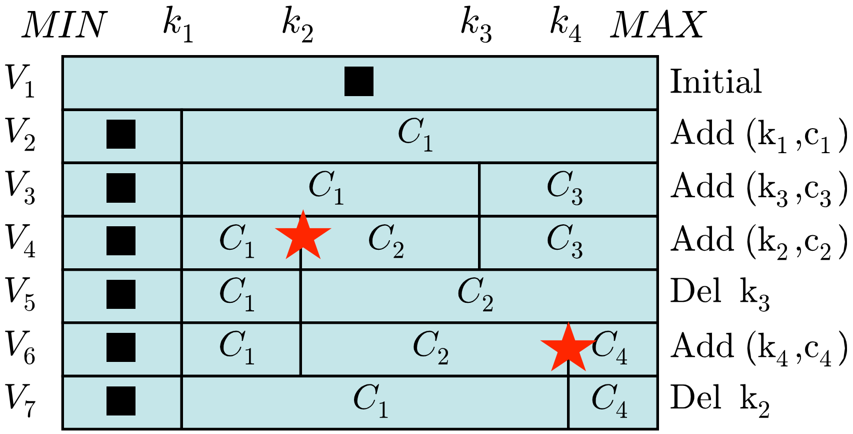
- Dictionary contents:
  - { $k_1 = c_1$, $k_2 = c_2$, $k_3 = c_3$, $k_4 = c_4$ }
- Divide key-space into intervals
- Tuples:
  - ([MIN,k_1),■)
  - ([k_1,k_2),c_1)
  - ([k_2,k_3),c_2)
  - ([k_3,k_4),c_3)
  - ([k_4,MAX),c_4)



"Key k_1 has value c_1, and there is no key in the dictionary between k_1 and k_2"

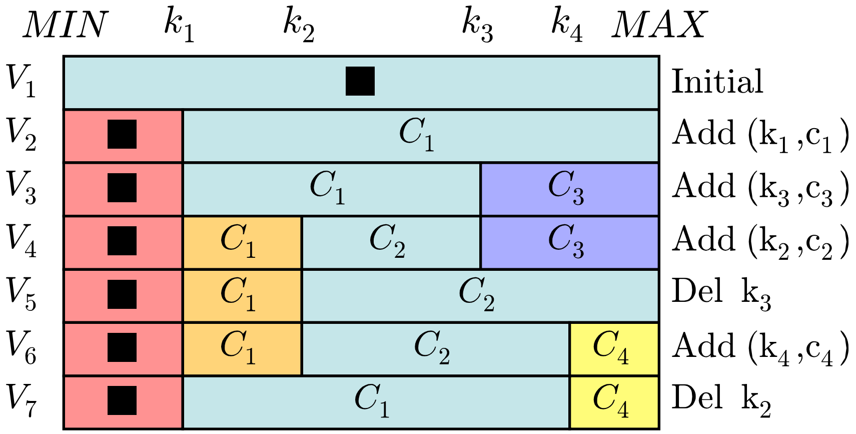
Making it persistent

- $(v_1,[k_1,k_2),c_1)$
  - “In snapshot v_1, key k_1 has value c_1, and there is no key in the dictionary between k_1 and k_2”



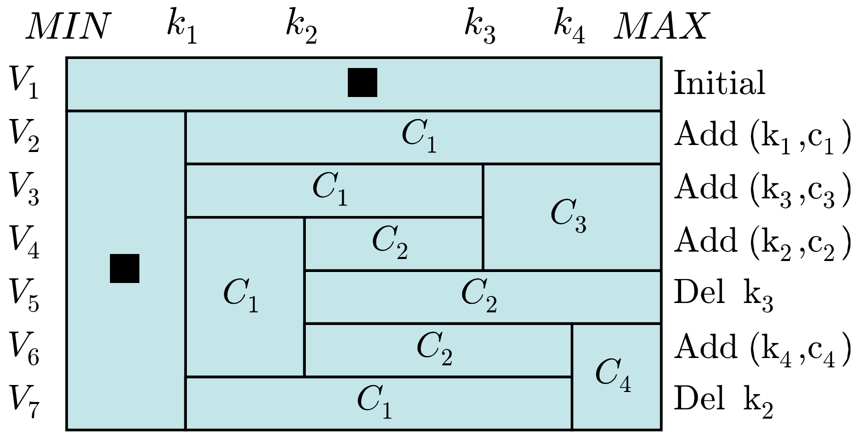
Observation

- Most tuples stay same between snapshots
- Every update
  - – Creates ≤ 2 tuples not in prior snapshot

