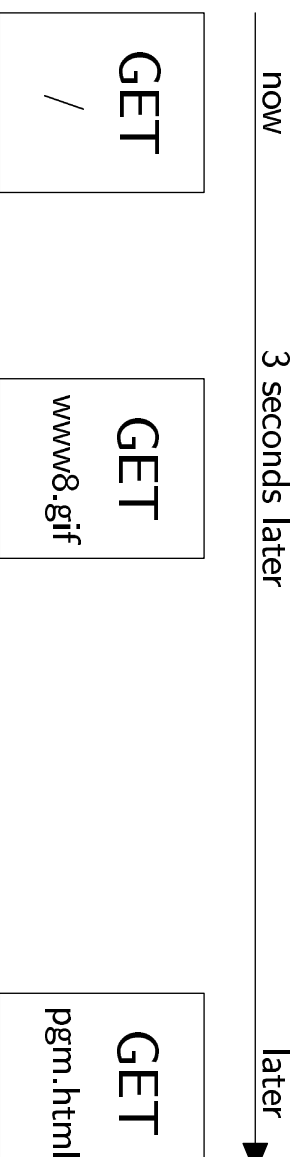


Managing TCP Connections Under Persistent HTTP

Edith Cohen Haim Kaplan Jeffrey D. Oldham
AT&T Labs—Research Stanford University
{edith,hkl}@research.att.com oldham@cs.stanford.edu

13 May 1999

An Example of Persistent HTTP Connections



HTTP/1.0

open

close

open

close

open

close

HTTP/1.0

open

close

open

close

open

close

HTTP/1.1
Persistent
Connections

open

GET
/

idle

GET
www8.gif

idle

close

open

GET
pgm.html

idle

close

HTTP/1.0

open

close

open

close

open

close

HTTP/1.1 Persistent Connections

open

GET

idle

GET
www8.gif

idle

close

open

GET
pgm.html

idle

close

Optimal Persistent Connections

open

GET

idle

GET
www8.gif

close

open

GET
pgm.html

close

Persistent HTTP Advantages and Disadvantages

Advantages for User: faster response time

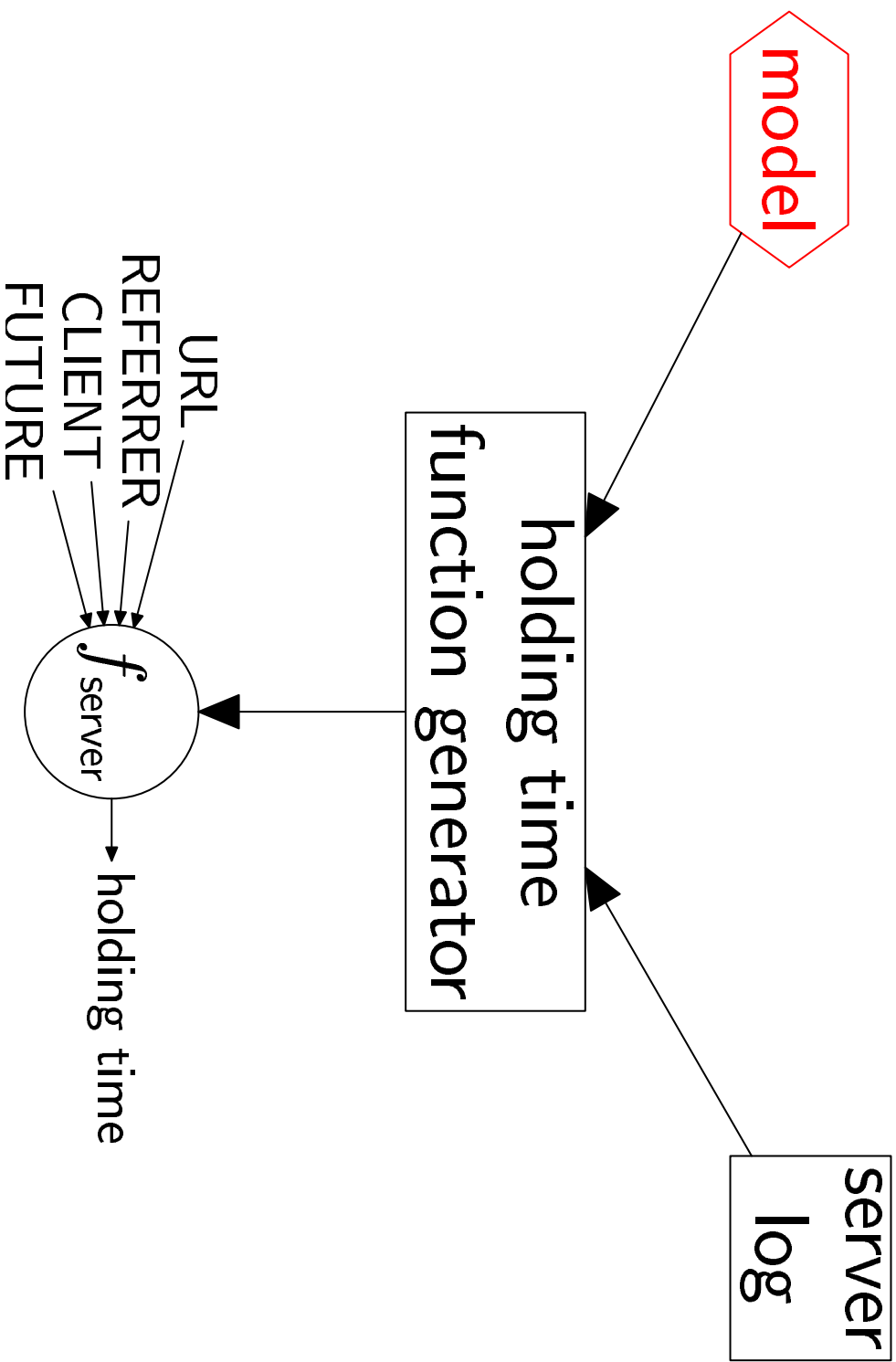
- fewer TCP handshakes
- can pipeline multiple requests
- TCP can better determine network congestion

