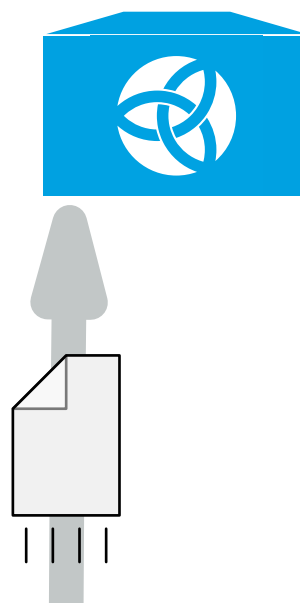


Cache Garbage Collection

in Knot Resolver

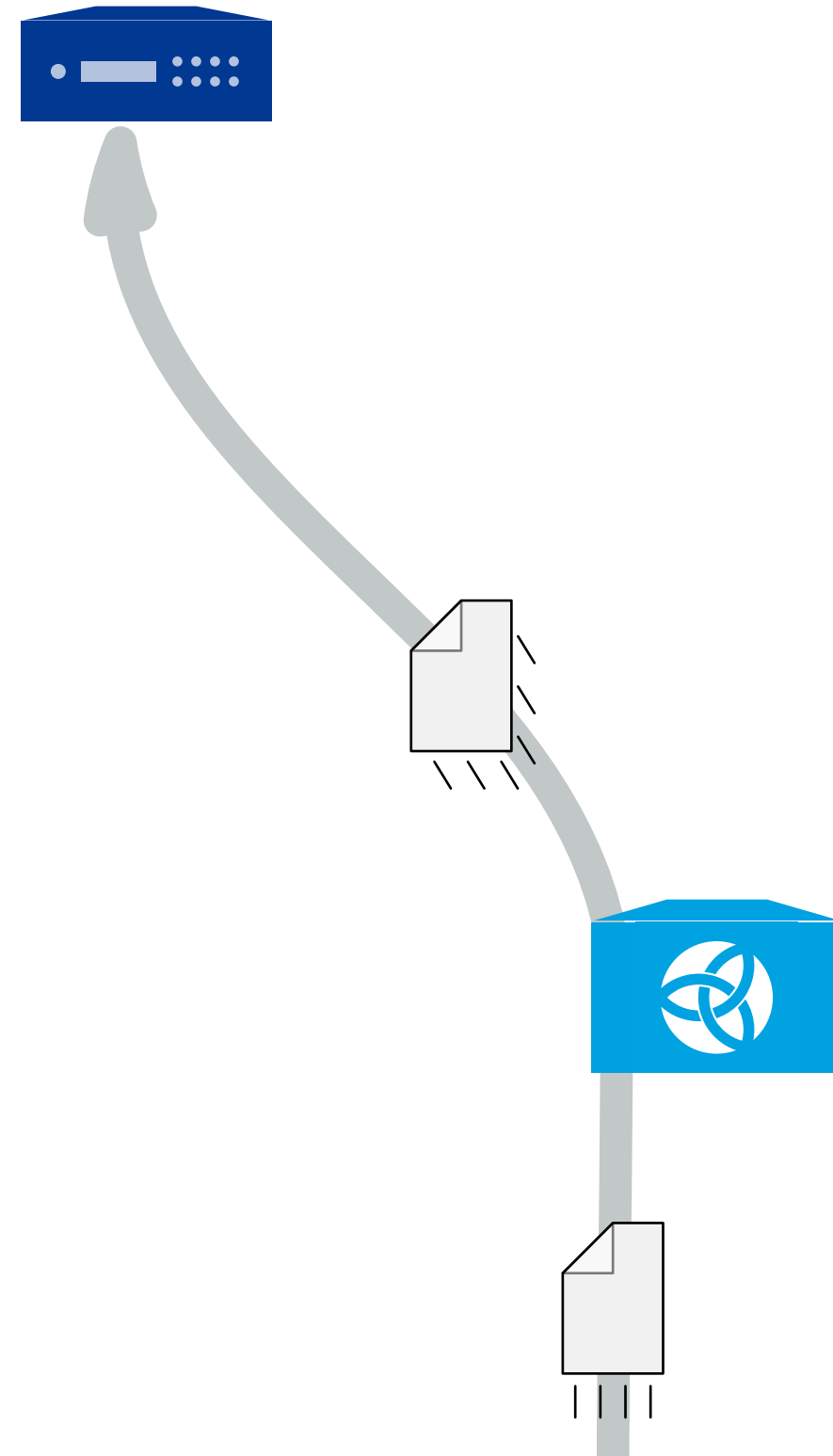
Overview

root nameserver

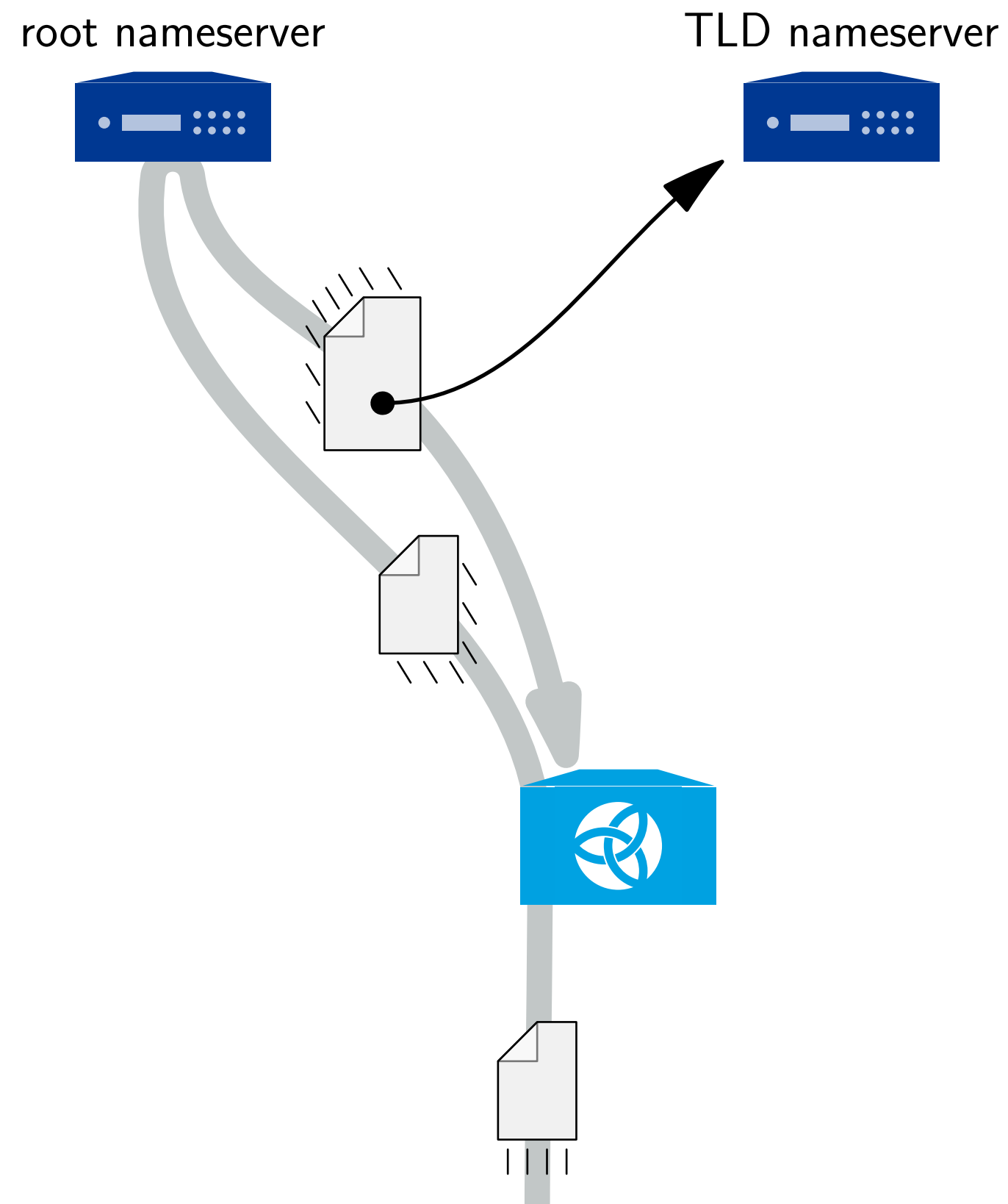


Overview

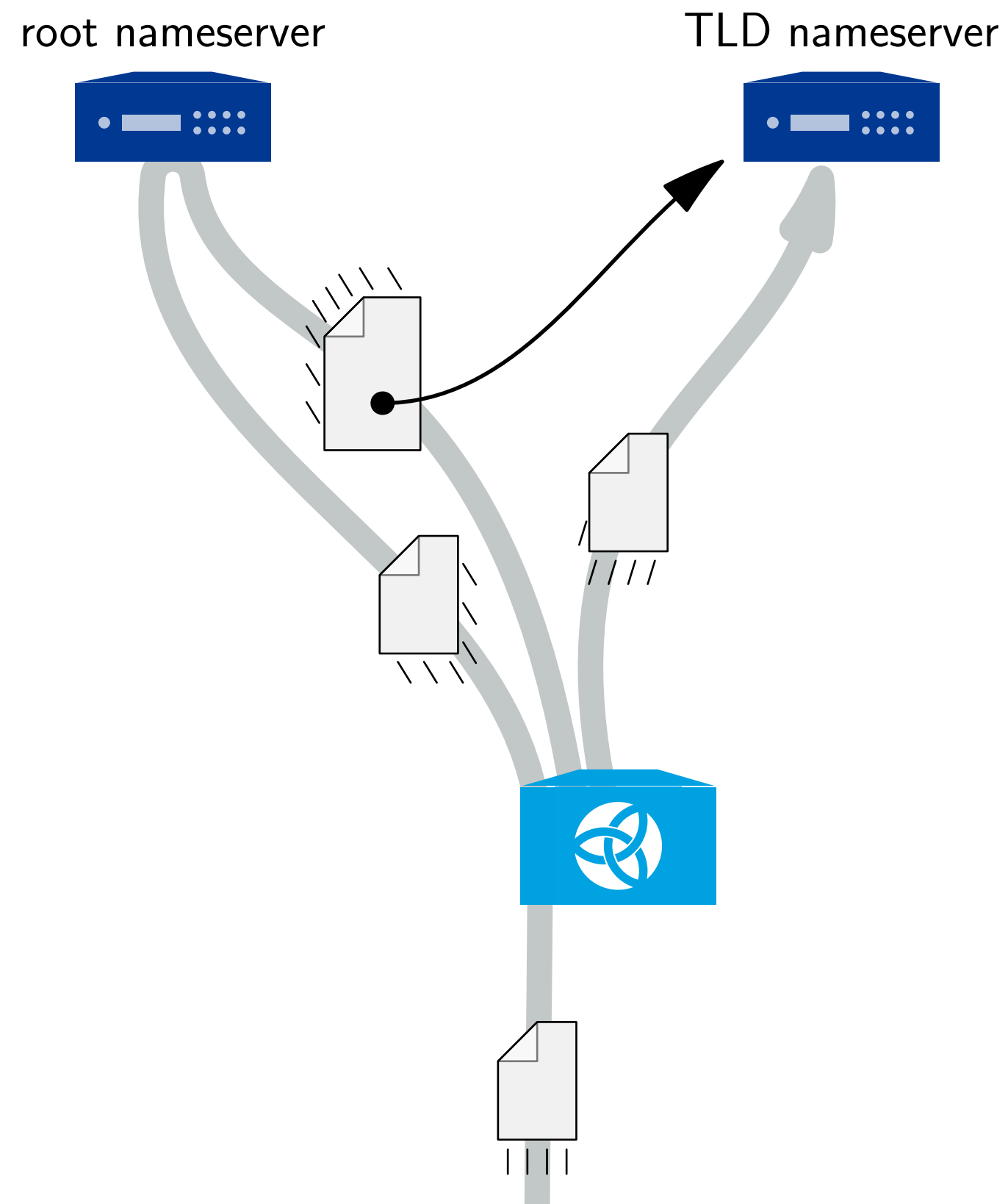
root nameserver



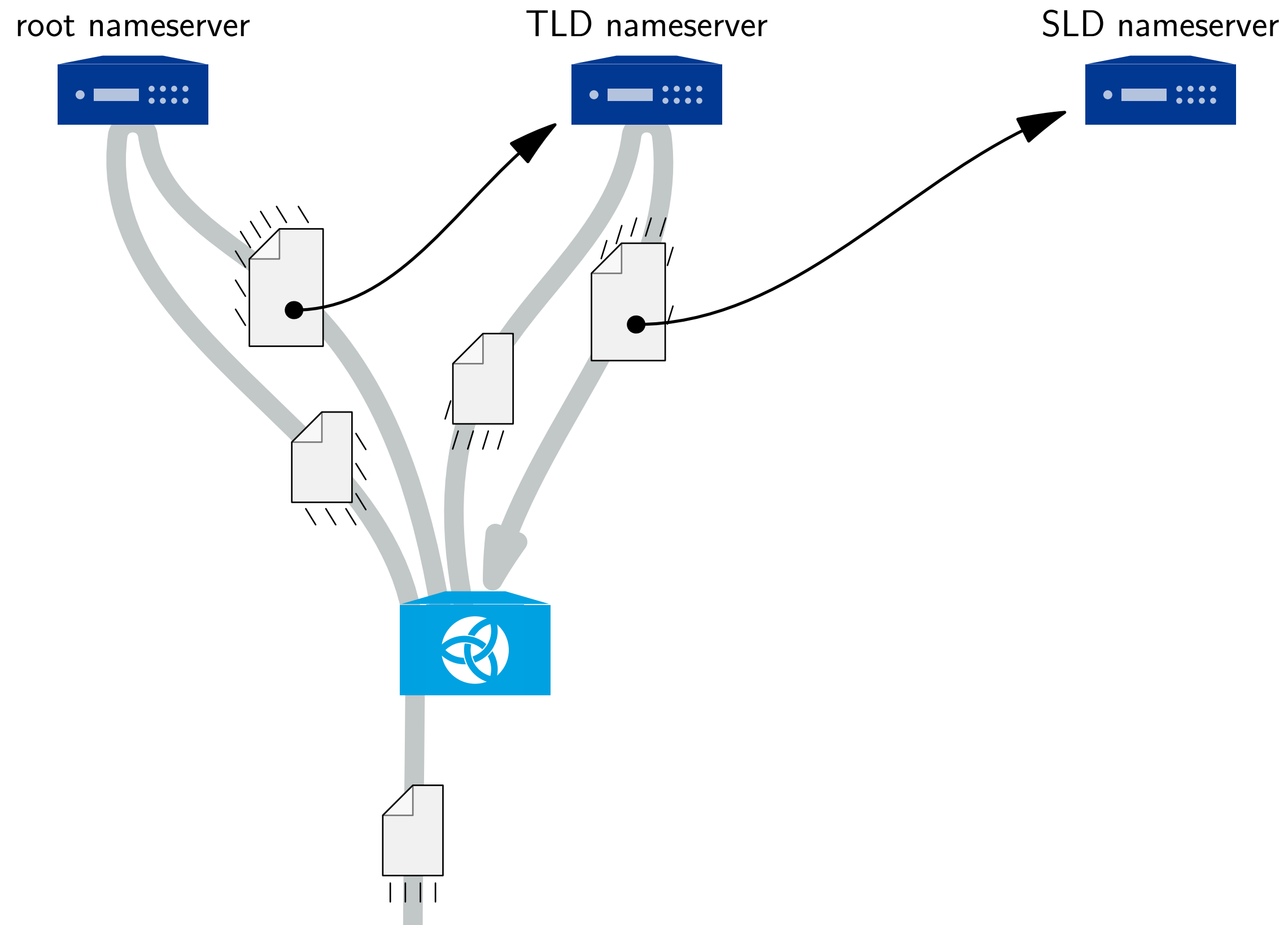
Overview



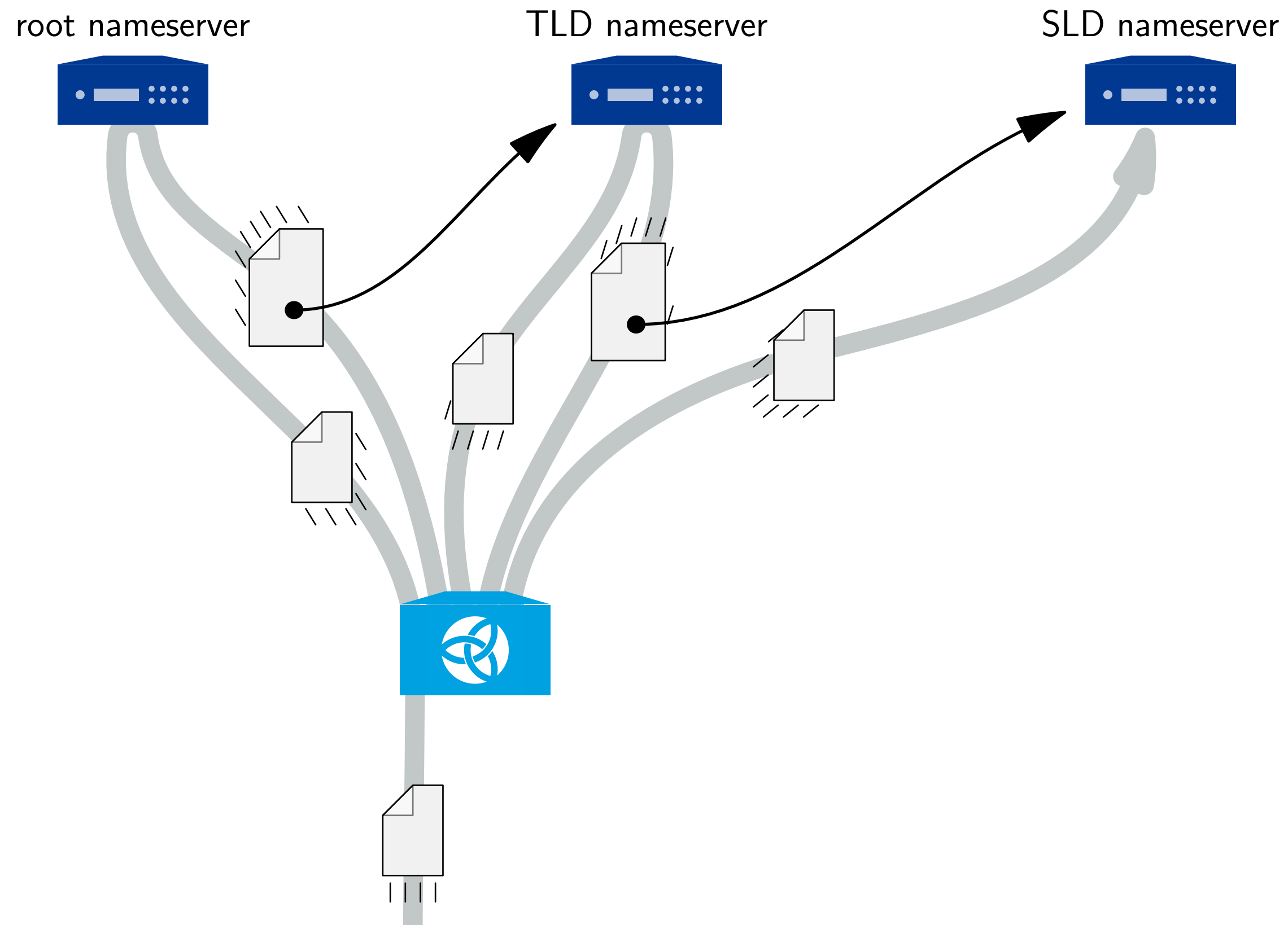
Overview



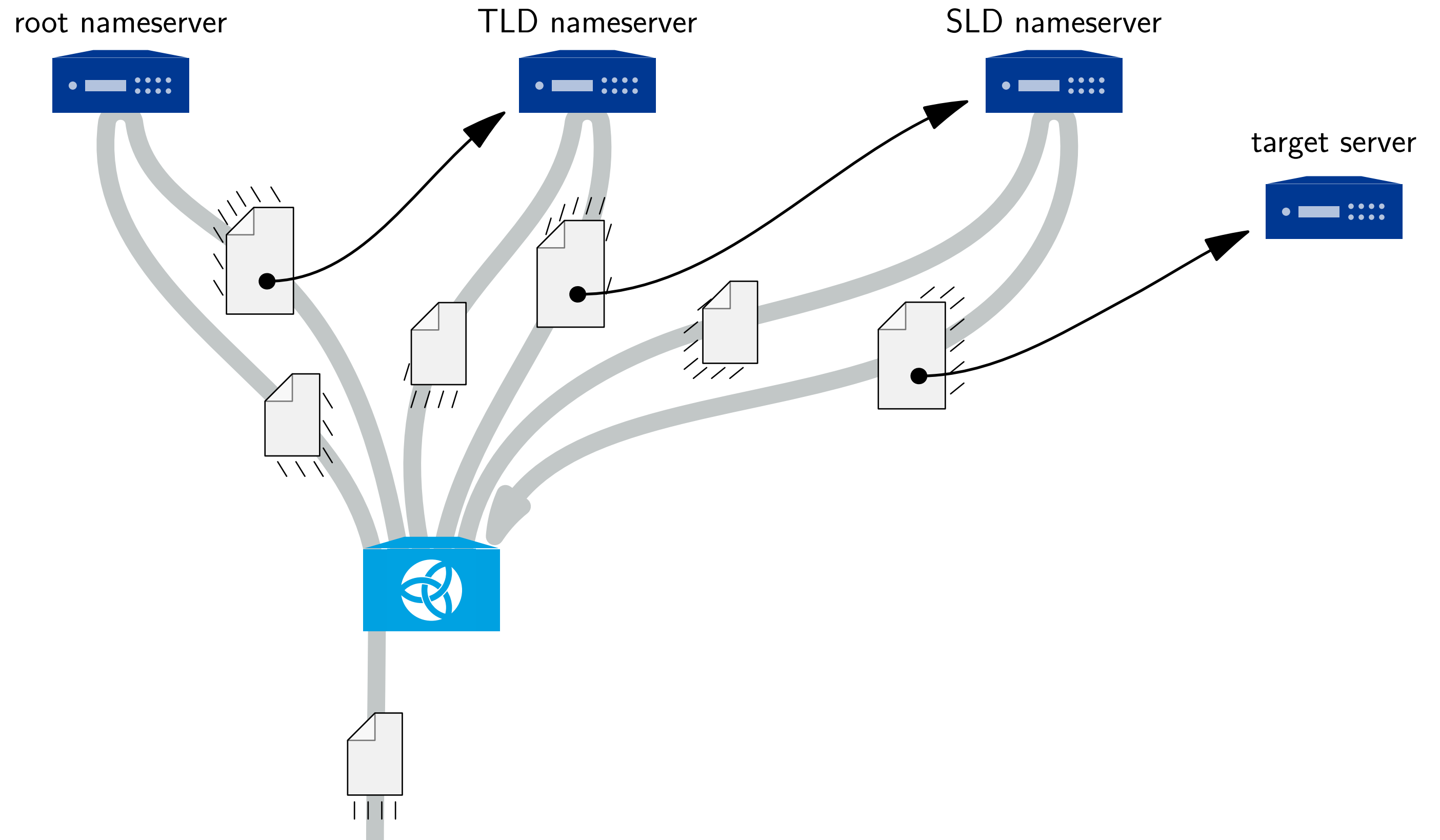
Overview



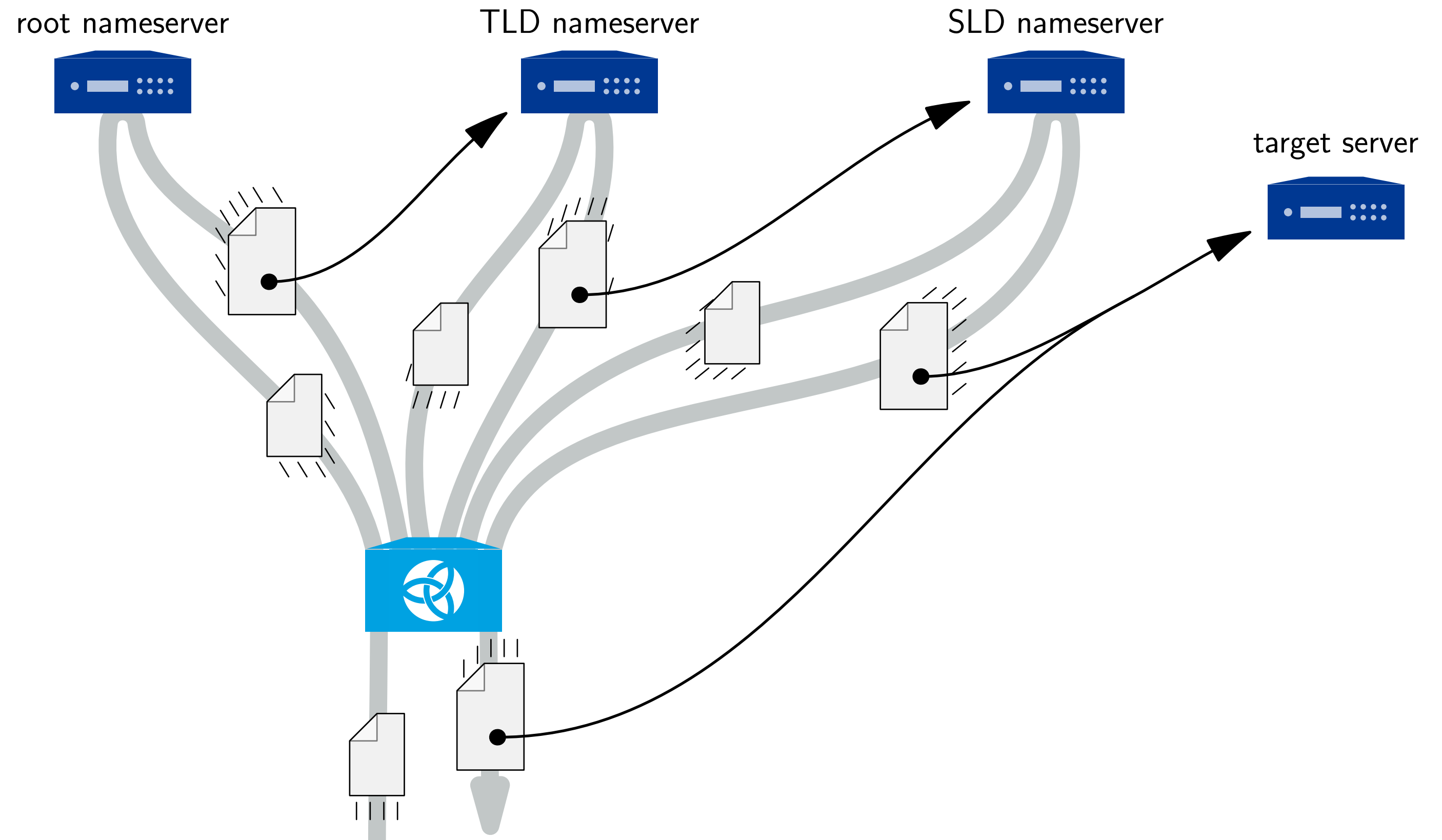
Overview



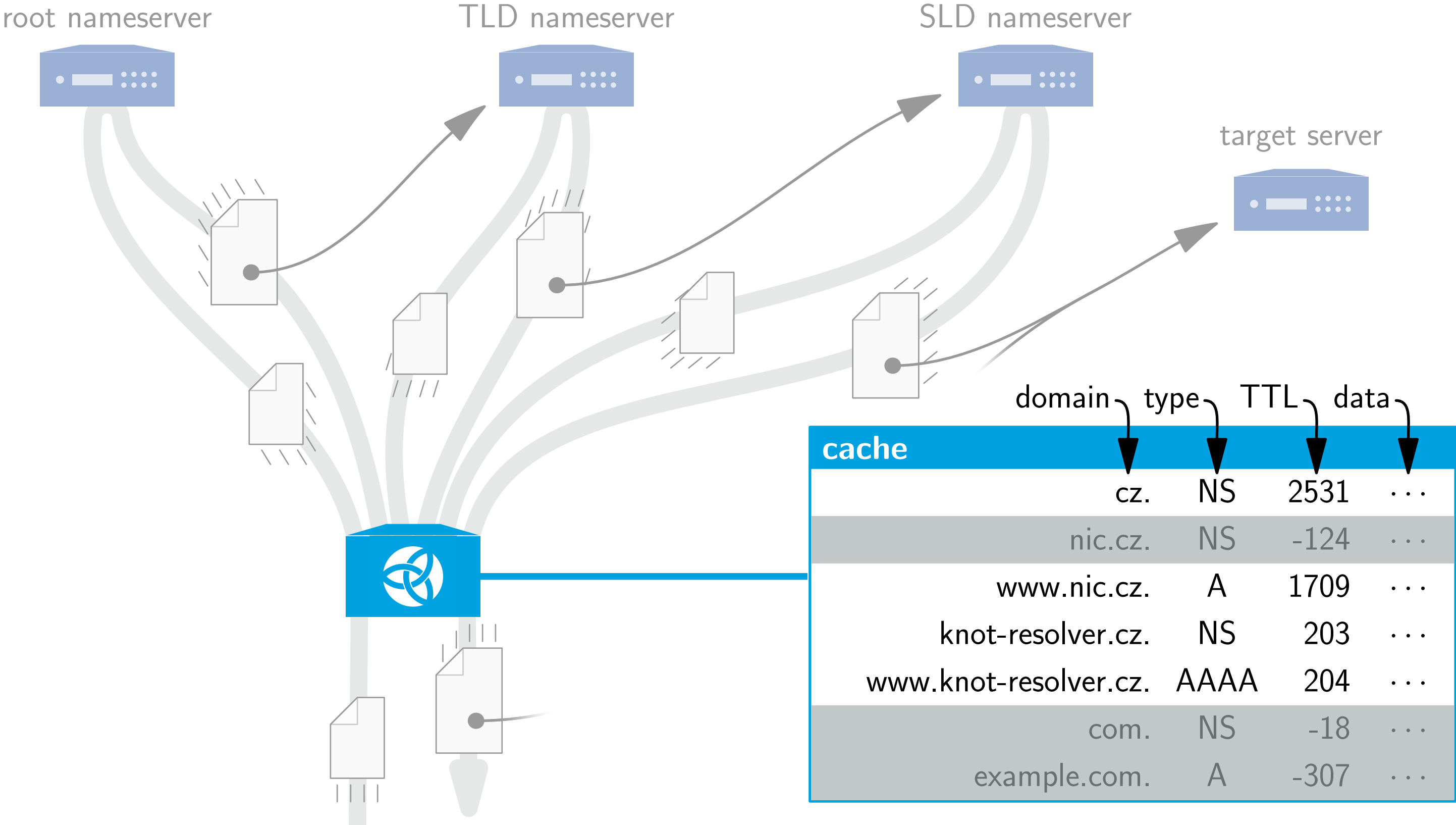
Overview



Overview

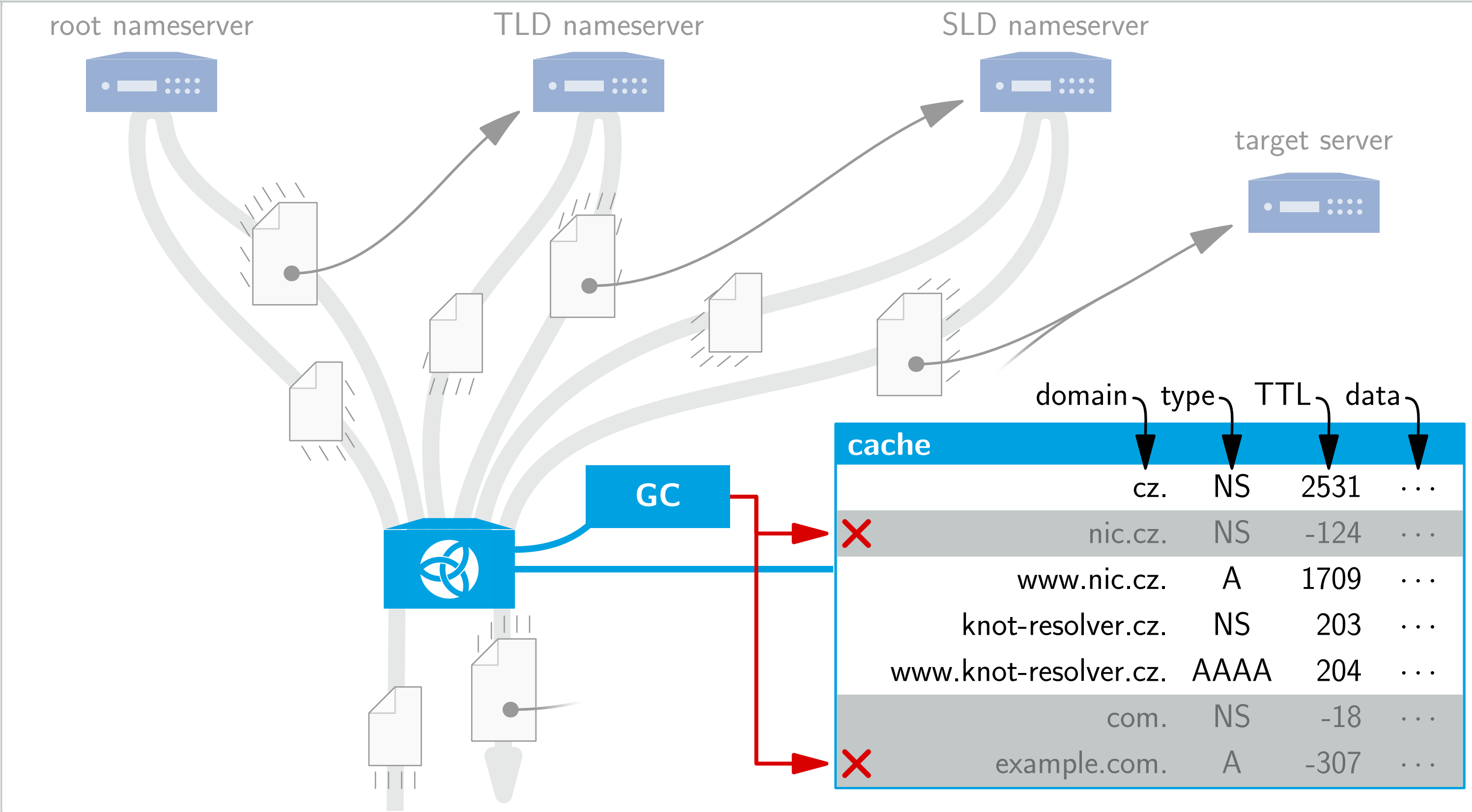


Overview



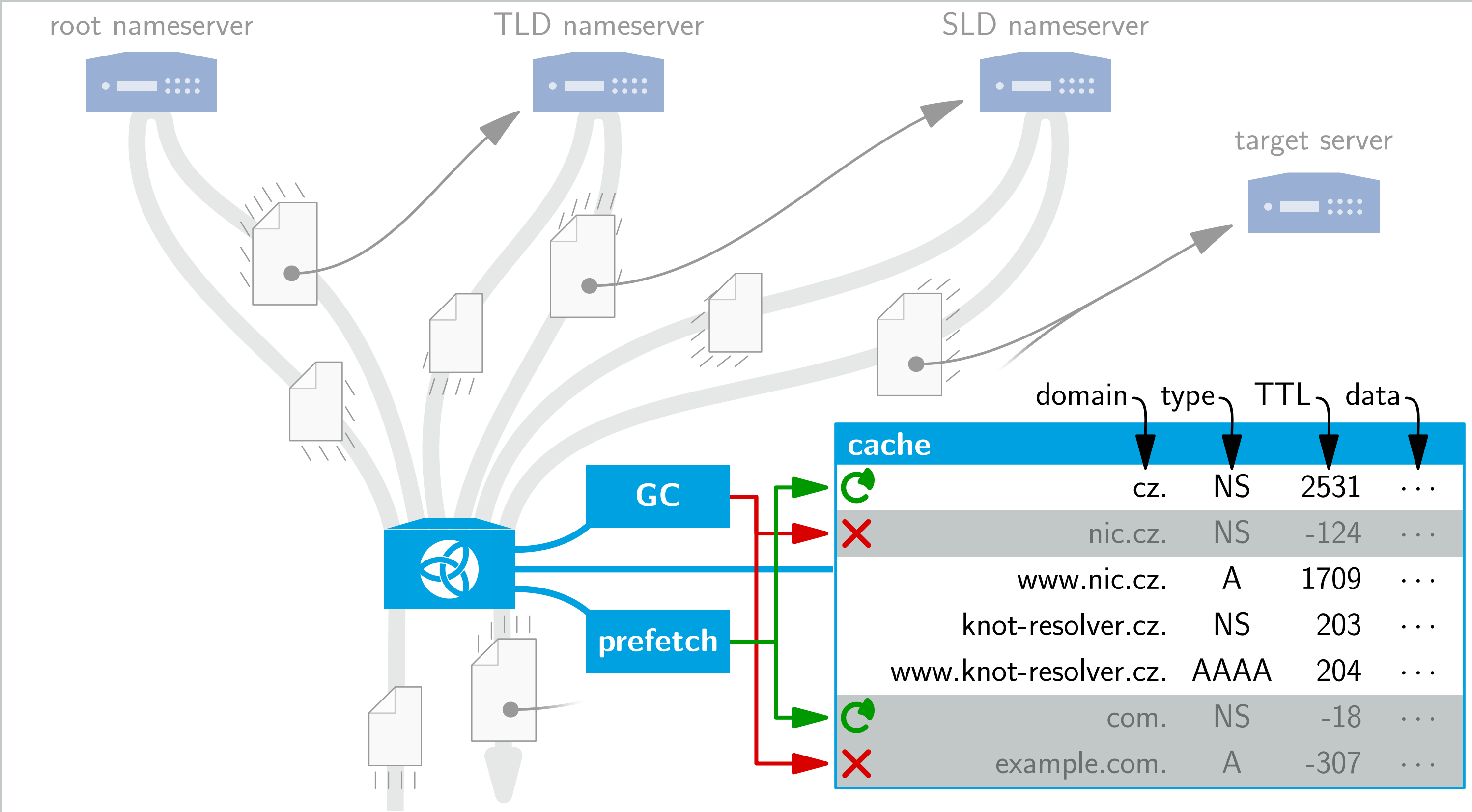
Overview

- garbage collector
 - improved
 - to be released soon



Overview

- garbage collector
 - improved
 - to be released soon
- prefetch expired
 - planned



Simple garbage collector

cache				domain	type	TTL	data
www.example.com. A 145		www.knot-dns.cz. AAAA 911 ...		nic.cz. DS 1156			
blog.nic.cz. AAAA 388 ..		www.nic.cz. AAAA 584 .		bird.nic.cz. AAAA 502 ..			
com. NS -18		cz. DNSKEY 1719					
www.turris.cz. A -416 ..		nic.cz. SOA 1475		www.mojeid.cz. A 285 .			
nic.cz. DNSKEY 553		net. NS 5274					
. NS 73650		www.knot-resolver.cz. AAAA 204					
. DNSKEY 51505							
example.com. NS 47626 ...		www.knot-dns.cz. A -451 ..		akademie.nic.cz. A 267 ..			
akademie.nic.cz. AAAA -368 ..		jaknainternet.cz. AAAA -434 .		www.mojeid.cz. AAAA 238 .			
www.turris.cz. AAAA -533 ...		example.com. A -307		knot-resolver.cz. NS 203			

Simple garbage collector

- once per sec:
 - >80% occupied
 - ⇒ 10% removed

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired
 - valid ≥3rd level

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired
 - valid ≥3rd level
 - valid 2nd level

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired
 - valid ≥3rd level
 - valid 2nd level
 - ...

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired
 - valid ≥3rd level
 - valid 2nd level
 - ...
- minor criteria
 - record type
 - oversized
 - randomness

cache			
domain	type	TTL	data
www.example.com.	A	145	
www.knot-dns.cz.	AAAA	911	...
nic.cz.	DS	1156	
blog.nic.cz.	AAAA	388	..
www.nic.cz.	AAAA	584	.
bird.nic.cz.	AAAA	502	..
com.	NS	-18	
cz.	DNSKEY	1719	
www.turris.cz.	A	-416	..
nic.cz.	SOA	1475	
www.mojeid.cz.	A	285	.
nic.cz.	DNSKEY	553	
net.	NS	5274	
.	NS	73650	
www.knot-resolver.cz.	AAAA	204	
.	DNSKEY	51505	
example.com.	NS	47626	...
www.knot-dns.cz.	A	-451	..