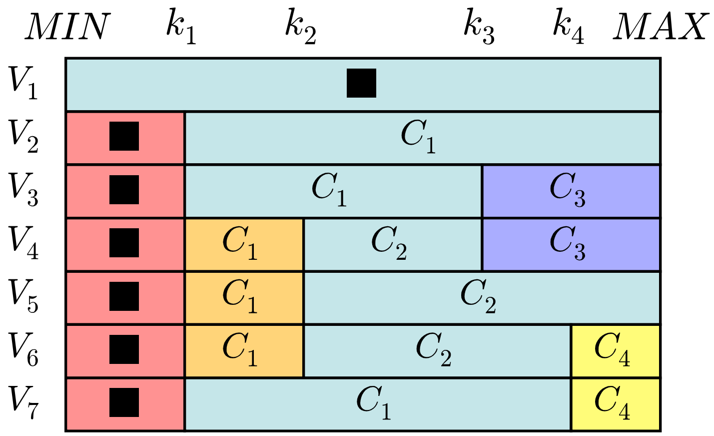


Tuple superseding

- Indicate a version range in each tuple
  - ([v_1,v_2+1], [k_1,k_2),c_1)
    - Which replaces ([v_1,v_2], [k_1,k_2),c_1)
    - At most 2 new tuples. Rest are replaced
  - Constant
    - Storage on server
  - Still have the same
    - Update time
    - Update size