Advantages for Server: less computation

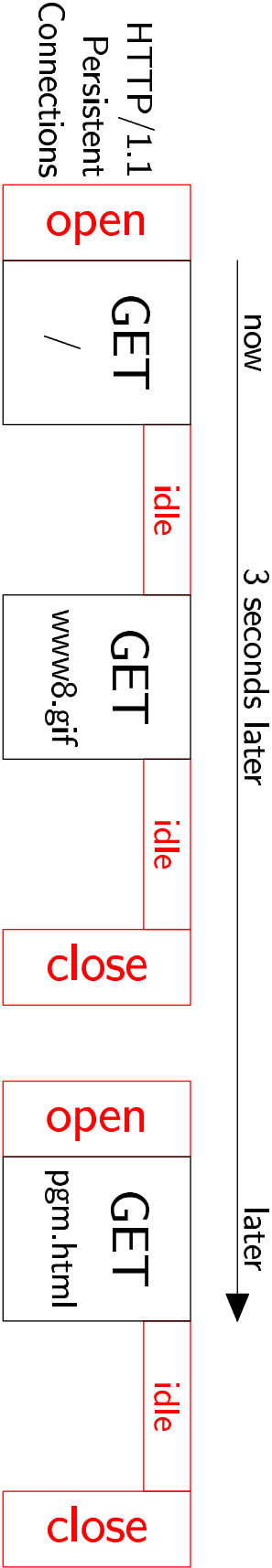
- less CPU time opening and closing TCP connections
- fewer simultaneous TCP control blocks

Disadvantages for Server: more idle open TCP connections

- TCP connections require memory
- many idle connections slow servers [Banga, Druschel, Mogul 1999]