akademie.nic.cz.	A	267	..
akademie.nic.cz.	AAAA	-368	..
jaknainternet.cz.	AAAA	-434	.
www.mojeid.cz.	AAAA	238	.
www.turris.cz.	AAAA	-533	...
example.com.	A	-307	
knot-resolver.cz.	NS	203	

Simple garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- first remove:
 - expired
 - valid ≥3rd level
 - valid 2nd level
 - ...
- minor criteria
 - record type
 - oversized
 - randomness

cache									
www.example.com. A 145				www.knot-dns.cz. AAAA 911 ...			nic.cz. DS 1156		
blog.nic.cz. AAAA 388 ..			www.nic.cz. AAAA 584 .			bird.nic.cz. AAAA 502 ..			
com. NS -18				cz. DNSKEY 1719					
www.turris.cz. A -416 ✗			nic.cz. SOA 1475			www.mojeid.cz. A 285 .			
nic.cz. DNSKEY 553					net. NS 5274				
. NS 73650					www.knot-resolver.cz. AAAA 204				
. DNSKEY 51505									
example.com. NS 47626 ...			www.knot-dns.cz. A -451 ✗			akademie.nic.cz. A 267 ..			
akademie.nic.cz. AAAA -368 ✗			jaknainternet.cz. AAAA -434 ✗			www.mojeid.cz. AAAA 238 .			
www.turris.cz. AAAA -533 ...			example.com. A -307				knot-resolver.cz. NS 203		

KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories

cache			
domain	type	TTL	data
nic.cz. DNSKEY 553	example.com. NS 47626 ...		
www.mojeid.cz. AAAA 238 .	www.knot-dns.cz. A -451 ..		
net. NS 5274	example.com. A -307		
akademie.nic.cz. A 267 ..	blog.nic.cz. AAAA 388 ..		
com. NS -18	www.turris.cz. AAAA -533 ...		
cz. DNSKEY 1719	jaknainternet.cz. AAAA -434 .		
. DNSKEY 51505			
www.example.com. A 145	www.knot-dns.cz. AAAA 911 ...		
. NS 73650	akademie.nic.cz. AAAA -368 ..		
www.knot-resolver.cz. AAAA 204	knot-resolver.cz. NS 203		

6 – 13 %

KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories

cache				domain	type	TTL	data
nic.cz.	DNSKEY	553		example.com.	NS	47626	...
www.mojeid.cz.	AAAA	238	.	www.knot-dns.cz.	A	-451	..
net.	NS	5274		example.com.	A	-307	
akademie.nic.cz.	A	267	..	blog.nic.cz.	AAAA	388	..
com.	NS	-18		www.turris.cz.	AAAA	-533	...
cz.	DNSKEY	1719		jaknainternet.cz.	AAAA	-434	.
.	DNSKEY	51505					
www.example.com.	A	145		www.knot-dns.cz.	AAAA	911	...
.	NS	73650		akademie.nic.cz.	AAAA	-368	..
www.knot-resolver.cz.	AAAA	204		knot-resolver.cz.	NS	203	

KRU		accesses
cz.	DNSKEY	1230
com.	NS	997
.	NS	991
nic.cz.	DNSKEY	964
akademie.nic.cz.	A	851
www.knot-dns.cz.	A	748
net.	NS	696
www.mojeid.cz.	AAAA	630
example.com.	A	628
www.turris.cz.	AAAA	622
www.example.com.	A	610
example.com.	NS	494
knot-resolver.cz.	NS	463
.	DNSKEY	432
www.knot-dns.cz.	AAAA	426