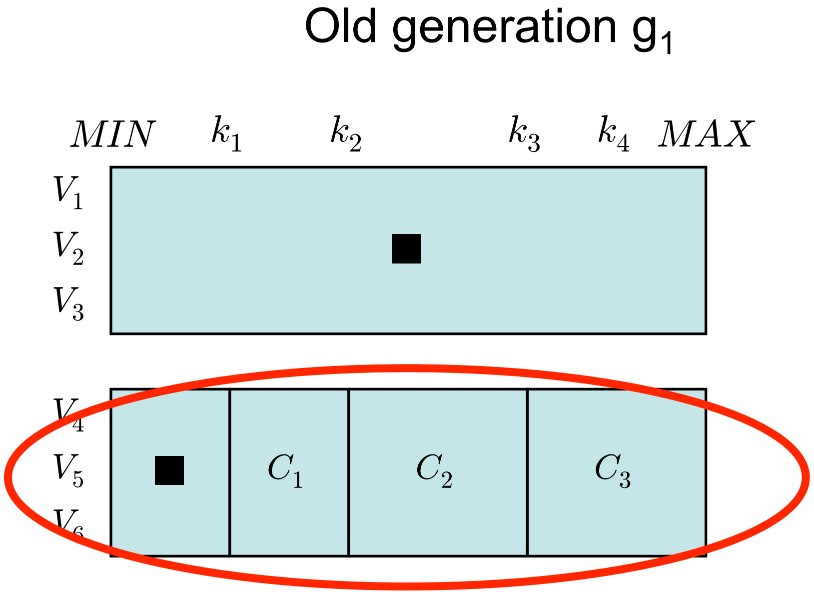
Insight: Speculation

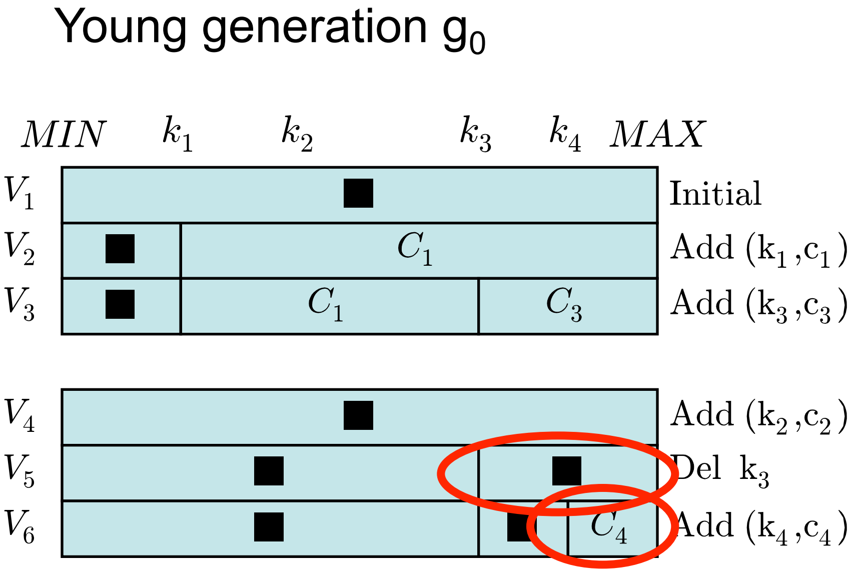


- Split PAD
  - – Speculative tuples
    - Older generation
    - Signed in every epoch
  - – Young generation
    - Correct mis-speculations
    - Signed every snapshot
    - Kept small, migrate keys into older generation
- O(G $n^{1/G}$) signatures per update
  - – Combines with lightweight signatures

Speculation: Updating the PAD

- (g_0,[v_1,v_2],[k_1,k_2),c_1)
  - "In generation g_0 and snapshots v_1 through v_2 key k_1 has value c_1, and there is no key in the dictionary between k_1 and k_2"





Reducing update costs

- Currently $O(G\ n^{1/G})$ update size
  - Requiring $O(G\ n^{1/G})$ work
- RSA accumulators [Benaloh and de Mare 93]
  - O(1)
    - Work on author
    - Update size
    - Lookup proof size
  - $O((G+1)\ n^{1/G}\ (\log n))$
    - Computation on server
    - Large constant factors

Comparing techniques

		Tree-based			Tuple-based		
		Path Copying	Cache Everywhere	Cache Median	Speculating+ Superseding	Superseding	Accumulators + Speculating
Updates	Time (Author)	O(log n)			$O(G * n^{1/G})$	O(n)	O(1)
	Time (Server)						$O(G * \log(n) * n^{1/G})$
	Size						O(1)
Storage	(per update)	O(log n)		O(1)	O(G)	O(1)	
Lookup	Time (Server)	O(log n)	O(log n * log v)	O(√n)	O(G * log n)	O(log n)	
	Size	O(log n)			O(G)	O(1)	

What about the real world?

		Tree-based	Tuple-based		
				perseding	Accumulators + Speculating
Updates	Time (Author)			O(n)	O(1)
	Time (Server)				$O(G * \log(n) * n^{1/G})$
	Size				O(1)
Storage	(per update)			O(1)	
Lookup	Time (Server)			O(log n)	
	Size	O(log n)	O(G)	O(1)	

Benchmarking PADs

Comprehensive implementation

- 21 algorithms
- Including all earlier designs
  - Path copy skiplists and path copy red-black trees [Anagnostopoulos et al.]

- Analysis also applies to non-persistent authenticated dictionaries

Algorithms

- Tree PADs – 12 designs
  - (4) Path copying, 3 caching strategies
  - (3) Red-black, Treap, and Skiplist
- Tuple PADs – 6 algorithms
  - (2) With and without speculation
  - (3) No-superseding, superseding, lightweight signatures
- Accumulator PADs – 3 algorithms

Implementation

- Hybrid of Python and C++
  - GMP for bignum arithmatic
  - OpenSSL for signatures
- Core 2 Duo CPU at 2.4 GHz
  - 4GB of RAM
  - 64-bit mode

(Not bad for circa 2007 hardware!)