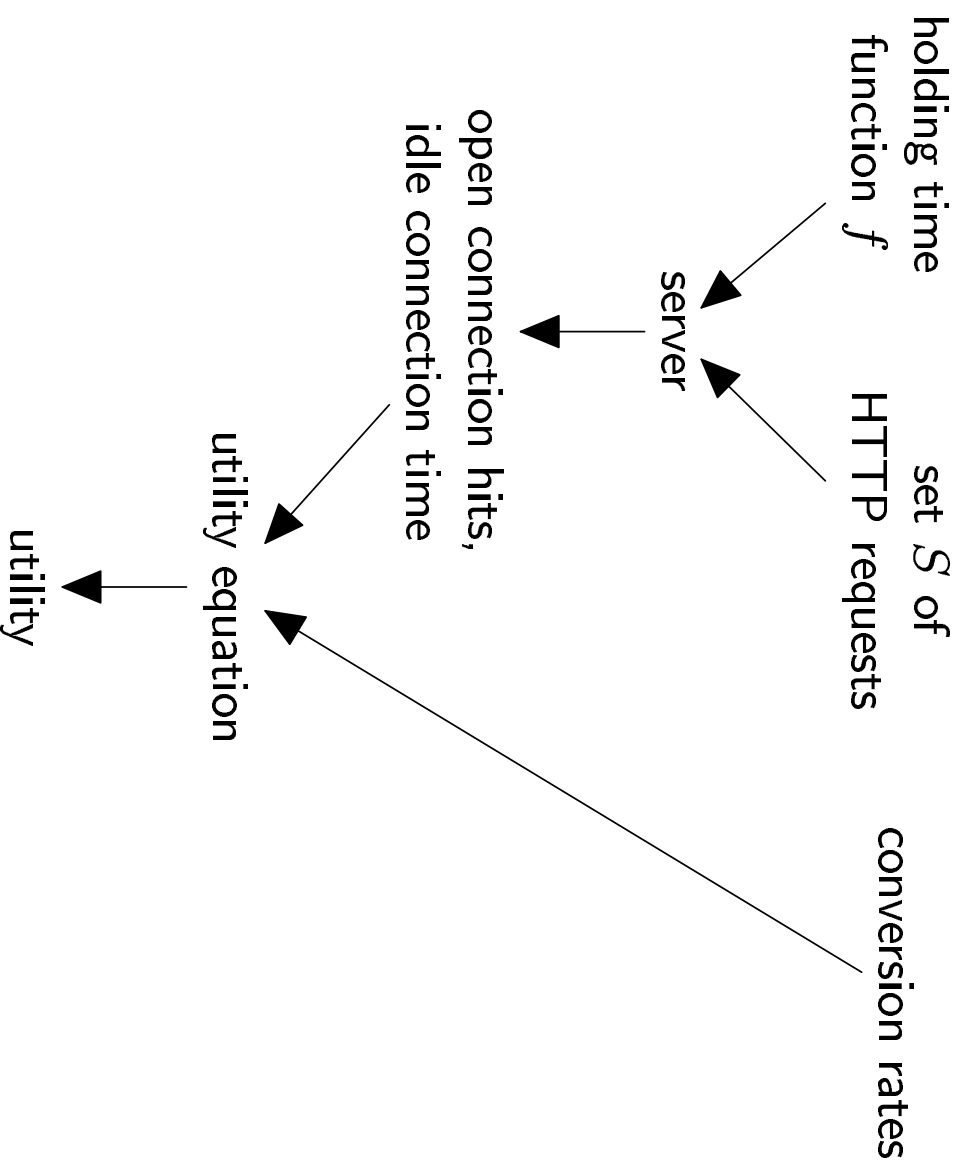


Holding Time Model



Utility Model

Goal: Maximize server's utility.



Utility Equation

utility = revenue – cost

revenue = hits $\cdot C_{h \rightarrow \$}$ + open-hits $\cdot C_{oh \rightarrow \$}$

cost = (hits – open-hits) \cdot TCP setup $\cdot C_{TCP \rightarrow \$}$ + idle-time $\cdot C_{it \rightarrow \$}$

Conversion Rates

$C_{h \rightarrow \$}$ = value of HTTP request

$C_{oh \rightarrow \$}$ = value of HTTP request to open connection

$C_{TCP \rightarrow \$}$ = cost of establishing TCP connection

$C_{it \rightarrow \$}$ = cost of idle connection

$$\text{utility} \approx C + C_o \cdot \text{open-hits} - C_i \cdot \text{idle-time}$$

Utility Equation

Which terms depend on holding time function?