6 – 13 %

KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - accesses per byte

cache				domain	type	TTL	data
nic.cz.	DNSKEY	553		example.com.	NS	47626	...
www.mojeid.cz.	AAAA	238	.	www.knot-dns.cz.	A	-451	..
net.	NS	5274		example.com.	A	-307	
akademie.nic.cz.	A	267	..	blog.nic.cz.	AAAA	388	..
com.	NS	-18		www.turris.cz.	AAAA	-533	...
cz.	DNSKEY	1719		jaknainternet.cz.	AAAA	-434	.
.	DNSKEY	51505					
www.example.com.	A	145		www.knot-dns.cz.	AAAA	911	...
.	NS	73650		akademie.nic.cz.	AAAA	-368	..
www.knot-resolver.cz.	AAAA	204		knot-resolver.cz.	NS	203	

KRU		accesses/size
cz. DNSKEY		32
akademie.nic.cz. A		31
. NS		26
com. NS		25
www.knot-dns.cz. A		25
nic.cz. DNSKEY		24
www.mojeid.cz. AAAA		21
example.com. A		20
www.turris.cz. AAAA		20
net. NS		18
www.example.com. A		17
example.com. NS		17
knot-resolver.cz. NS		15
www.knot-dns.cz. AAAA		13
akademie.nic.cz. AAAA		9

6 – 13 %

KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte

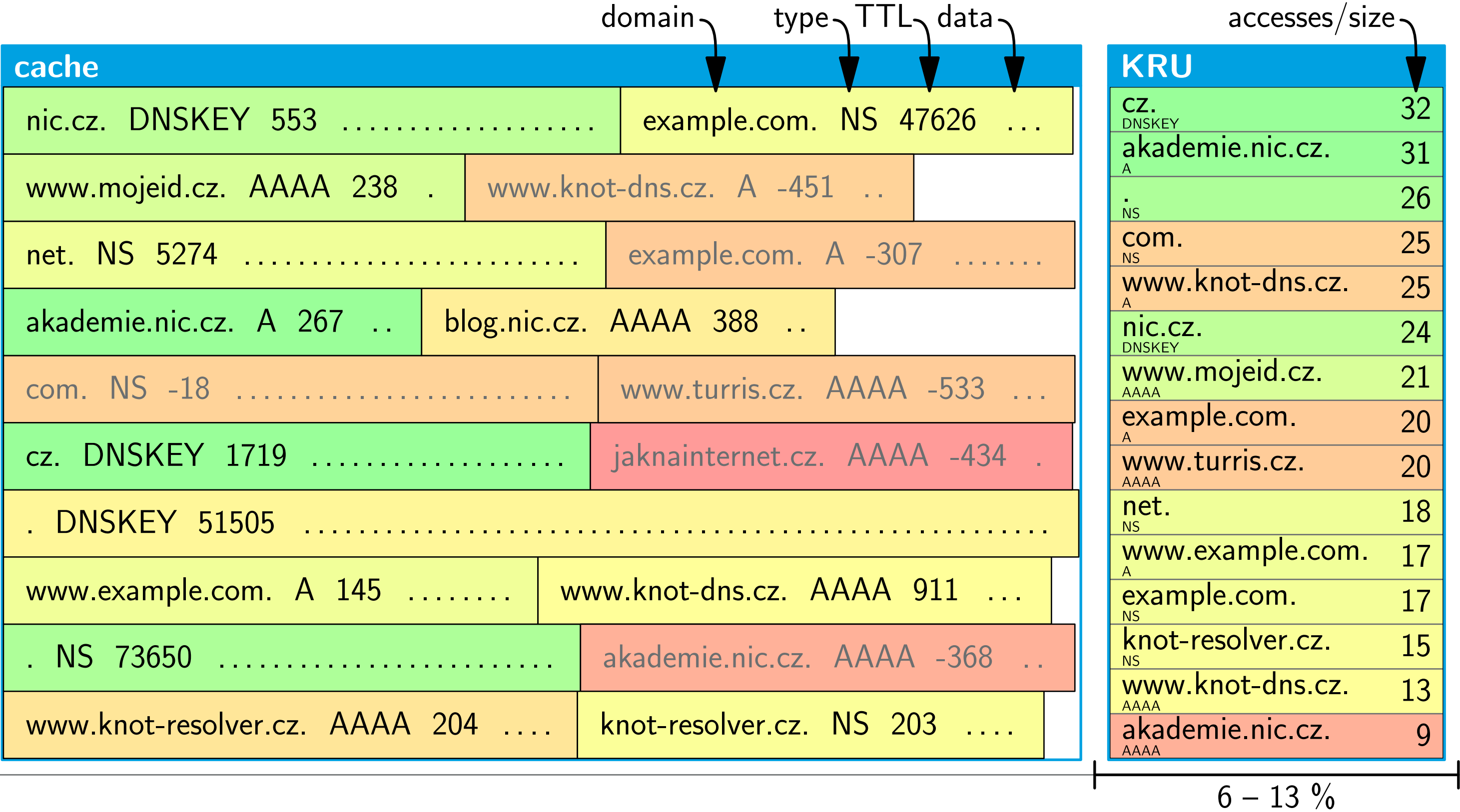
cache				domain	type	TTL	data
nic.cz.	DNSKEY	553		example.com.	NS	47626	...
www.mojeid.cz.	AAAA	238	.	www.knot-dns.cz.	A	-451	..
net.	NS	5274		example.com.	A	-307	
akademie.nic.cz.	A	267	..	blog.nic.cz.	AAAA	388	..
com.	NS	-18		www.turris.cz.	AAAA	-533	...
cz.	DNSKEY	1719		jaknainternet.cz.	AAAA	-434	.
.	DNSKEY	51505					
www.example.com.	A	145		www.knot-dns.cz.	AAAA	911	...
.	NS	73650		akademie.nic.cz.	AAAA	-368	..
www.knot-resolver.cz.	AAAA	204		knot-resolver.cz.	NS	203	

KRU		accesses/size
cz.	DNSKEY	32
akademie.nic.cz.	A	31
.	NS	26
com.	NS	25
www.knot-dns.cz.	A	25
nic.cz.	DNSKEY	24
www.mojeid.cz.	AAAA	21
example.com.	A	20
www.turris.cz.	AAAA	20
net.	NS	18
www.example.com.	A	17
example.com.	NS	17
knot-resolver.cz.	NS	15
www.knot-dns.cz.	AAAA	13
akademie.nic.cz.	AAAA	9

6 – 13 %

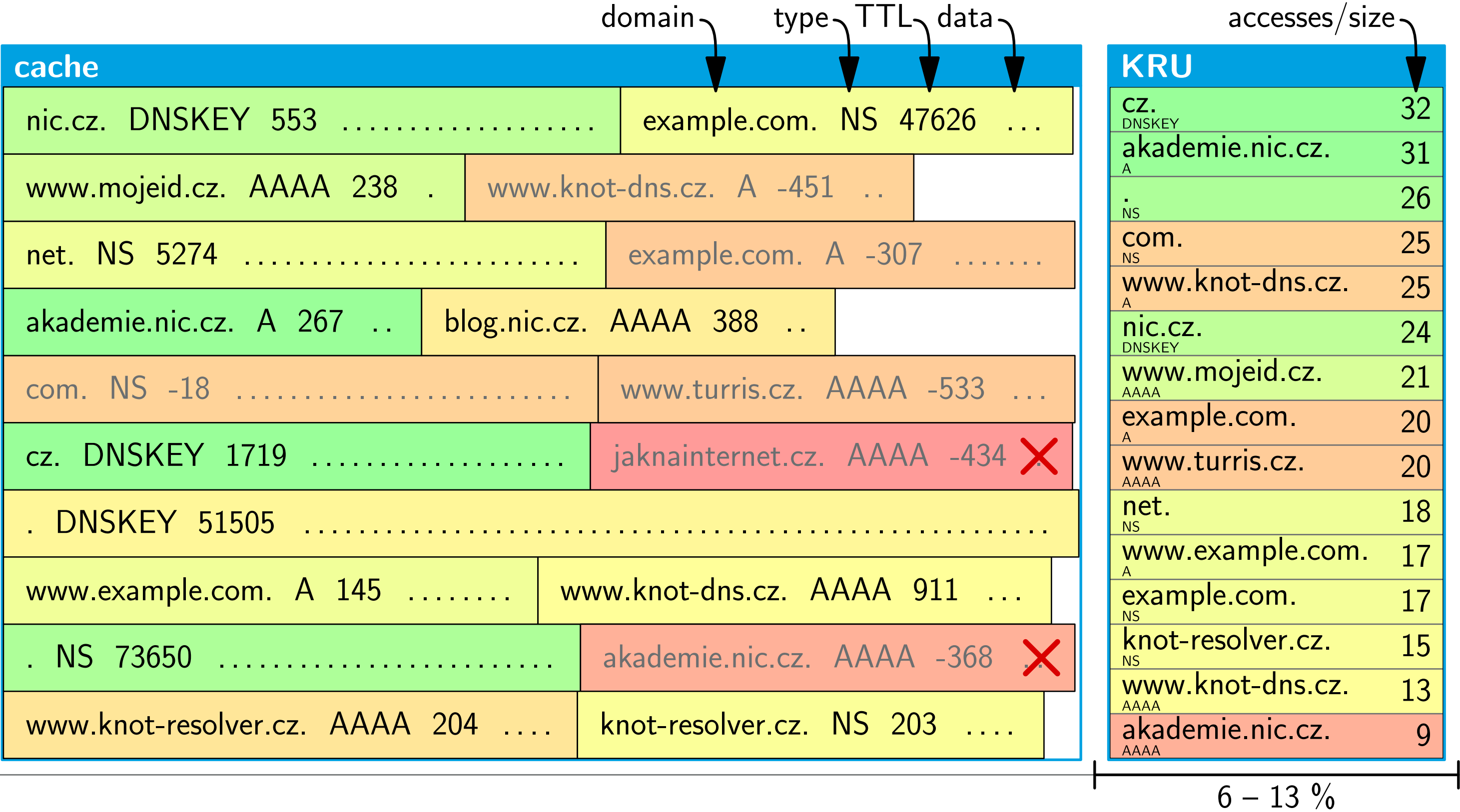
KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte



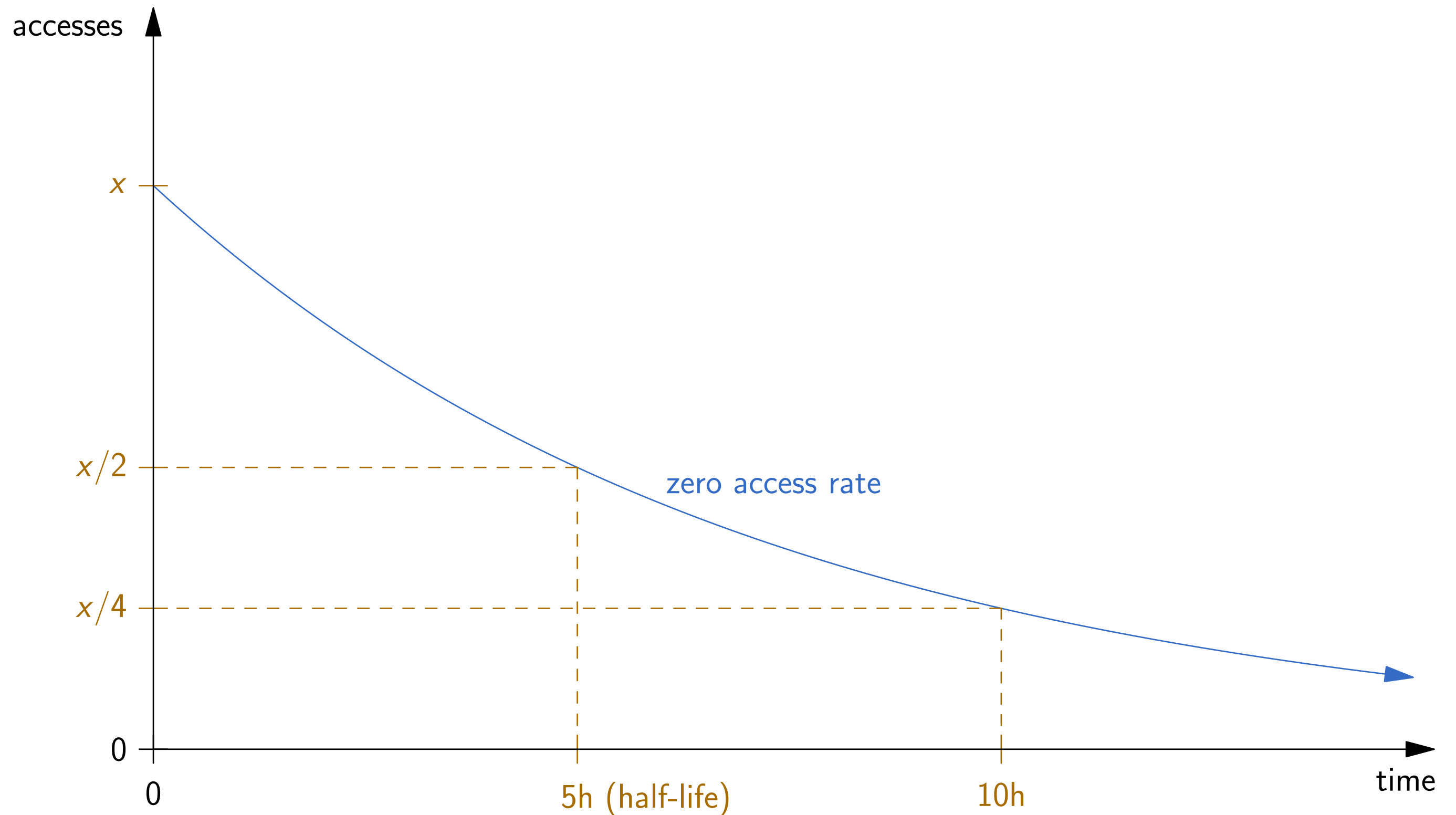
KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte



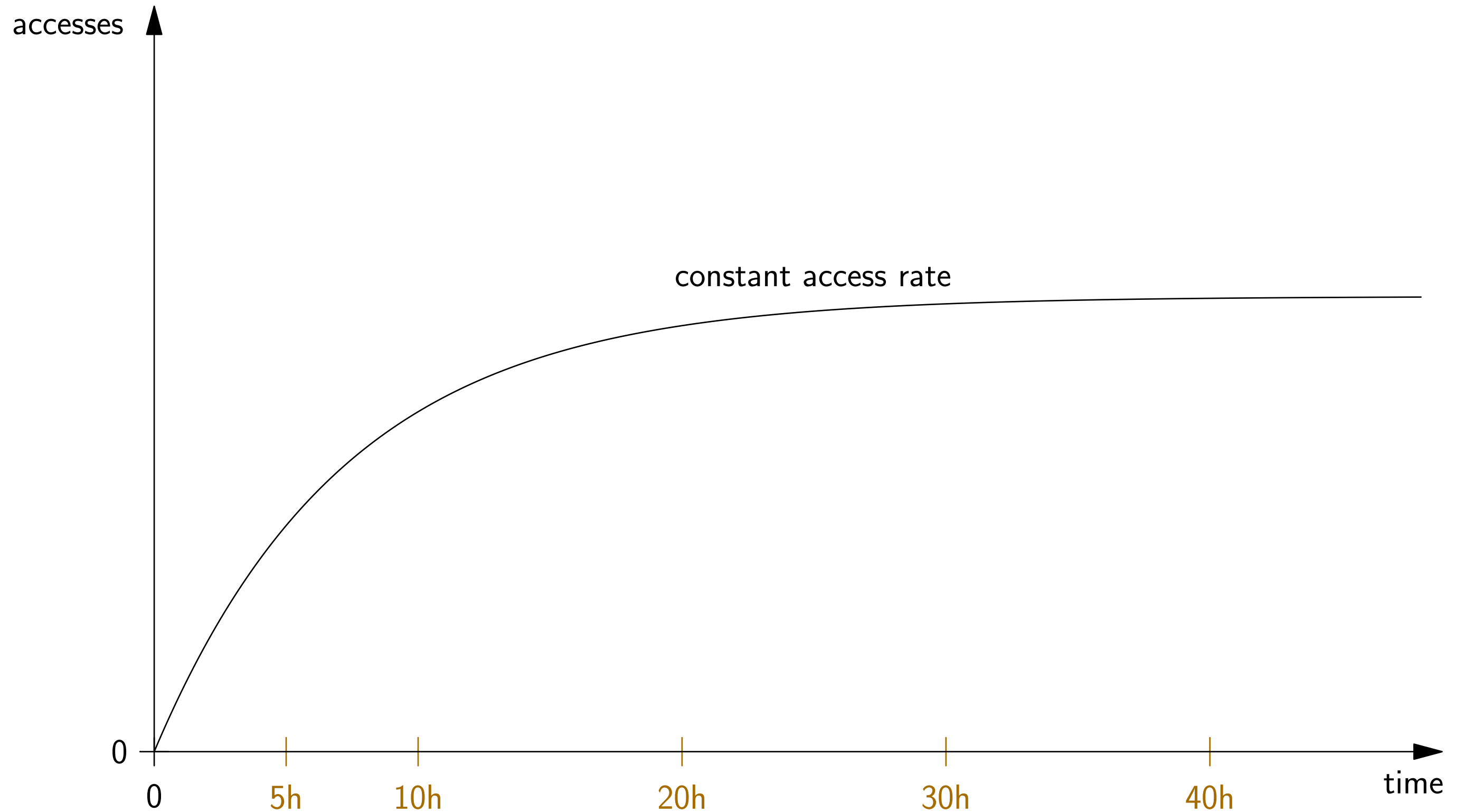
KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte
- counters
 - exponential decay



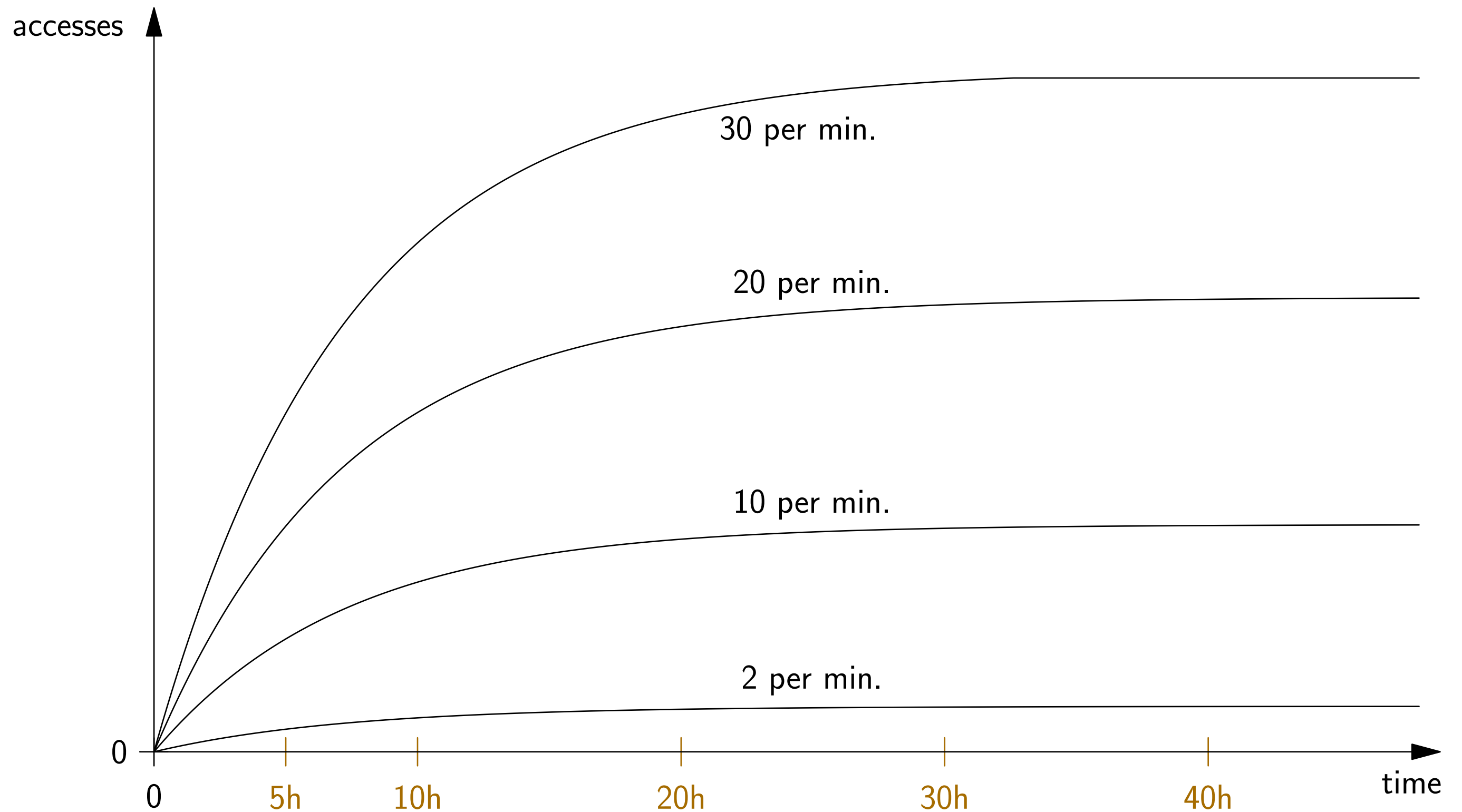
KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte
- counters
 - exponential decay
 - regular accesses converge