Benchmark

- ‘Growing benchmark’
  - Insert 10,000 keys with a snapshot after every insert
- Play a trace of price changes of luxury goods
  - 27 snapshots
  - 14000 keys
  - 39000 updates

Tree PADs

- Comparing algorithms
  - Red-black
    - Smallest proofs, least RAM, highest performance
  - Skiplists do the worst
- Comparing repositories
  - Path copying
  - Sarnak-Tarjan nodes cache everywhere
    - Same performance
    - 40% of the RAM

Cache median vs Cache everywhere

- 100,000 keys

	Update Size	Update Rate	Lookup Size	Lookup Rate	Memory usage
Cache median	.15kb	730/sec	1.5kb	196/sec	205MB
Cache everywhere	.15kb	730/sec	1.5kb	7423/sec	358MB

The costs of an algorithm

- Care about the monetary costs
- Use prices from cloud computing providers
  - In 2007, 200kb was worth 1sec of CPU time
    - Worth about $ .000030 = 3000μ¢

Monetary analysis

- Evaluate
  - Absolute costs per operation
    - CPU time and bandwidth
  - Relative contribution of
    - CPU
    - Bandwidth

Tree PAD caching strategies

- 37x slower, but only costs 2x as much
  - Sending a lookup reply
    - 1.5kb, costing **18μ¢**
  - Generating a lookup reply
    - Cache median: 5ms, costing **16μ¢**
    - Cache everywhere .13ms : **.4μ¢**

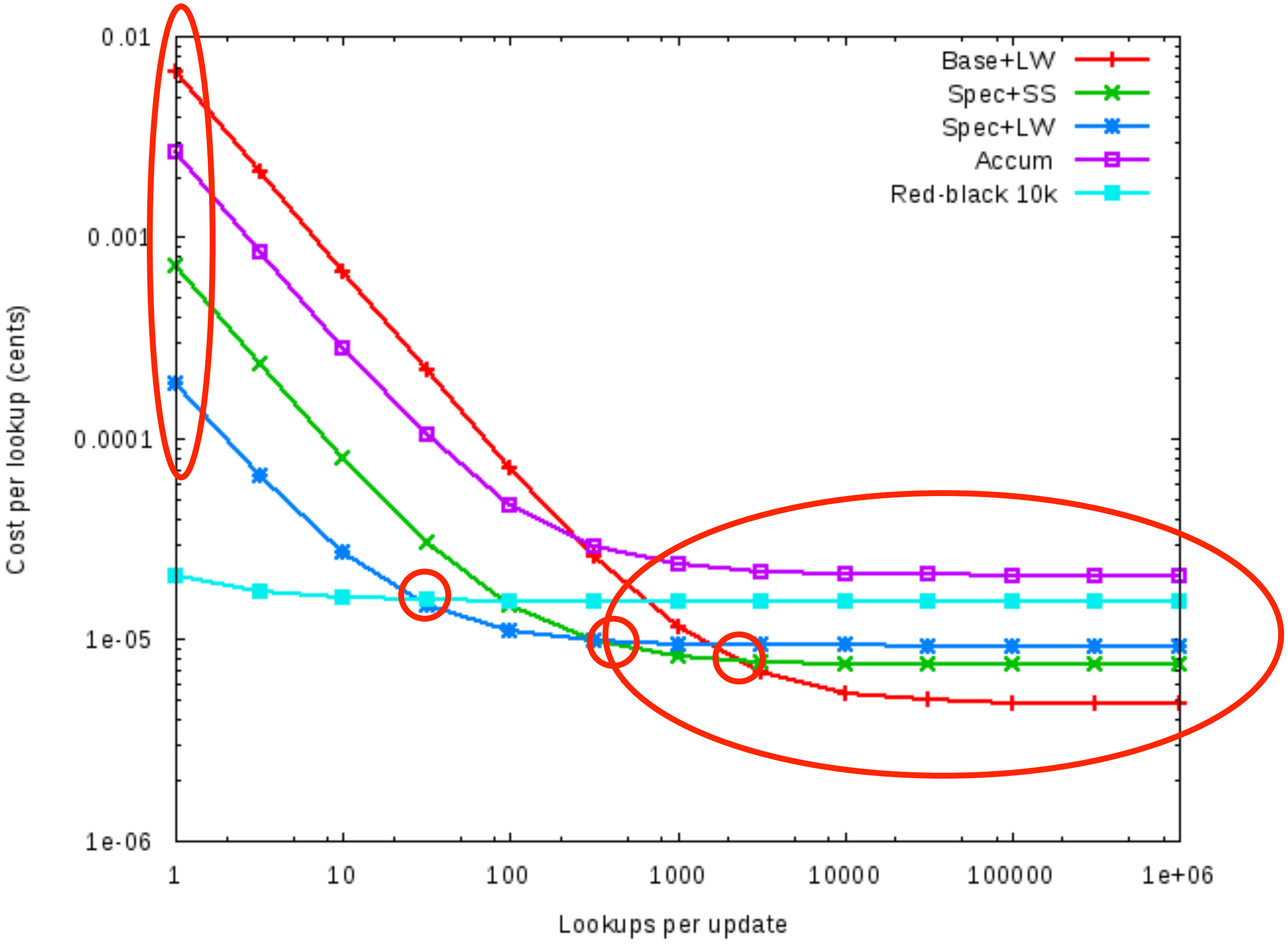
	Lookup size	Lookup rate	Cost per lookup	Memory usage
Cache median	1.5kb	196/sec	34 μ¢	205MB
Cache everywhere	1.5kb	7423/sec	18 μ¢	358MB