utility = revenue – cost

revenue = hits · $C_{h \rightarrow \$}$ + open-hits · $C_{oh \rightarrow \$}$

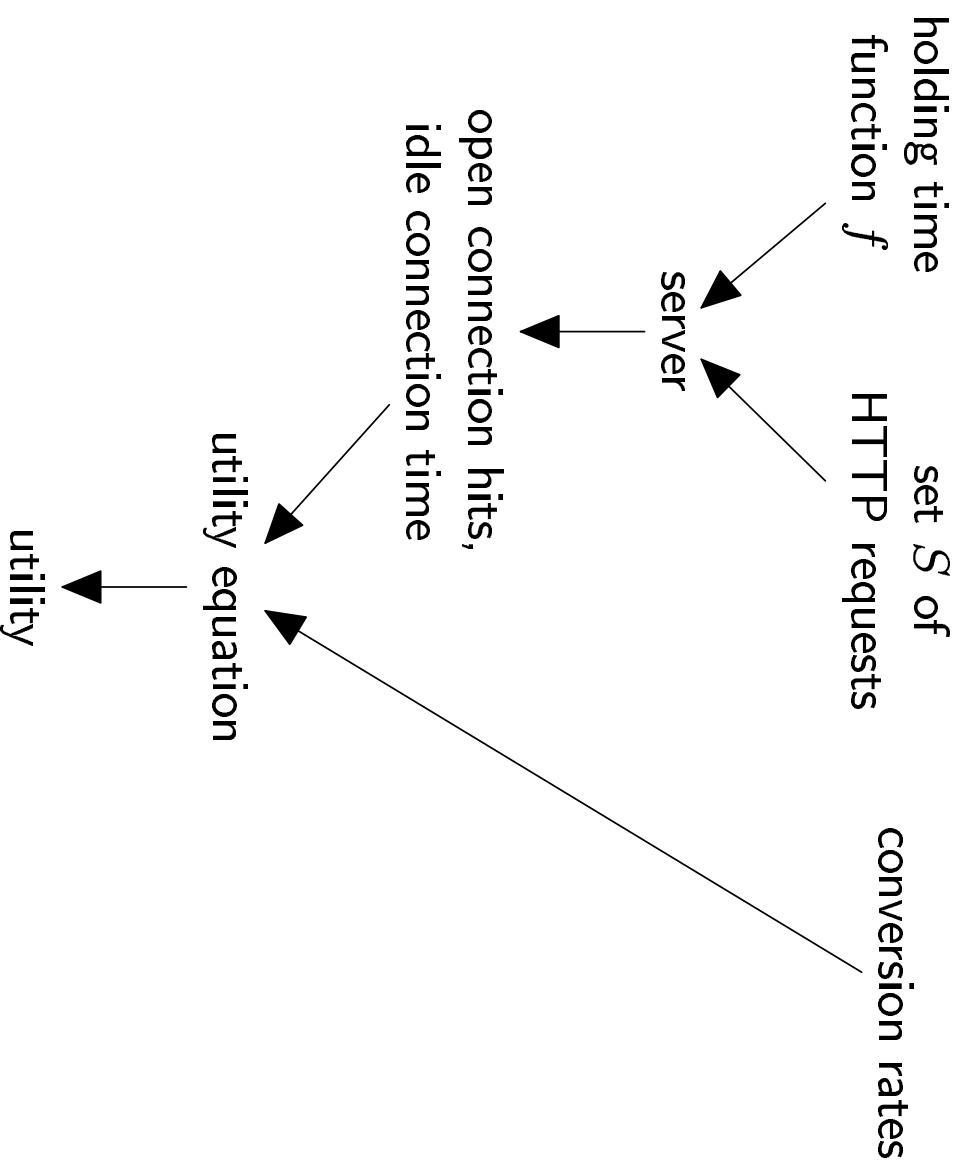
cost = (hits – open-hits) · TCP setup · $C_{TCP \rightarrow \$}$ + idle-time · $C_{it \rightarrow \$}$