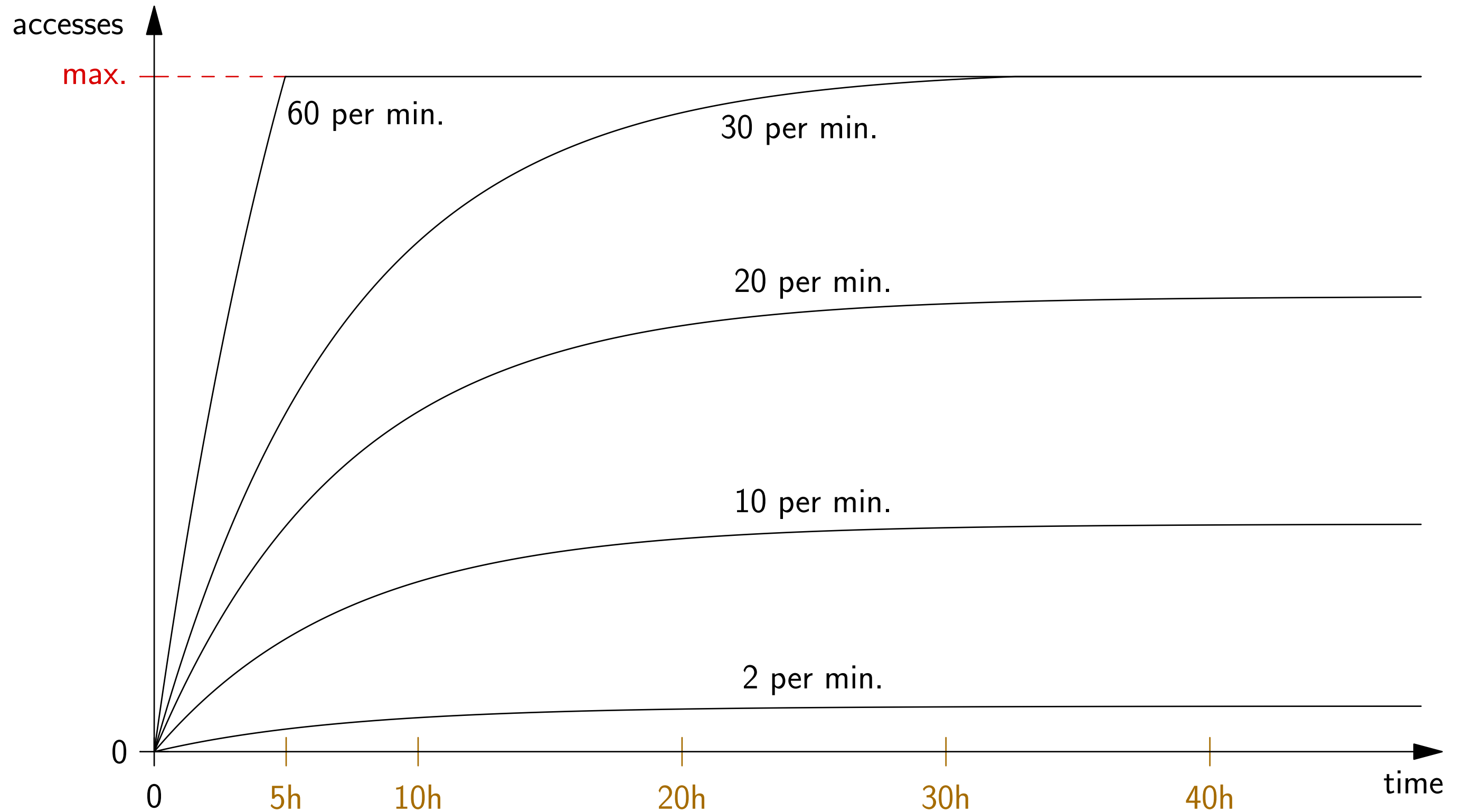
KRU-based garbage collector

- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte
- counters
 - exponential decay
 - regular accesses converge

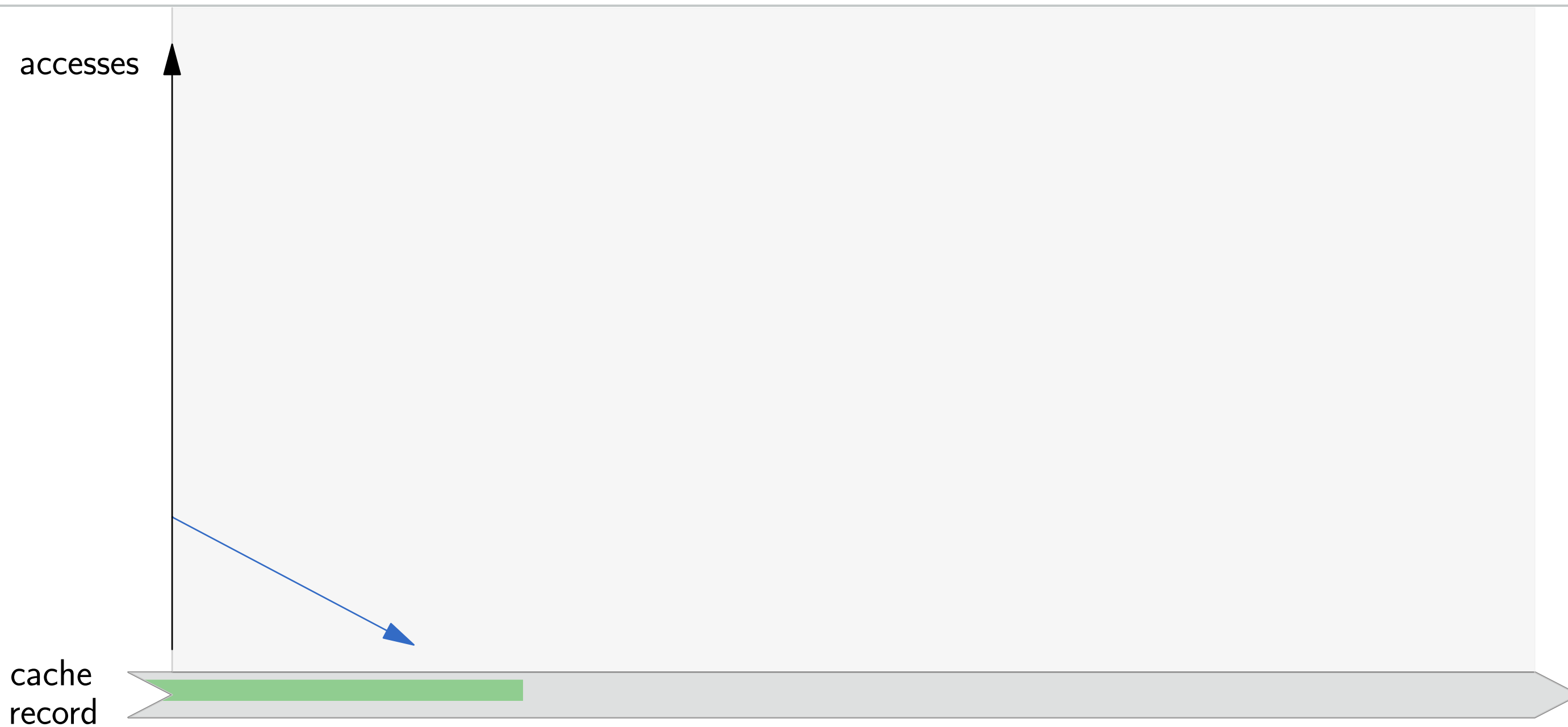


KRU-based garbage collector

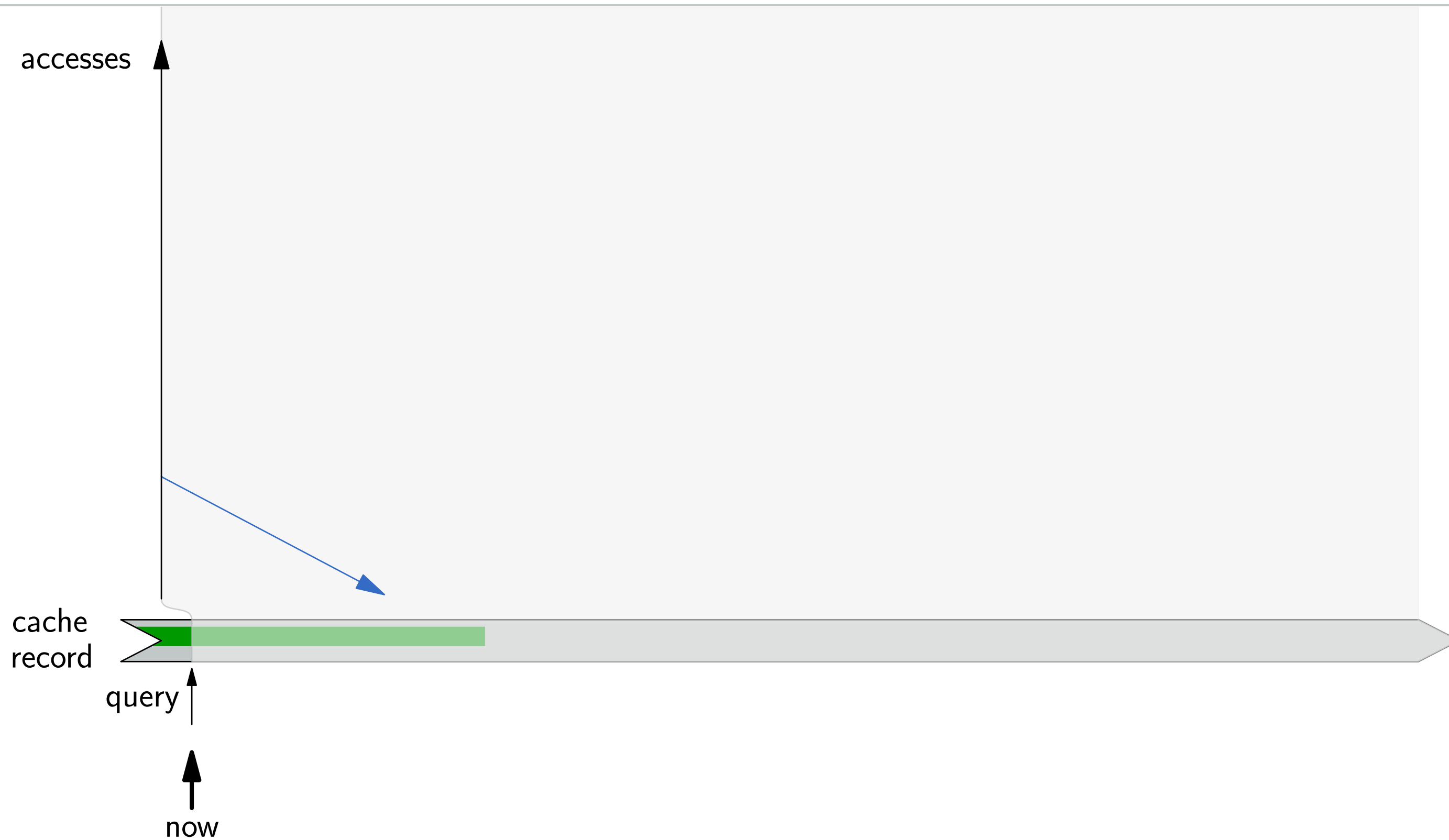
- once per sec:
 - >80% occupied
⇒ 10% removed
- lin. categorization
 - 100 categories
- criteria
 - validity
 - accesses per byte
- counters
 - exponential decay
 - regular accesses converge