Evaluating the monetary costs of updates and lookups

- Tuple PADs
  - Extremely cheap lookups
  - Expensive updates
- Tree PADs
  - Cheap lookups
  - Cheap updates

“What is the cost per lookup if there are *k* lookups for each update for different values of *k*.”

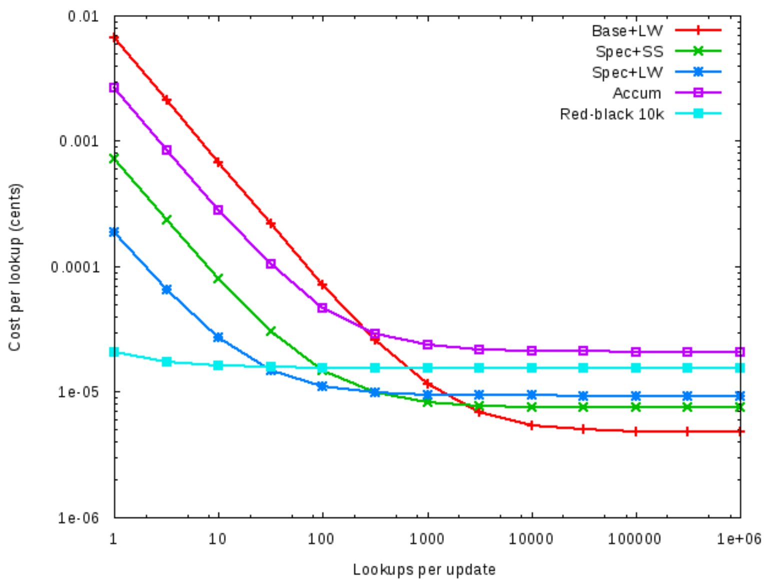
Costs per lookup on growing benchmark



These results

- Could not be presented without looking at costs of bandwidth and CPU time
- Constant factors matter

- Accumulators
  - Lookup proof >1kb
    - Just as big as red-black
  - Expensive updates



PAD designs

- Presented
  - New PAD designs
    - Improved tree PAD designs
    - New tuple PAD designs
      - Constant storage and constant sized lookup proofs
  - Comprehensive evaluation of PAD designs
    - Monetary analysis
- Focused on efficiency and the real-world

Conclusion

- Presented two tamper evident algorithms
  - New PAD designs
    - Comprehensive evaluation
    - Monetary analysis
  - Tamper-evident history
    - New extensions for fast digital signatures
- Focused on efficiency in the real-world
- Code and technical reports

http://tamperevident.cs.rice.edu