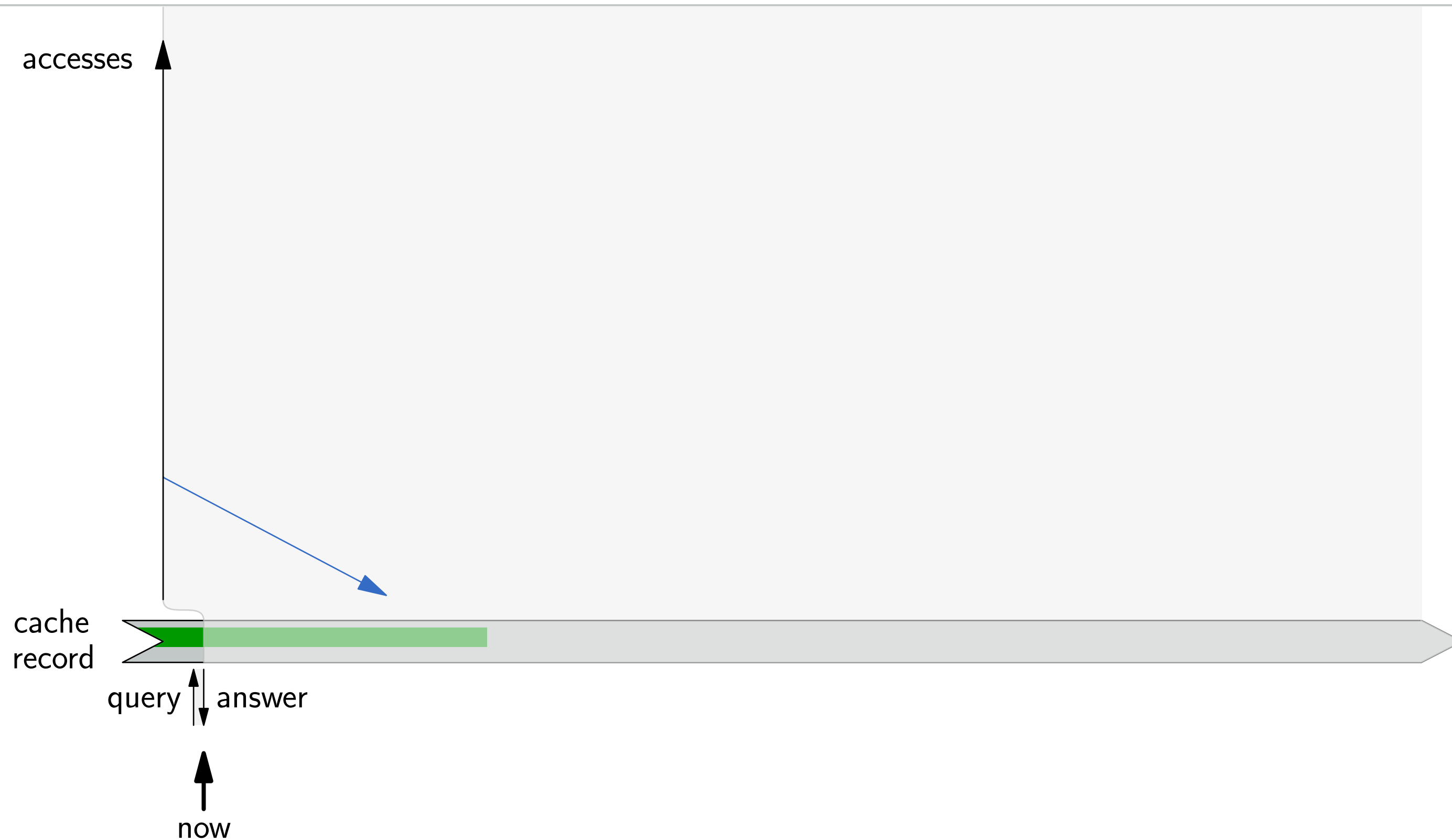
Prefetching expired records



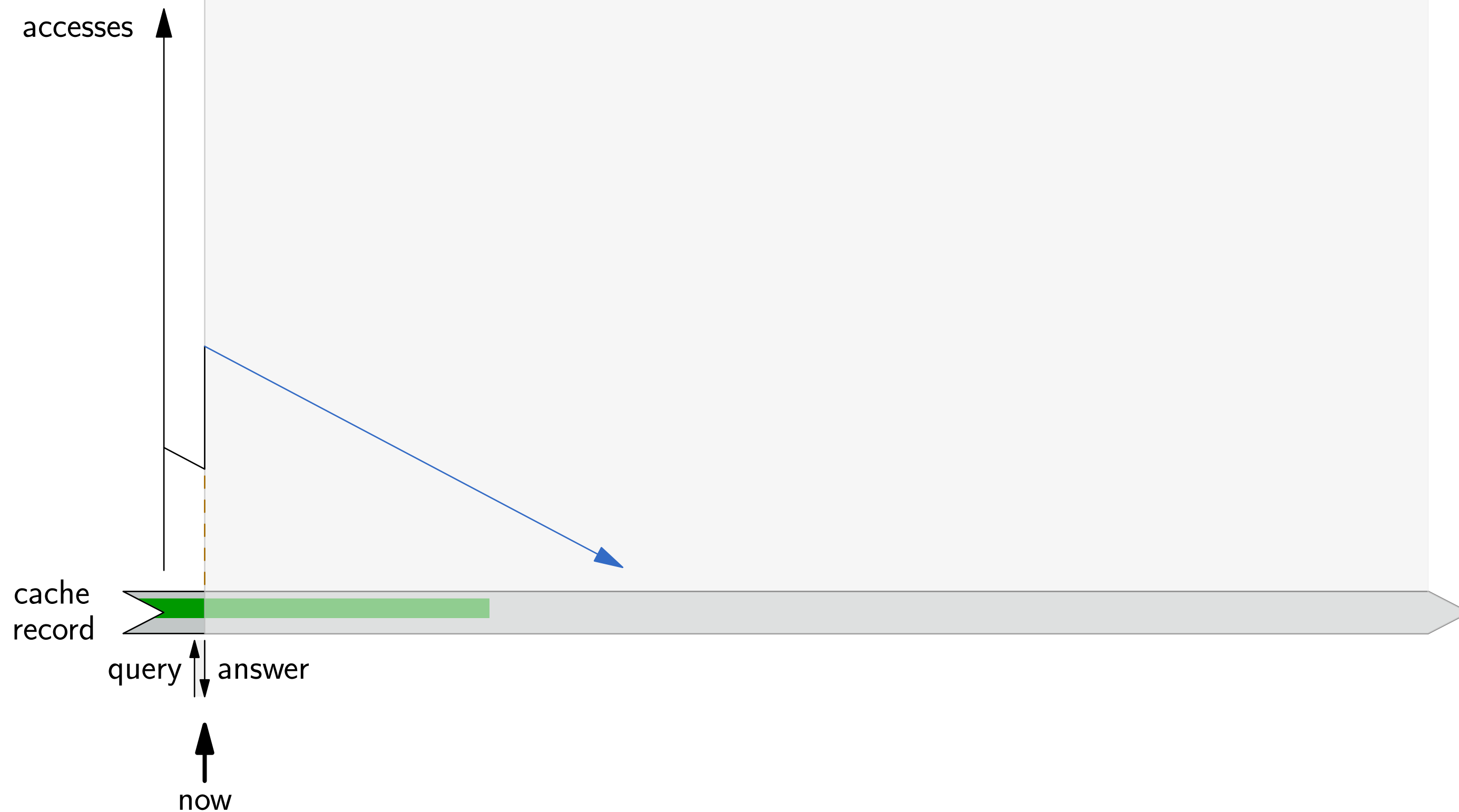
Prefetching expired records



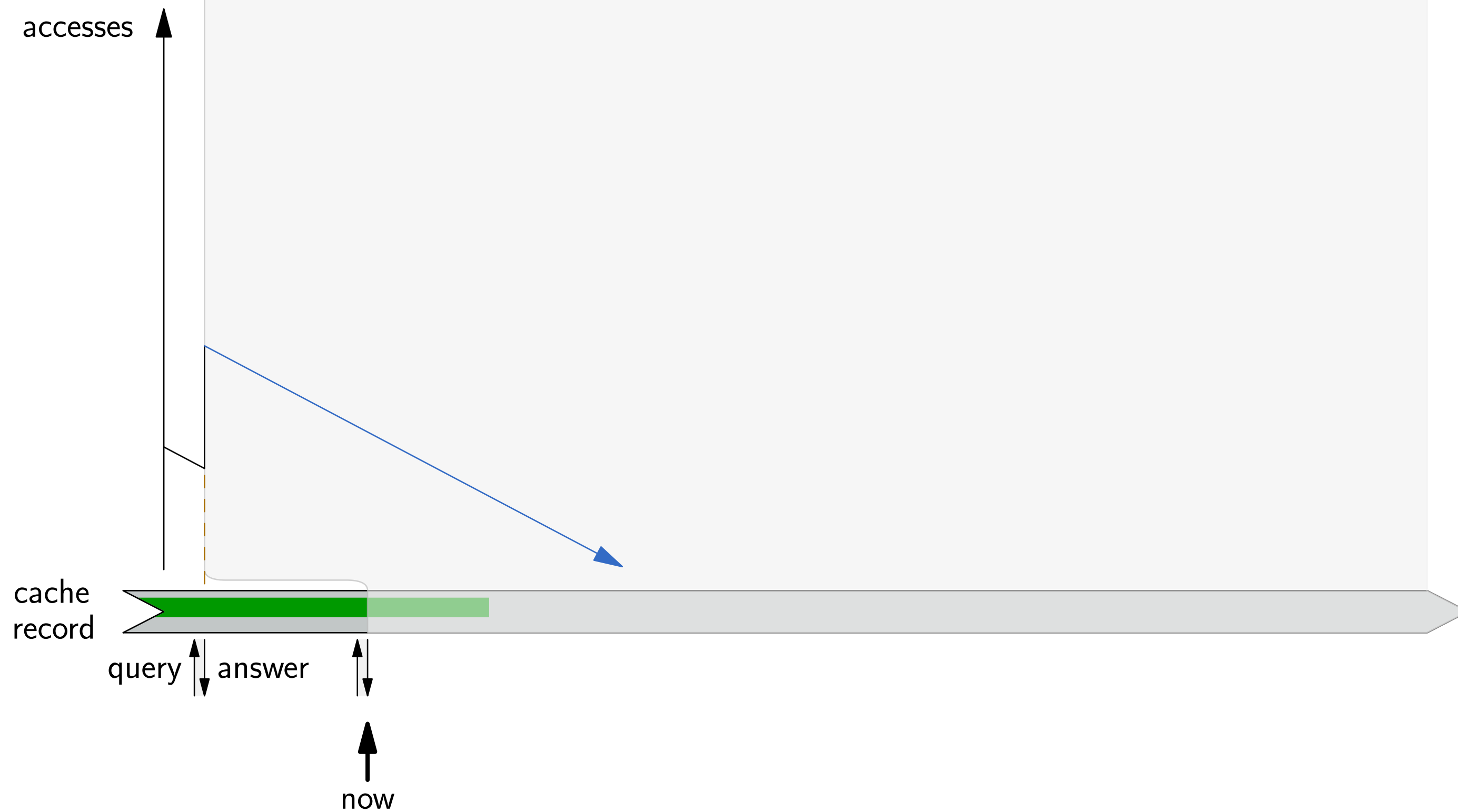
Prefetching expired records



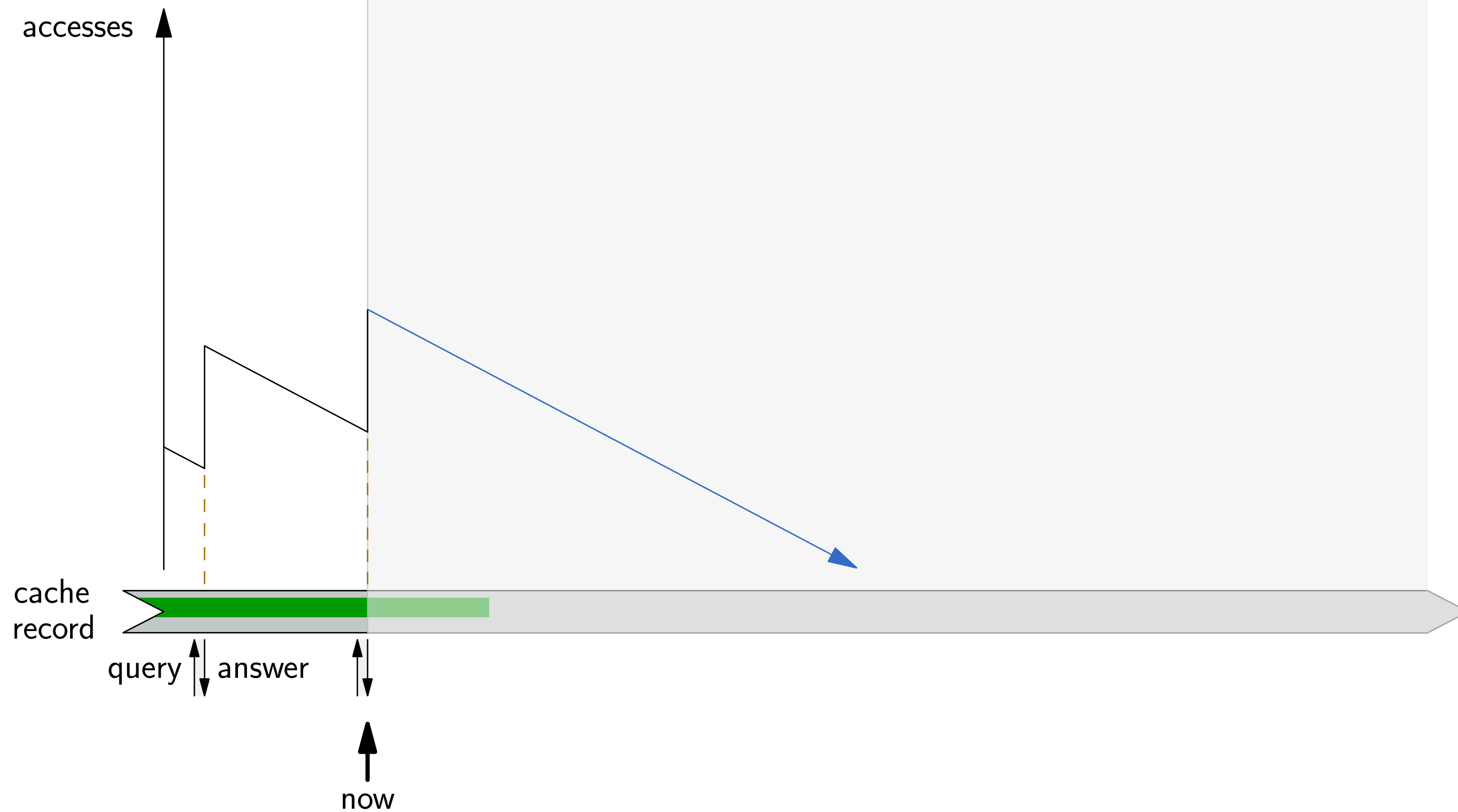
Prefetching expired records



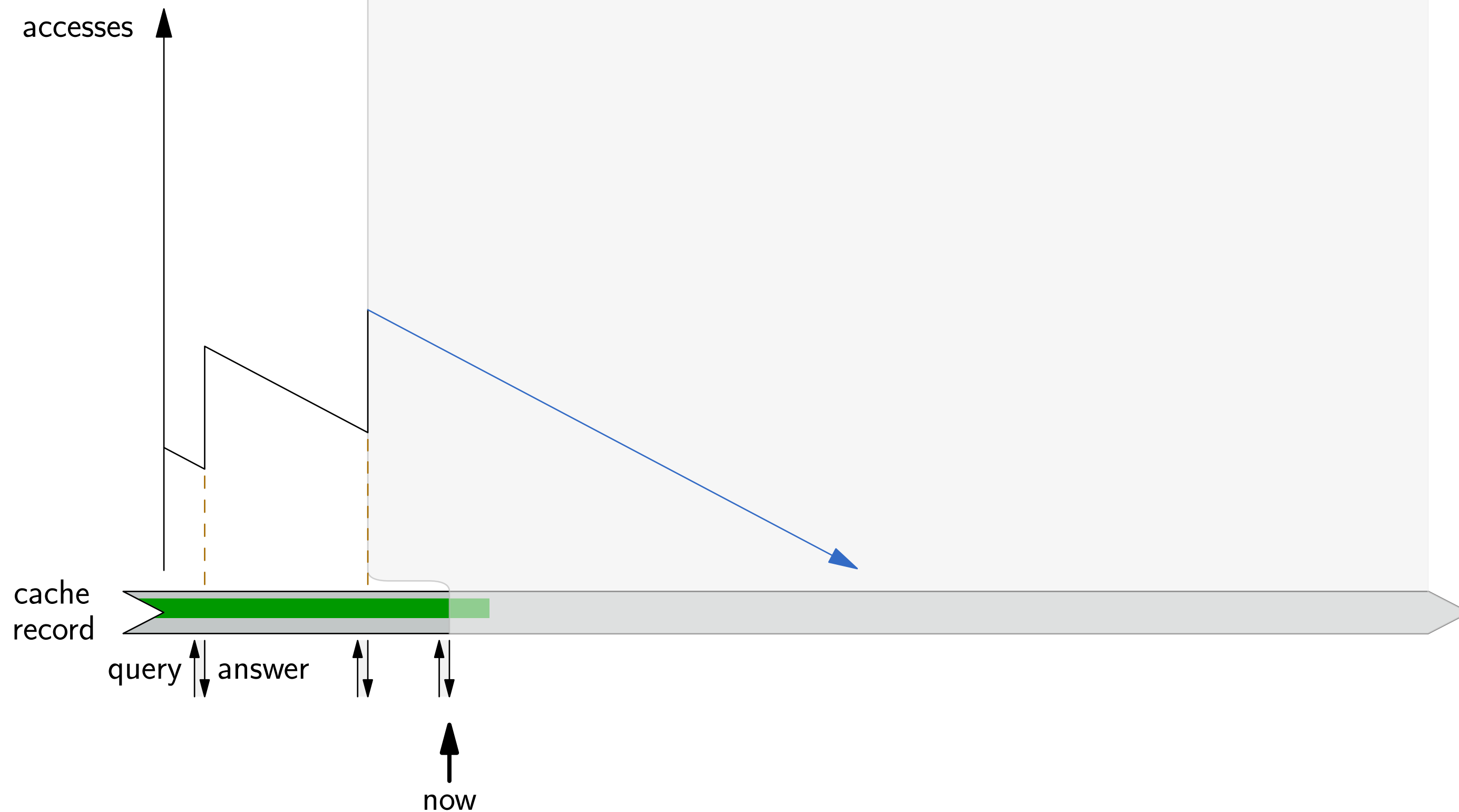
Prefetching expired records



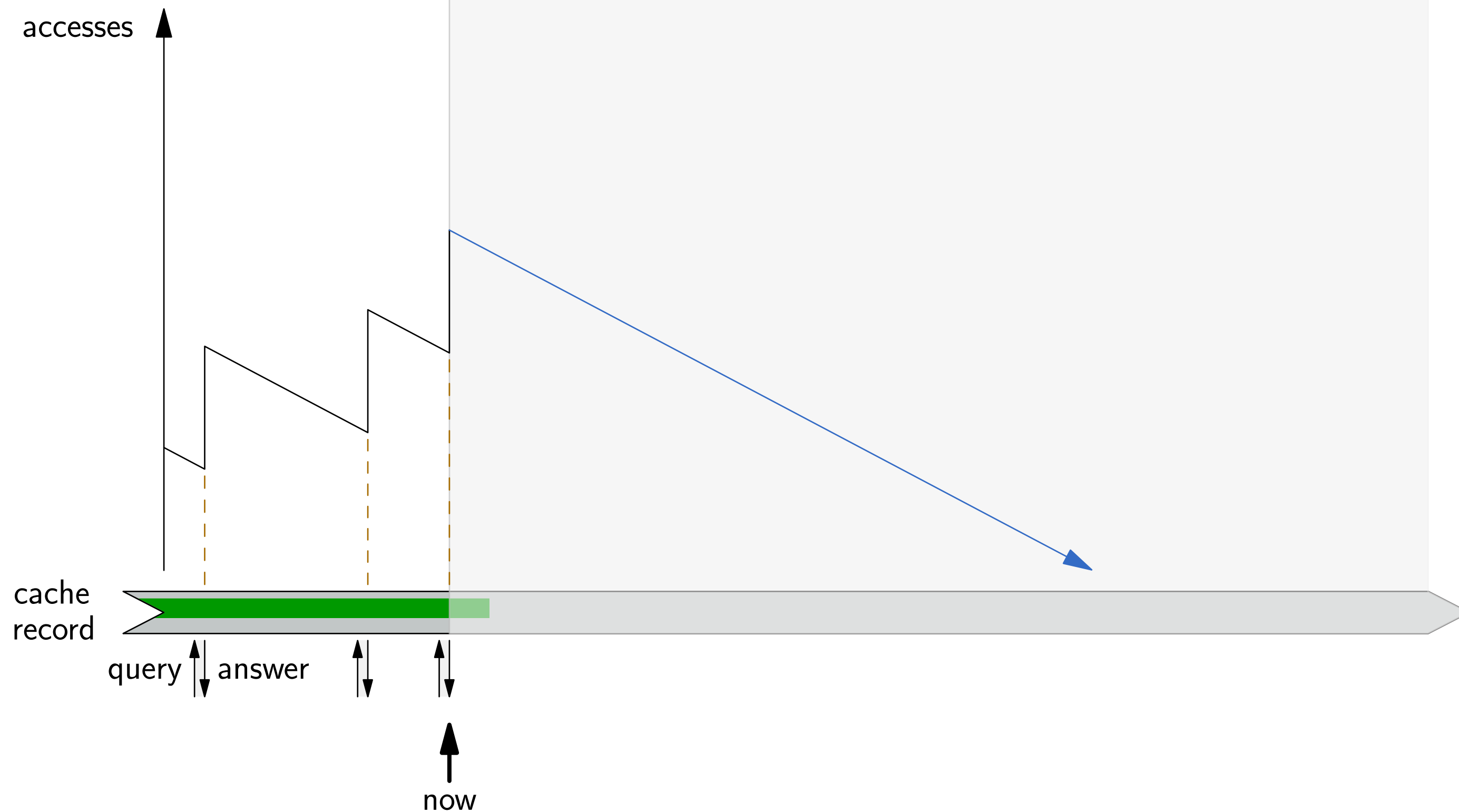
Prefetching expired records



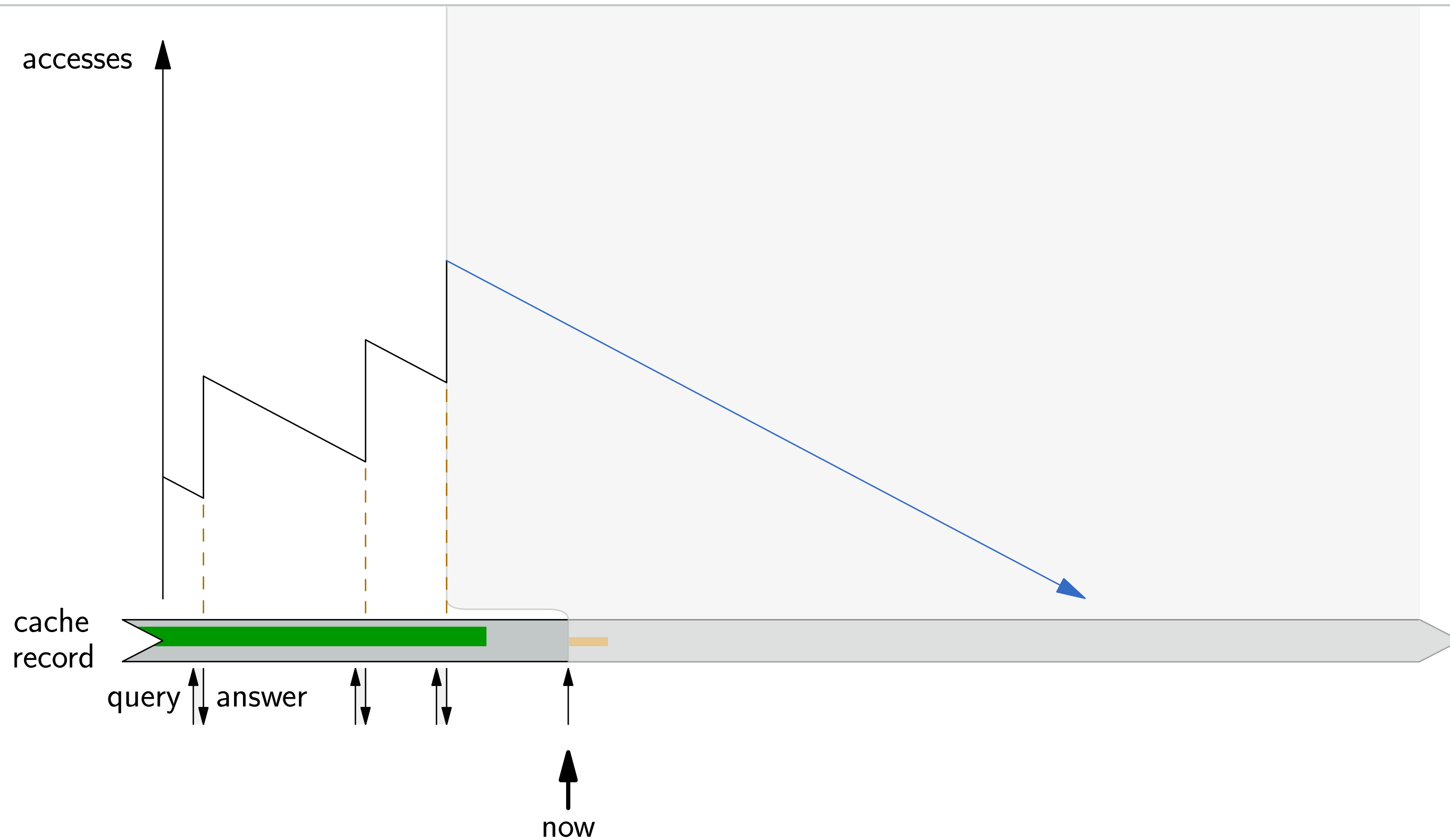
Prefetching expired records



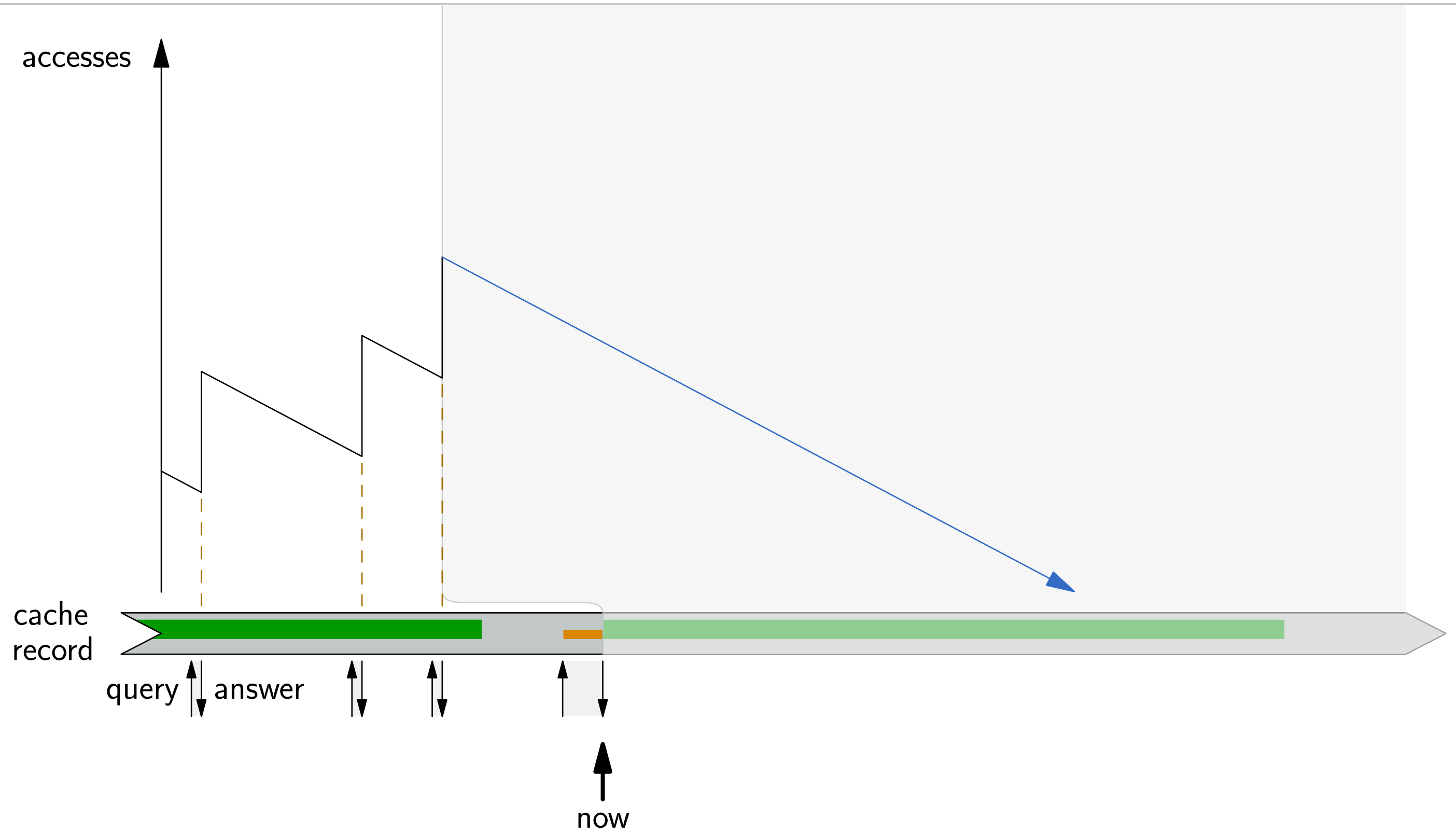
Prefetching expired records



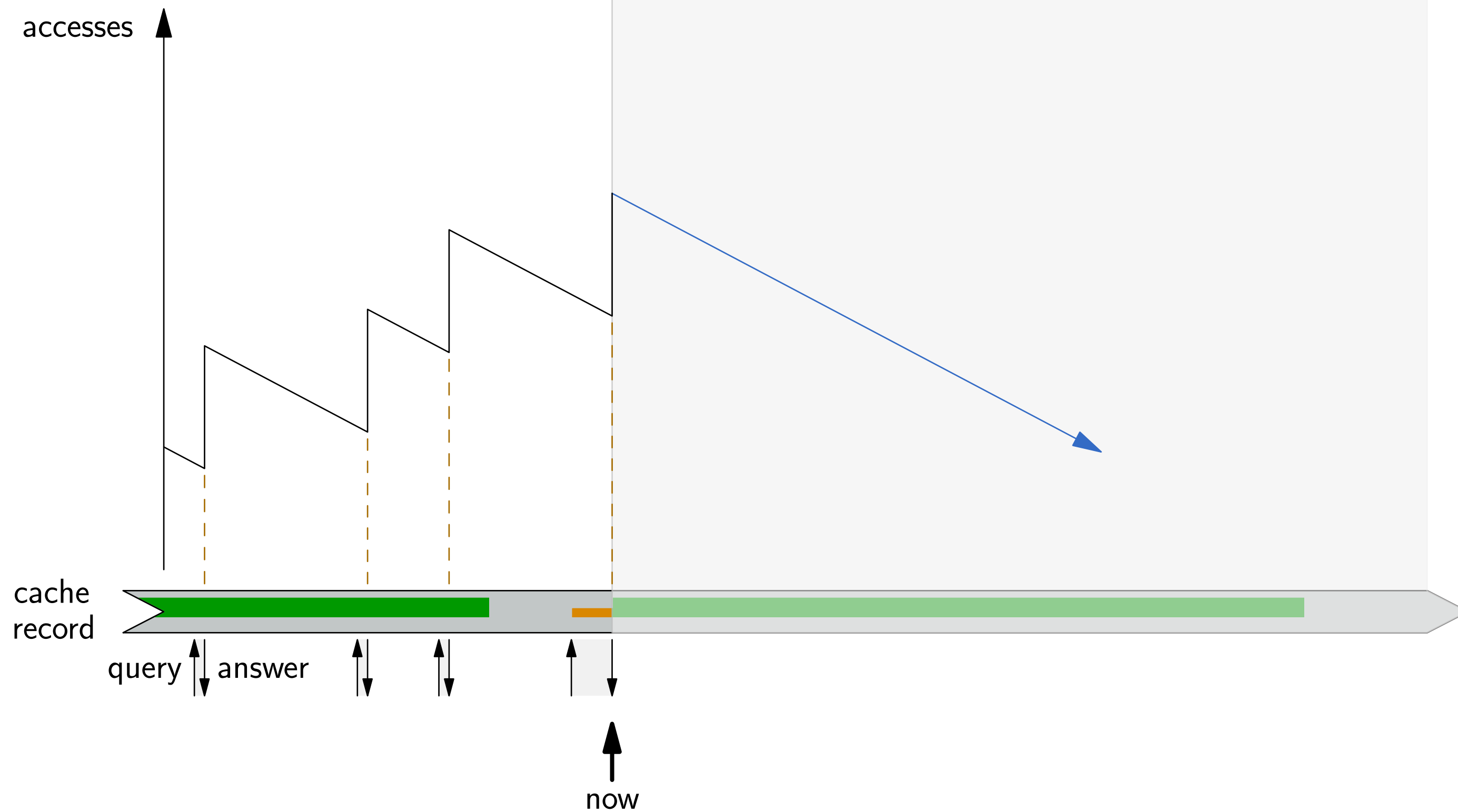
Prefetching expired records



Prefetching expired records

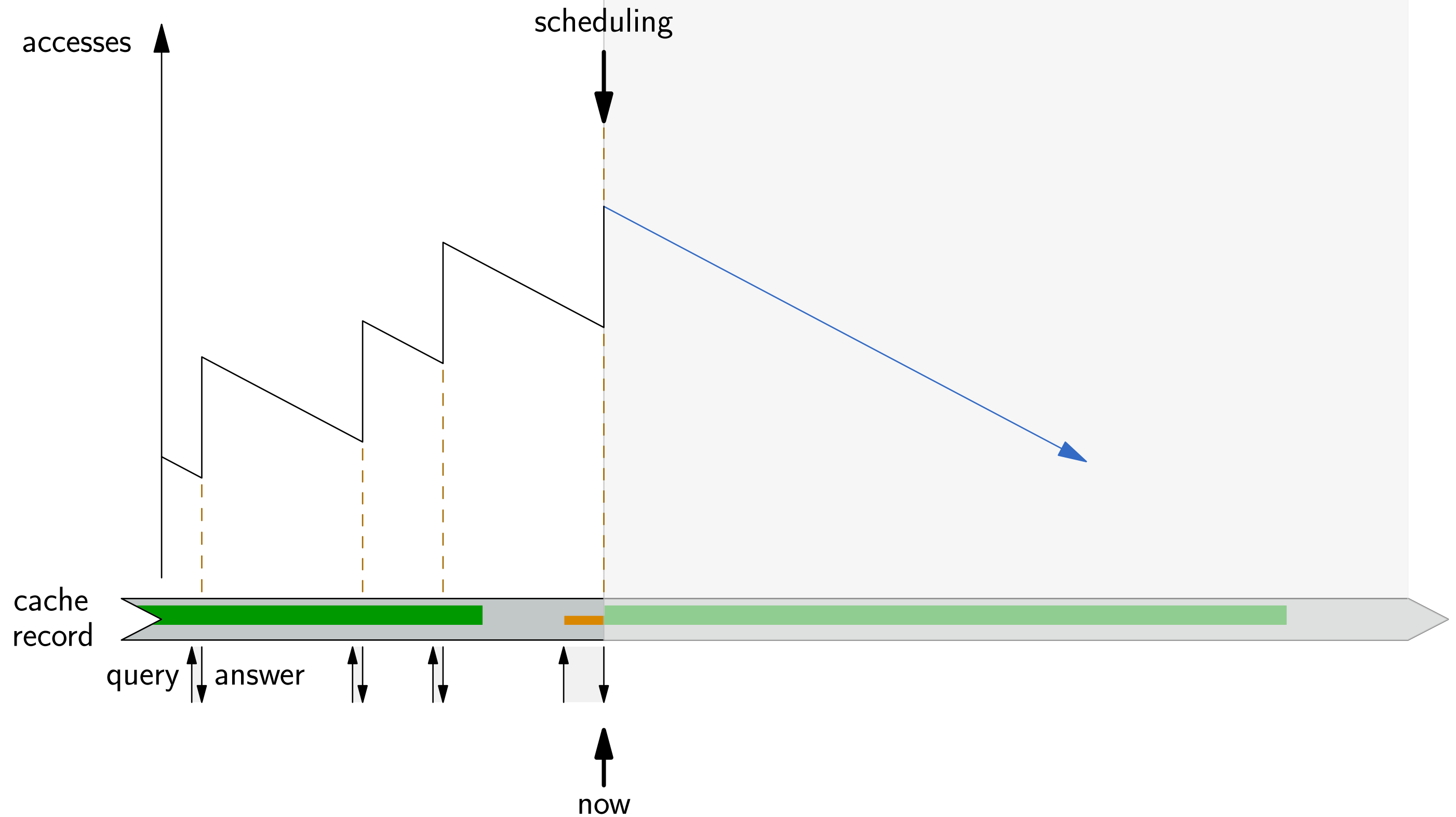


Prefetching expired records



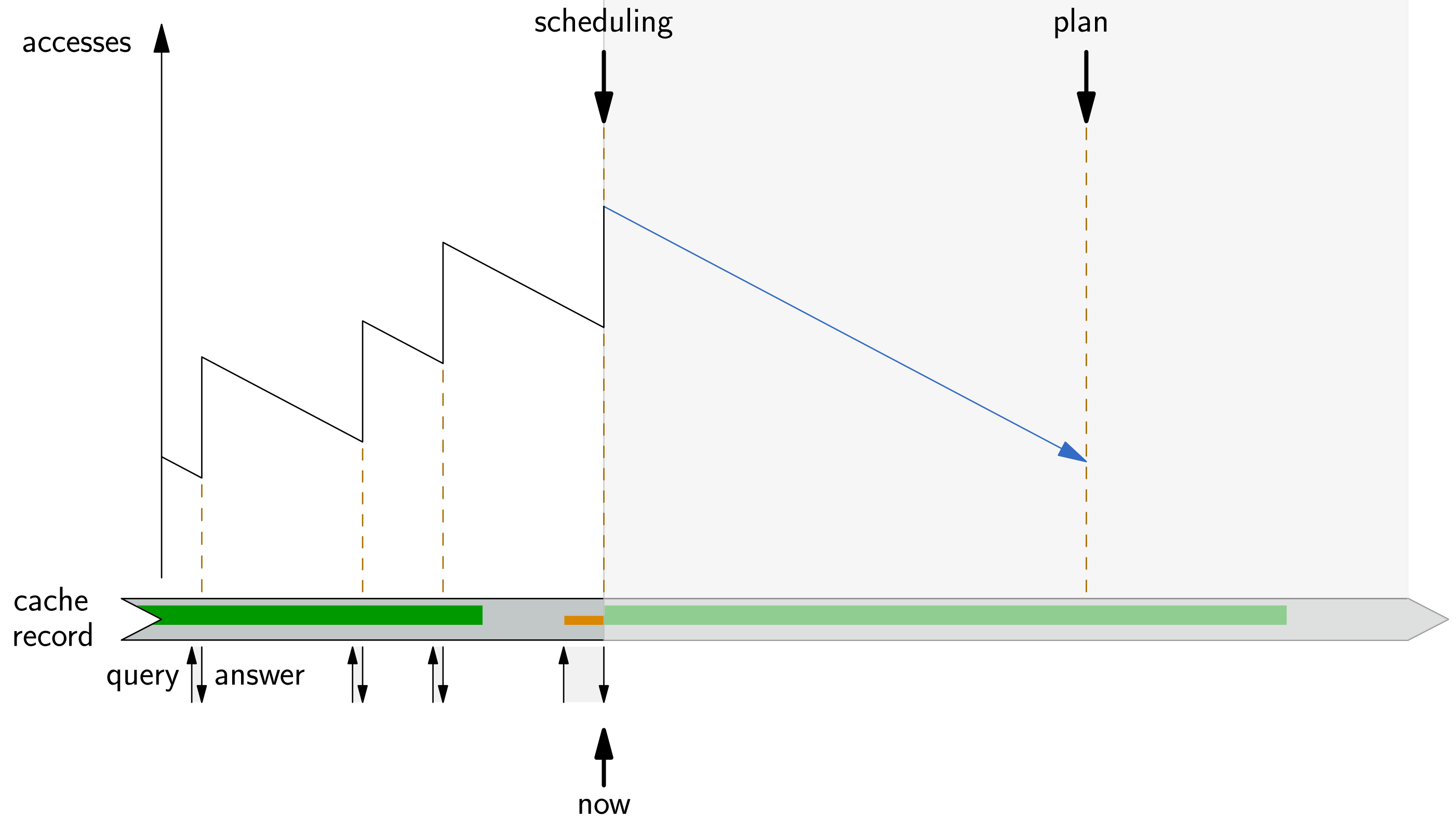
Prefetching expired records

- scheduling:
 - during access



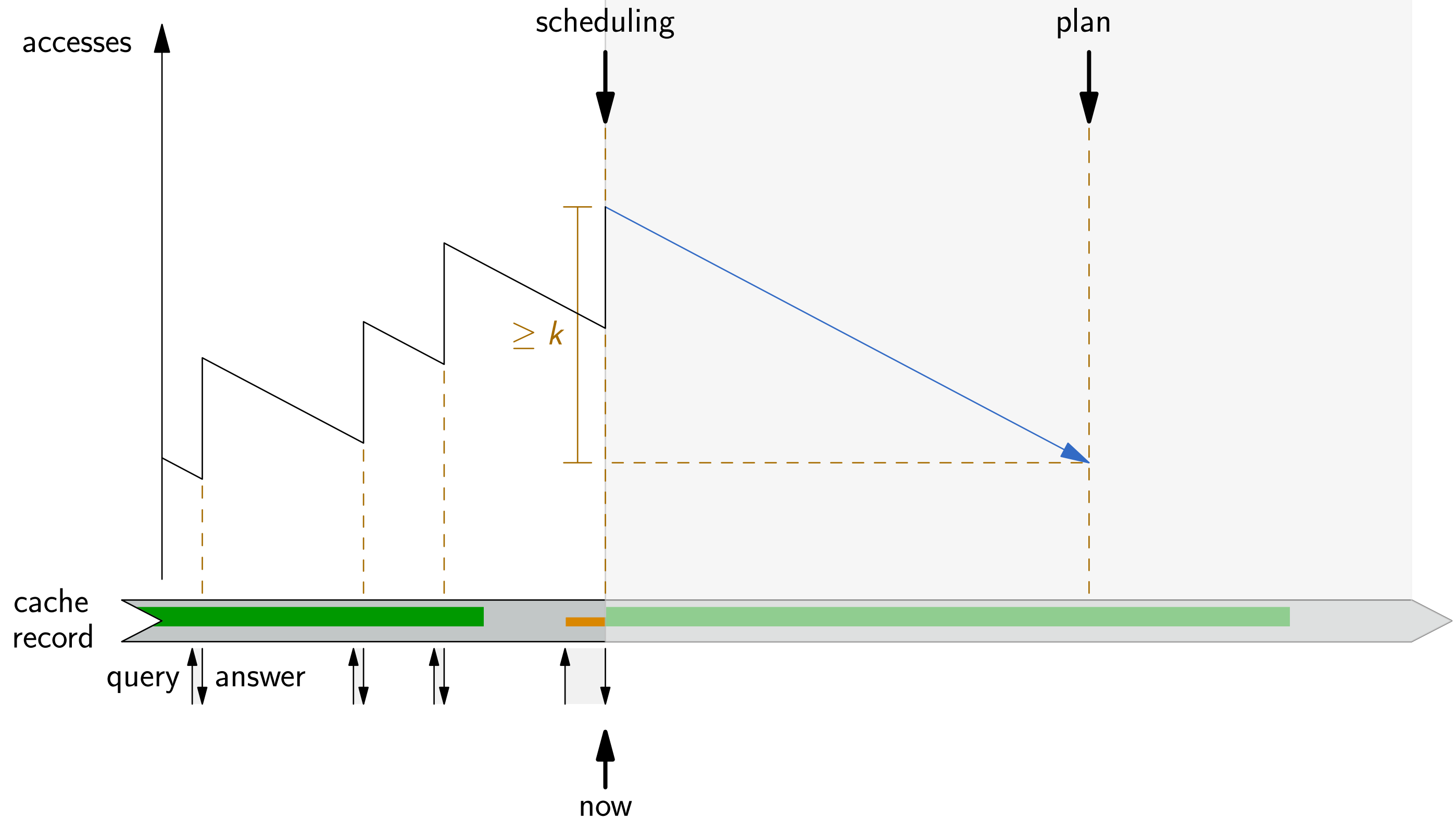
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL



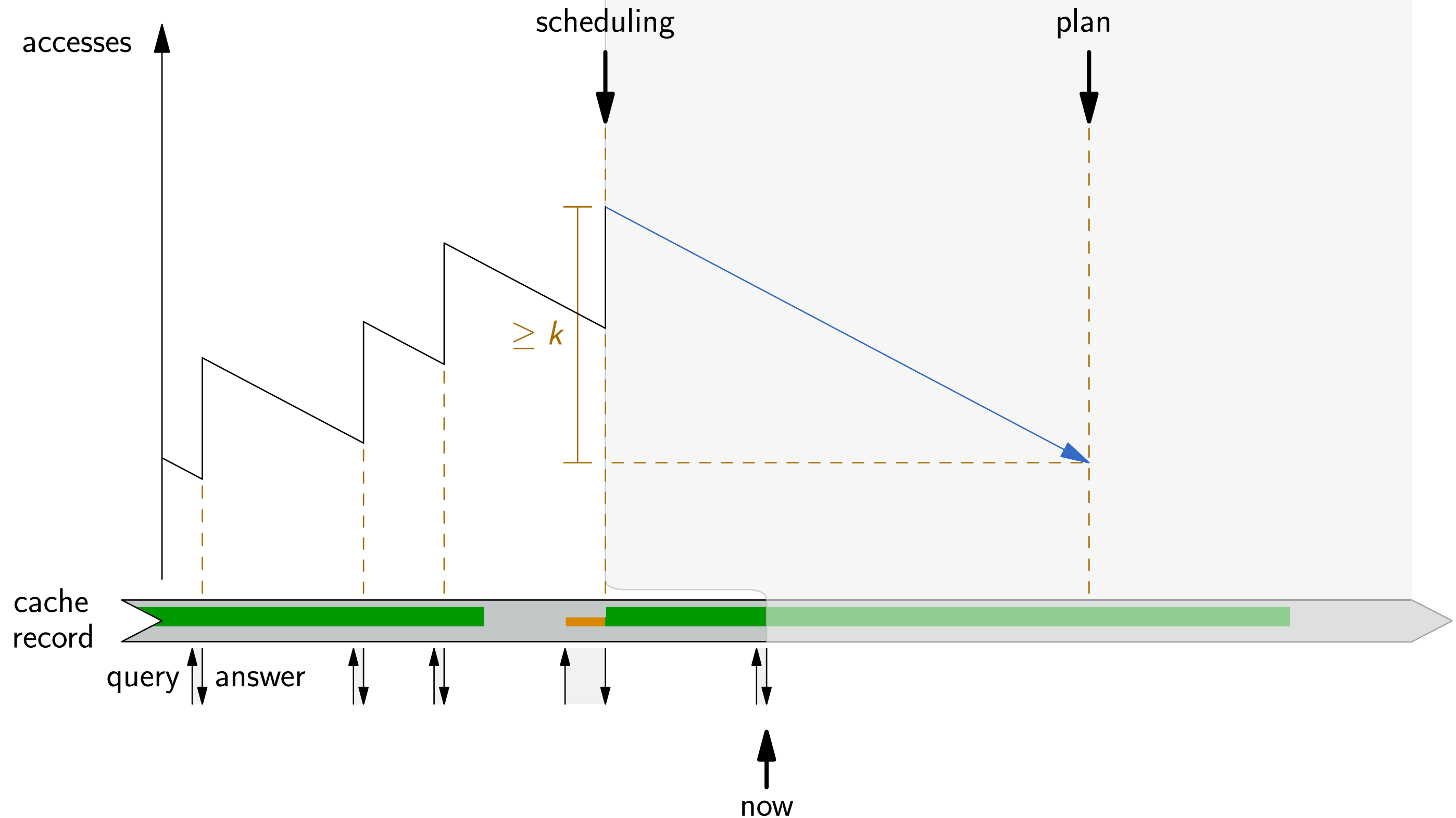
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses



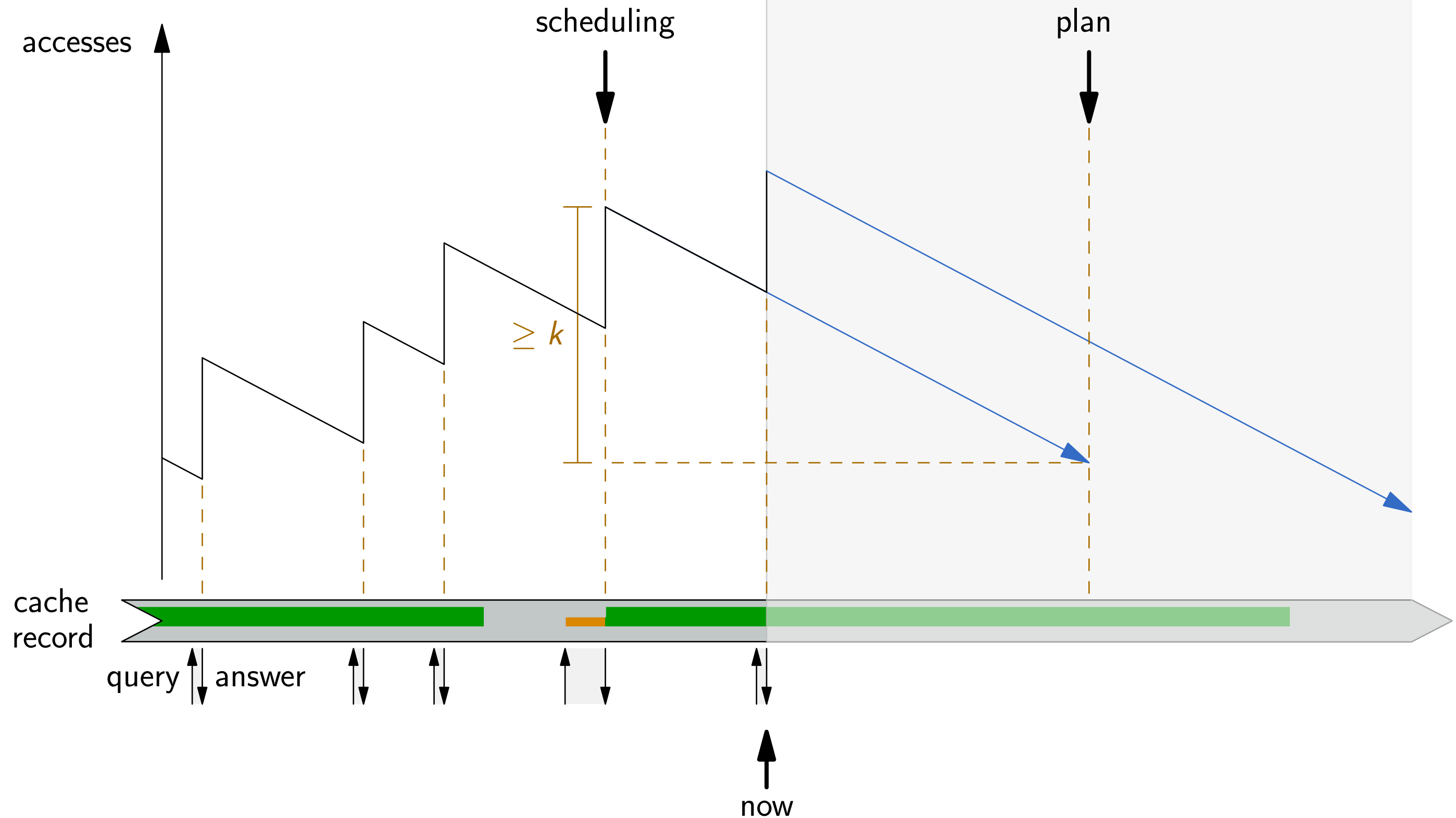
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses



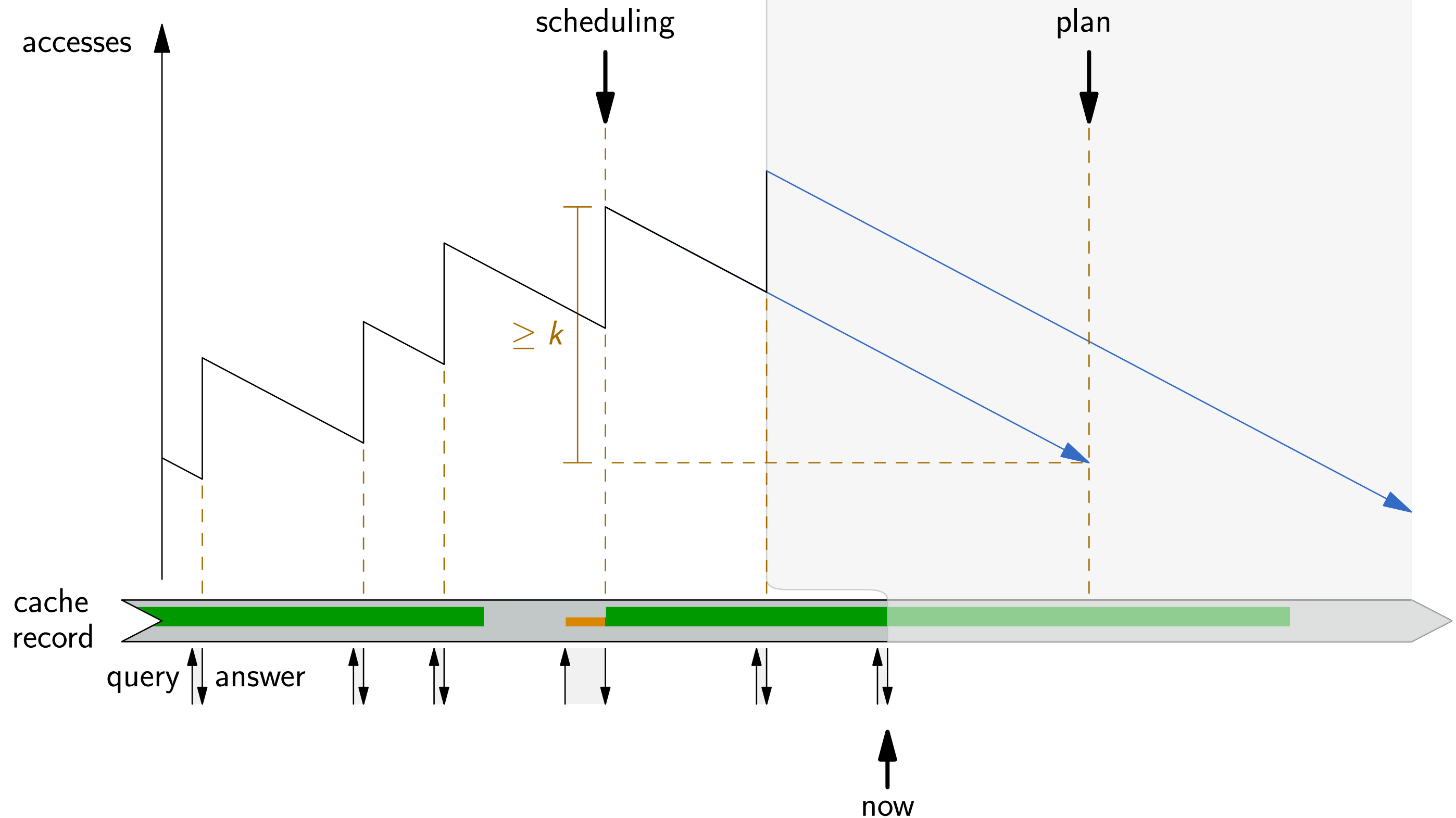
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses



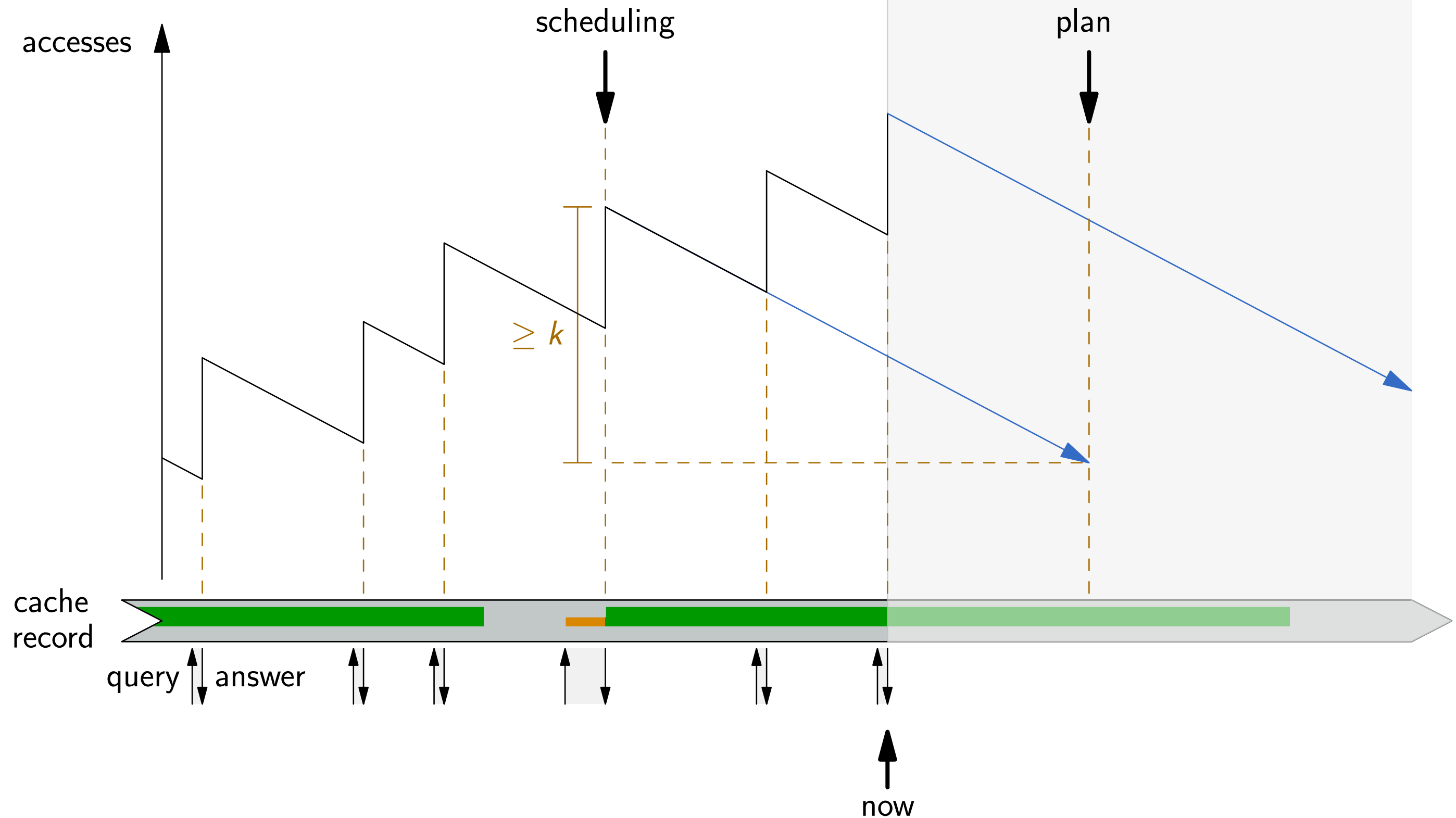
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses



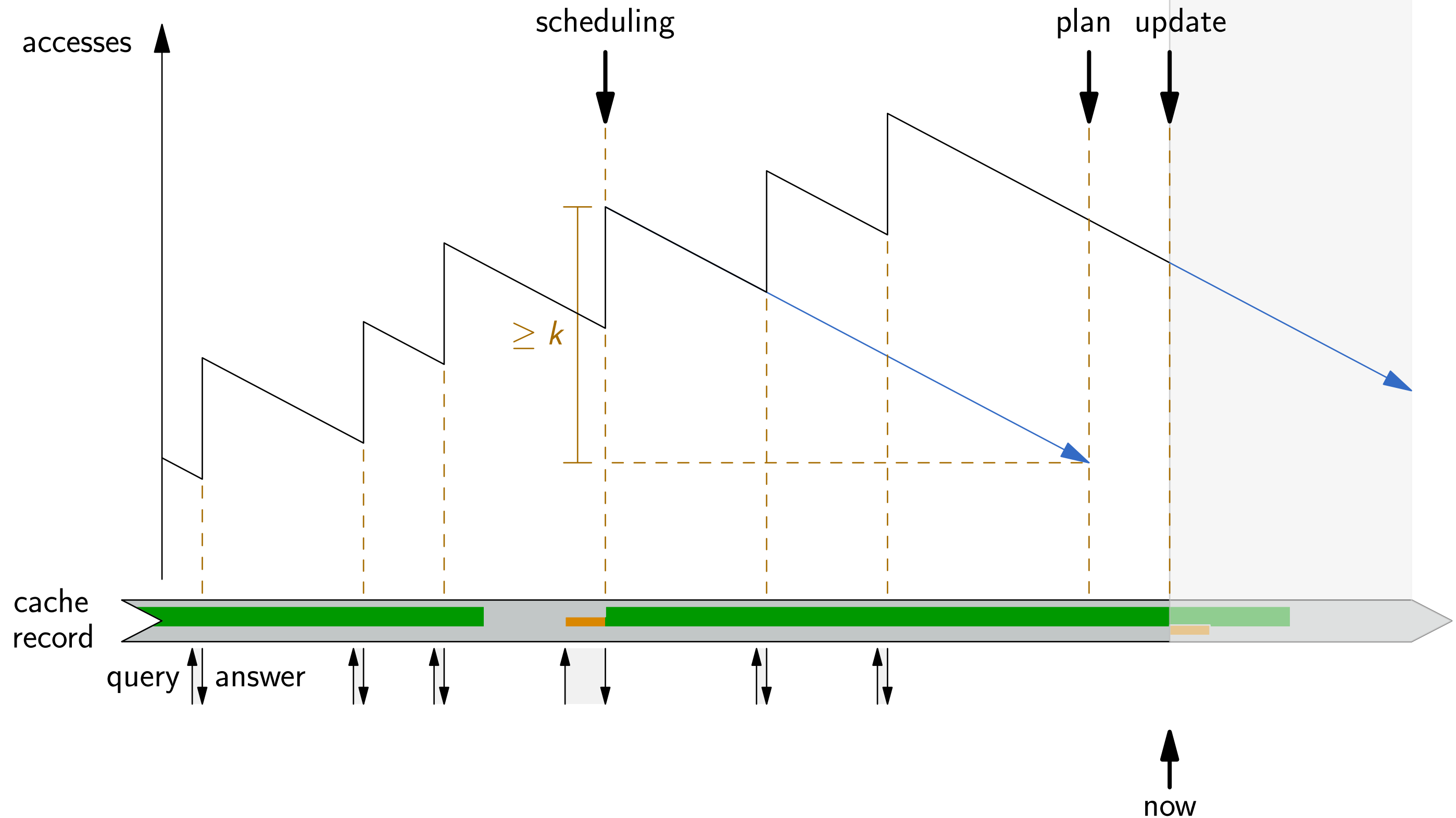
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses



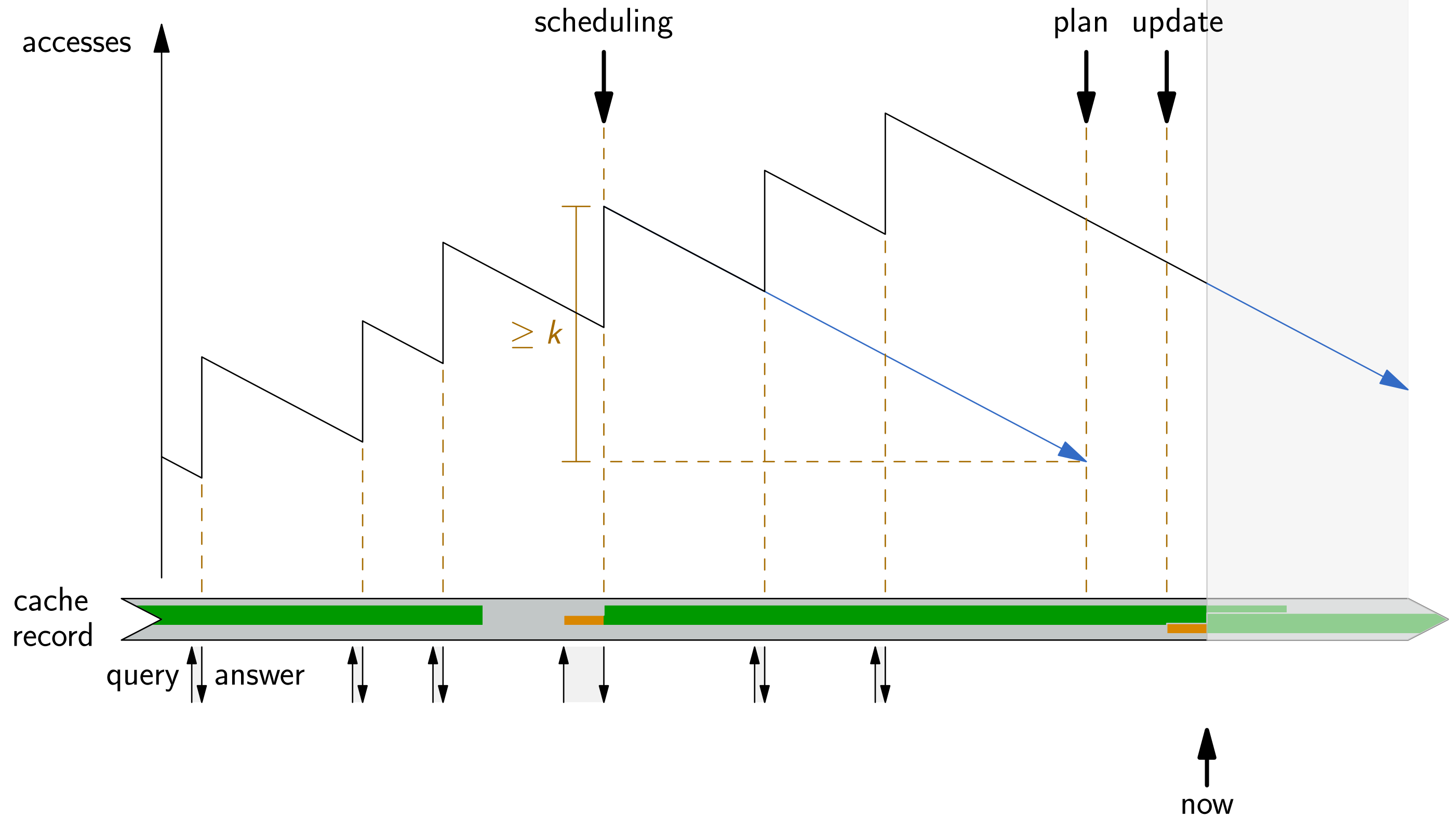
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle



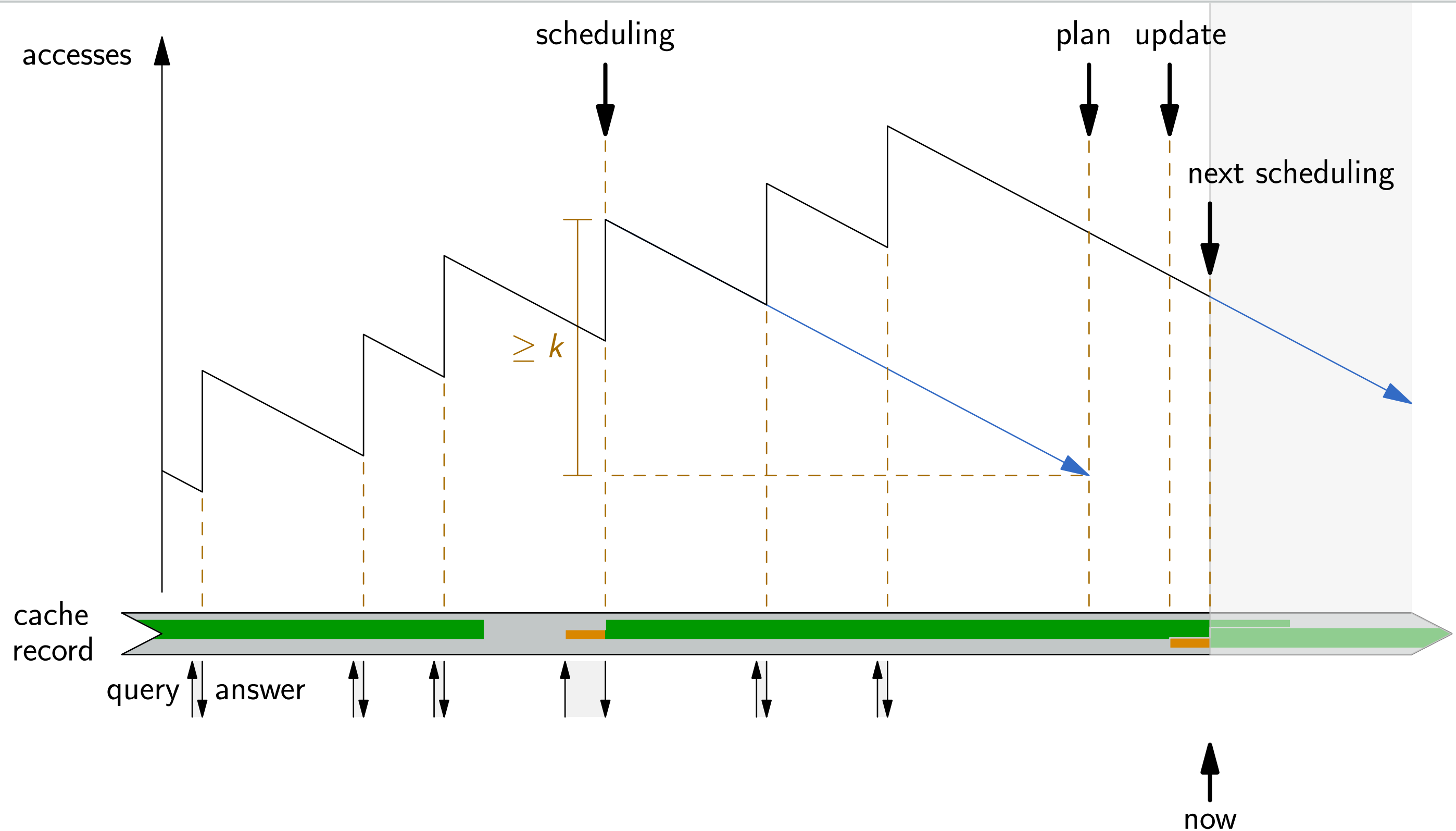
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle



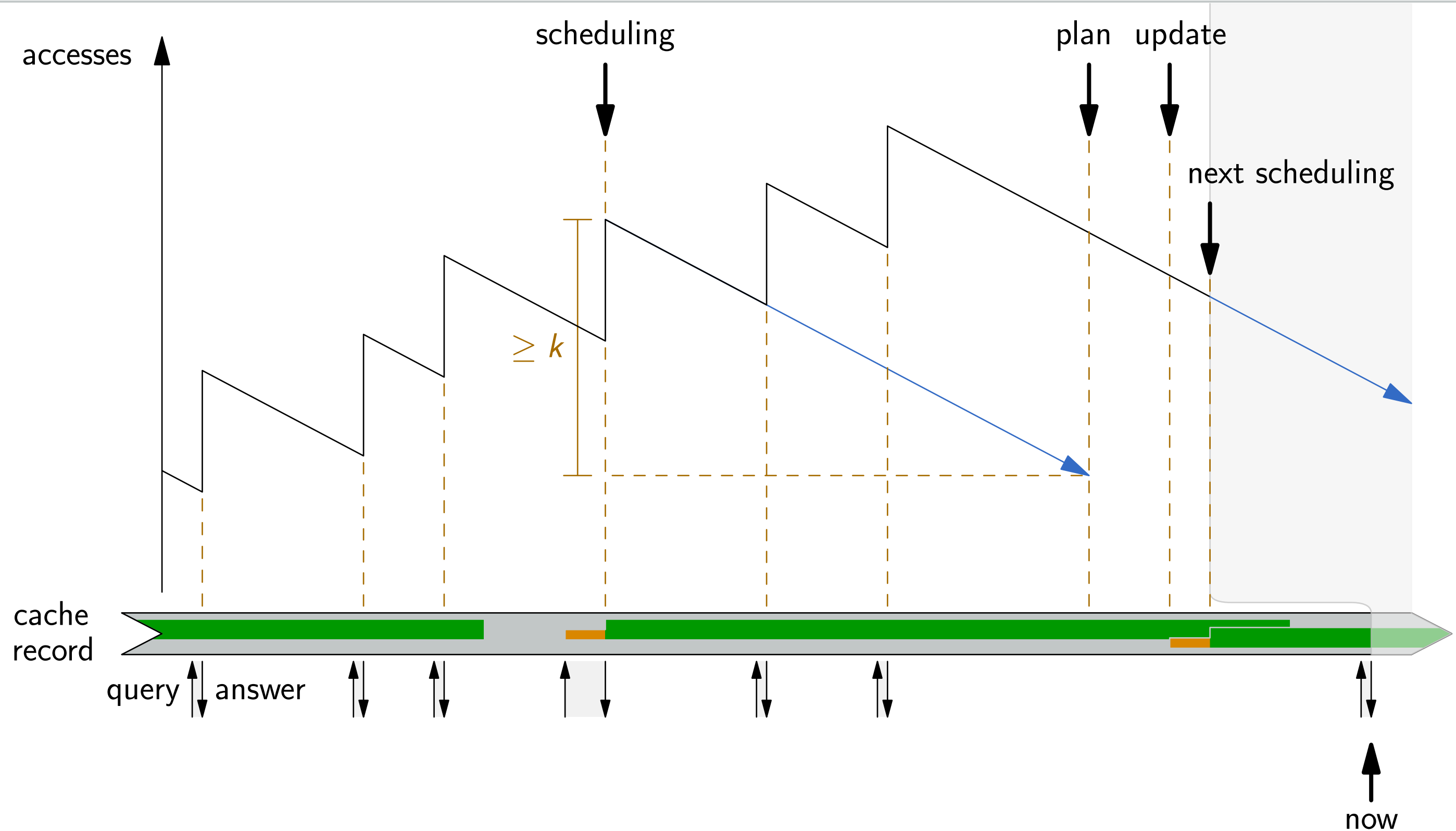
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle



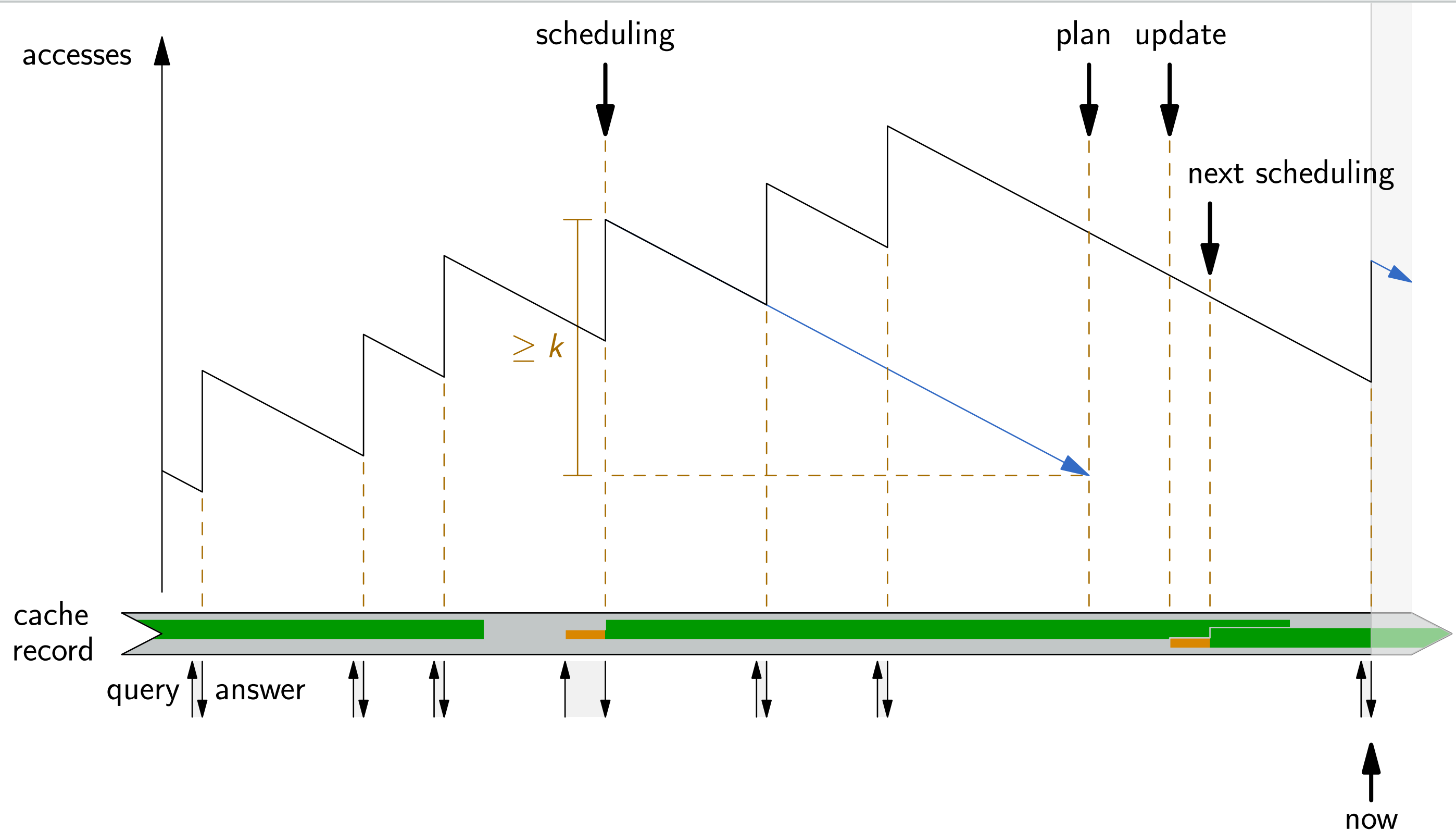
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle



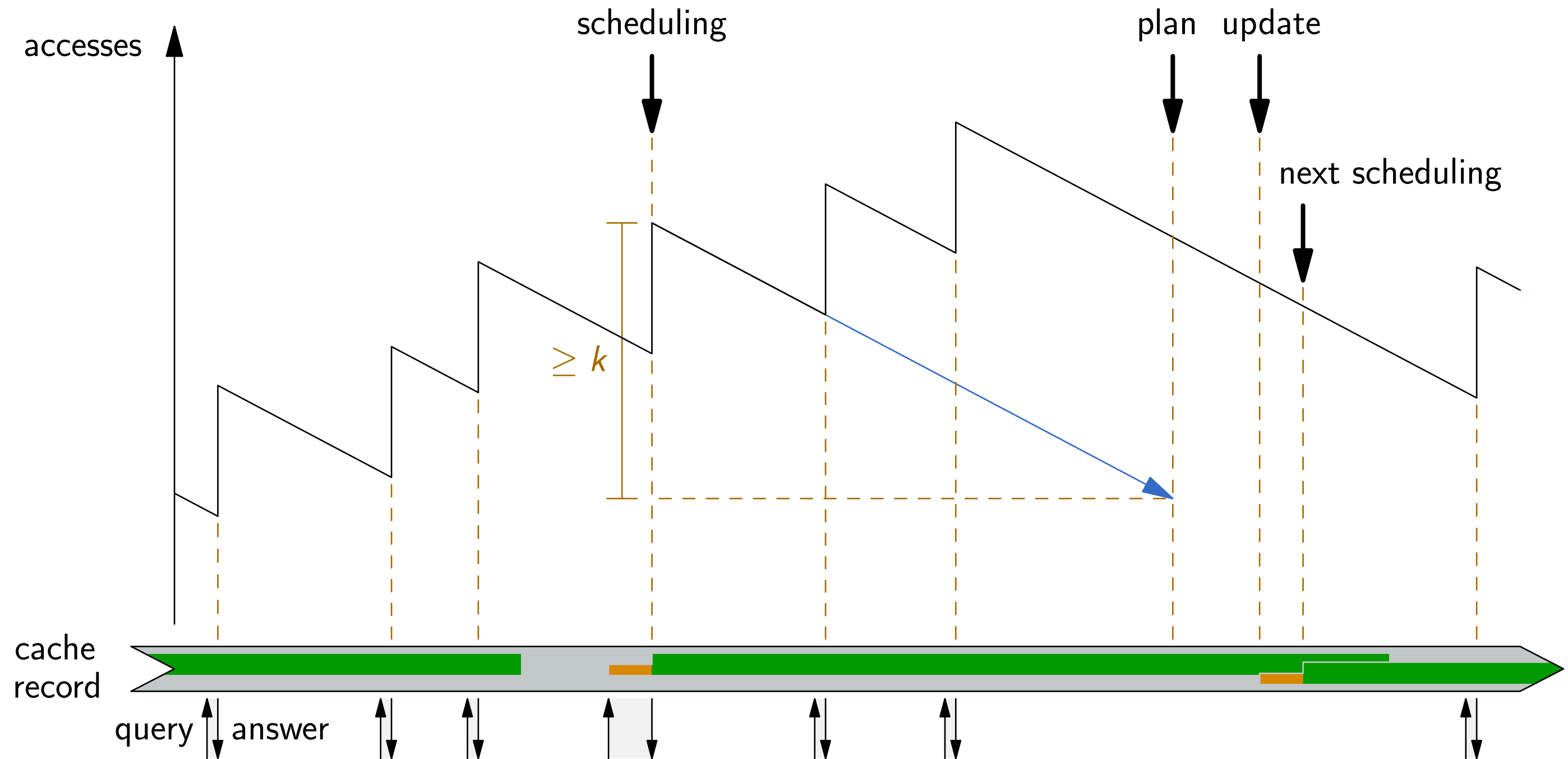
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle



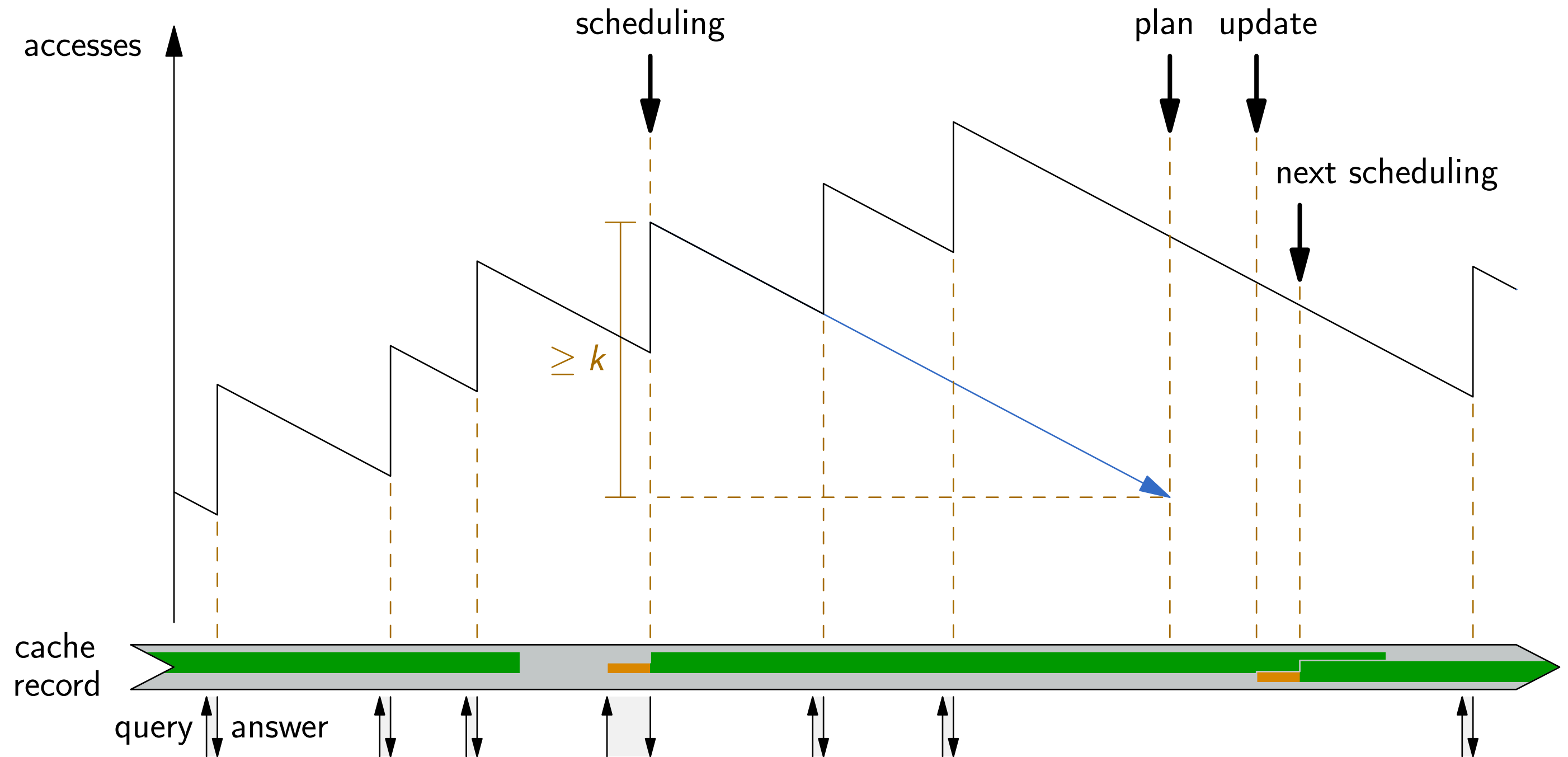
Prefetching expired records

- scheduling:
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- real update:
 - after plan when idle



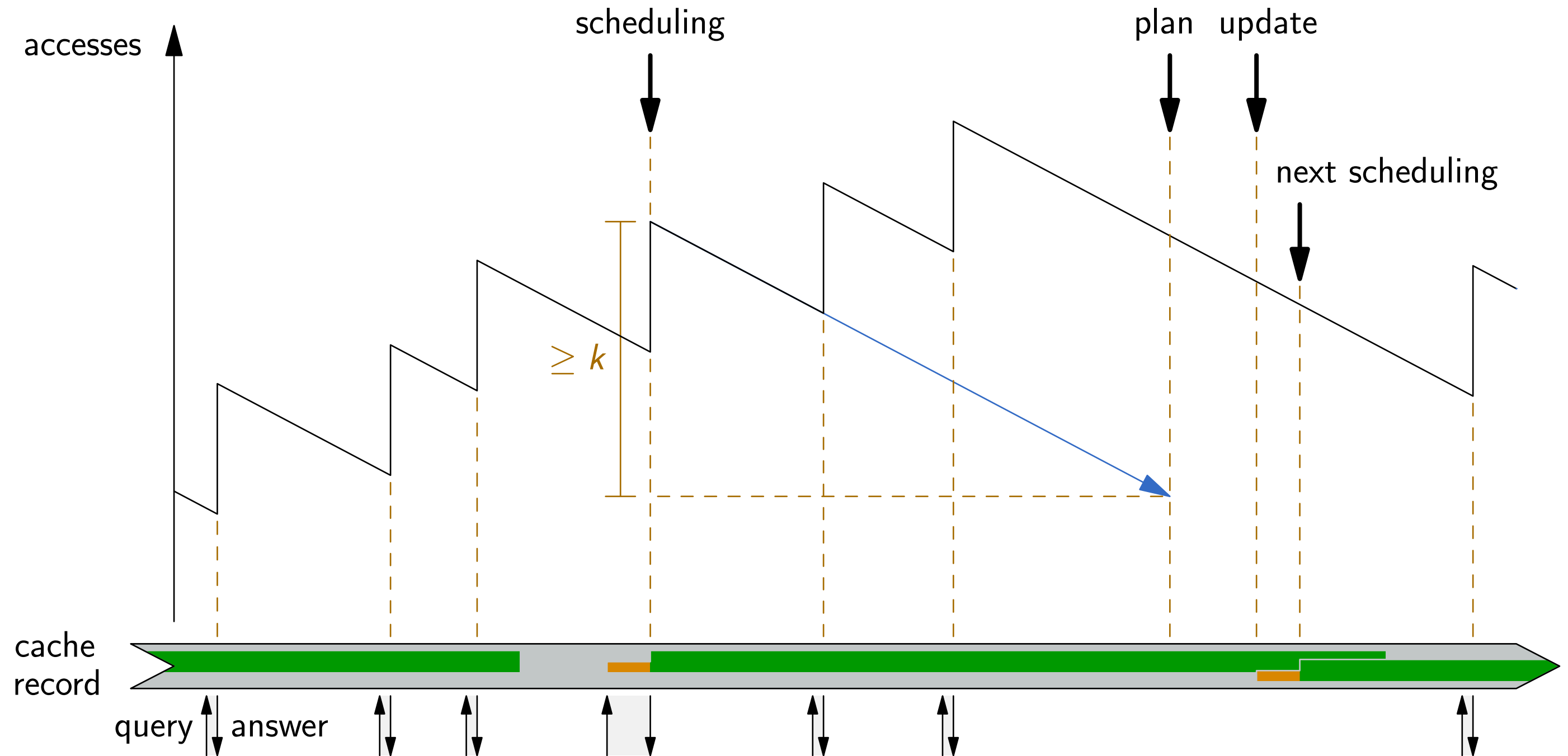
Prefetching expired records

- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle
- **GC interactions:**
 - only existing
 - cancel during GC

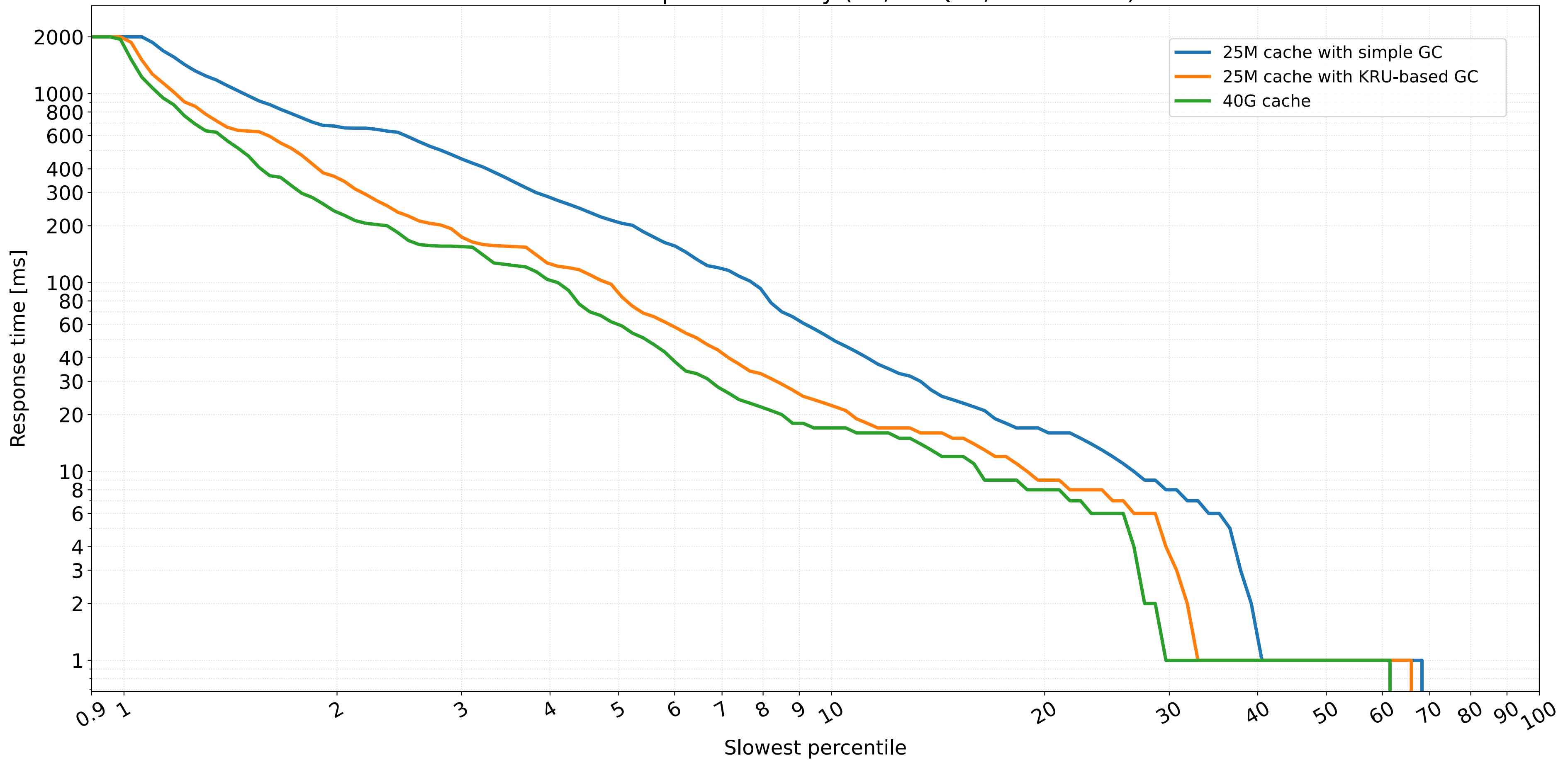


Prefetching expired records

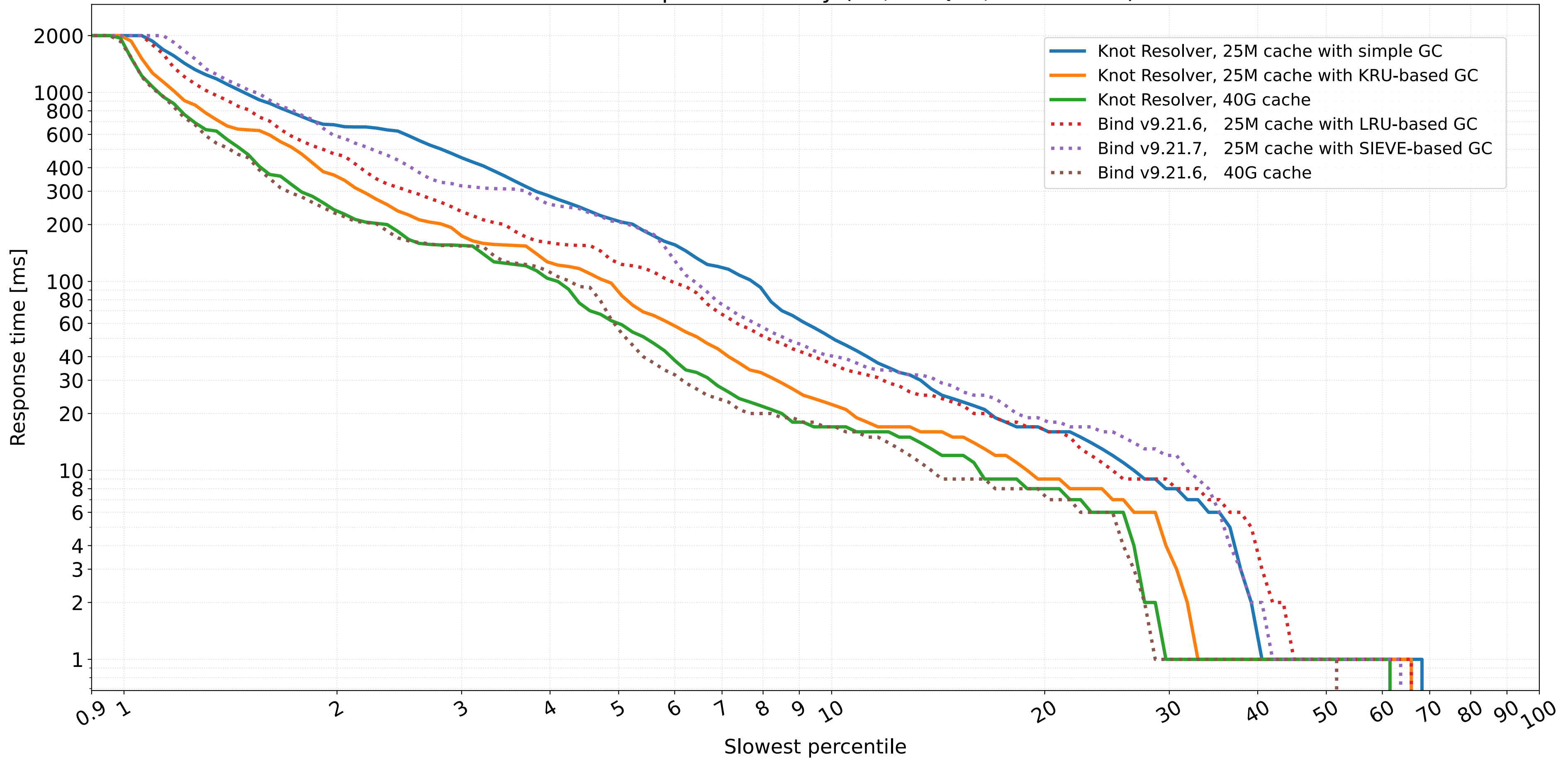
- **scheduling:**
 - during access
 - plan: 5s to TTL
 - if max 1 prefetch per k past accesses
- **real update:**
 - after plan when idle
- **GC interactions:**
 - only existing
 - cancel during GC
- **GC log: size of**
 - removed
 - expired unscheduled
 - valid unscheduled
 - scheduled
 - canceled



Measured Response Latency (9h, 2k QPS, 15 threads)



Measured Response Latency (9h, 2k QPS, 15 threads)



Summary

- original GC
 - main criteria
 - validity
 - domain level
- KRU-based GC
 - stats use 6–13%
 - main criteria
 - validity
 - accesses
 - soon released
- measurements ⇒
- prefetch
 - criteria
 - max 1 prefetch per k accesses
 - design proposal