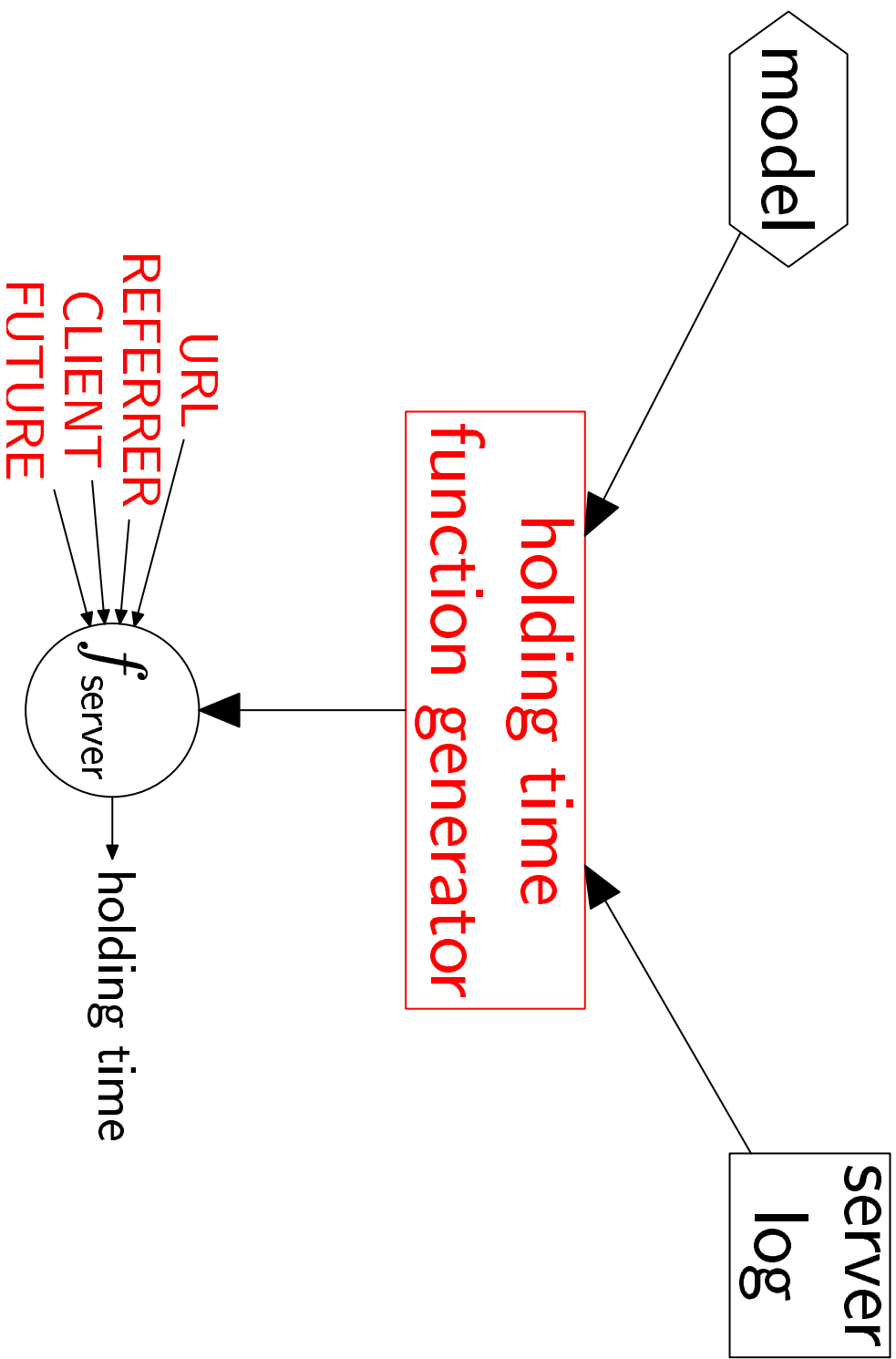
equivalent to

utility $\approx C + C_o \cdot \text{open-hits} - C_i \cdot \text{idle-time}$

Utility Model

Goal: Maximize server's utility.





Basing Holding Times on HTTP Request Attributes

Current Page has Embedded Images

- probable request within few seconds
- short holding time

Client Predictability:

- request stock market information every 15 minutes
- request Toronto weather information

Types of Attribute-Based Functions

Relevant Parameters

RESOURCE: use current URL

REFERER: use source for current URL's address

CLIENT: use current client

FUTURE: use knowledge of future requests

OPT: uses FUTURE

- closes connection if idle time more costly than next open-hit
- best possible function
- unimplementable

FIXED: ignores all parameters

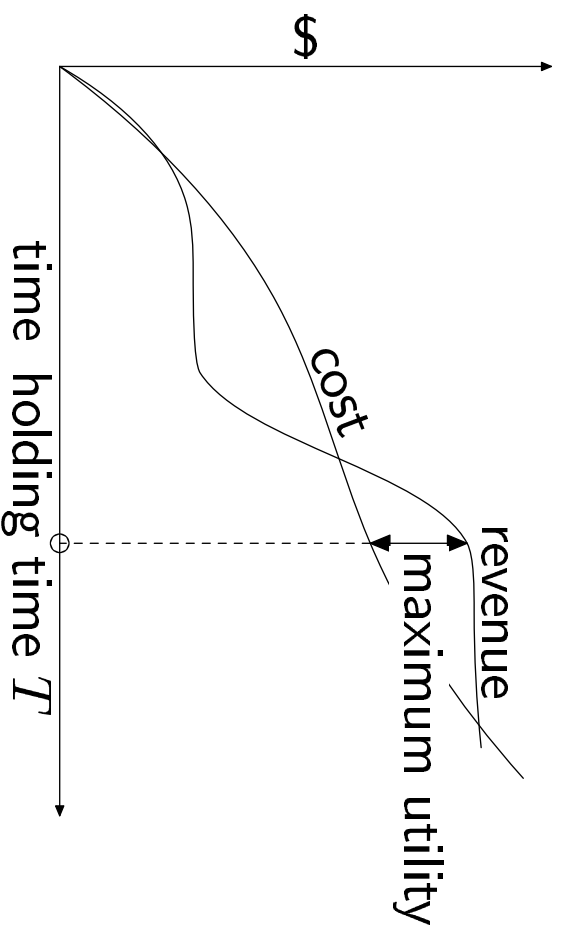
- every holding time equals t seconds
- simple to implement
- HTTP/1.1 suggests this function
- Apache v.1.3 uses 15 seconds

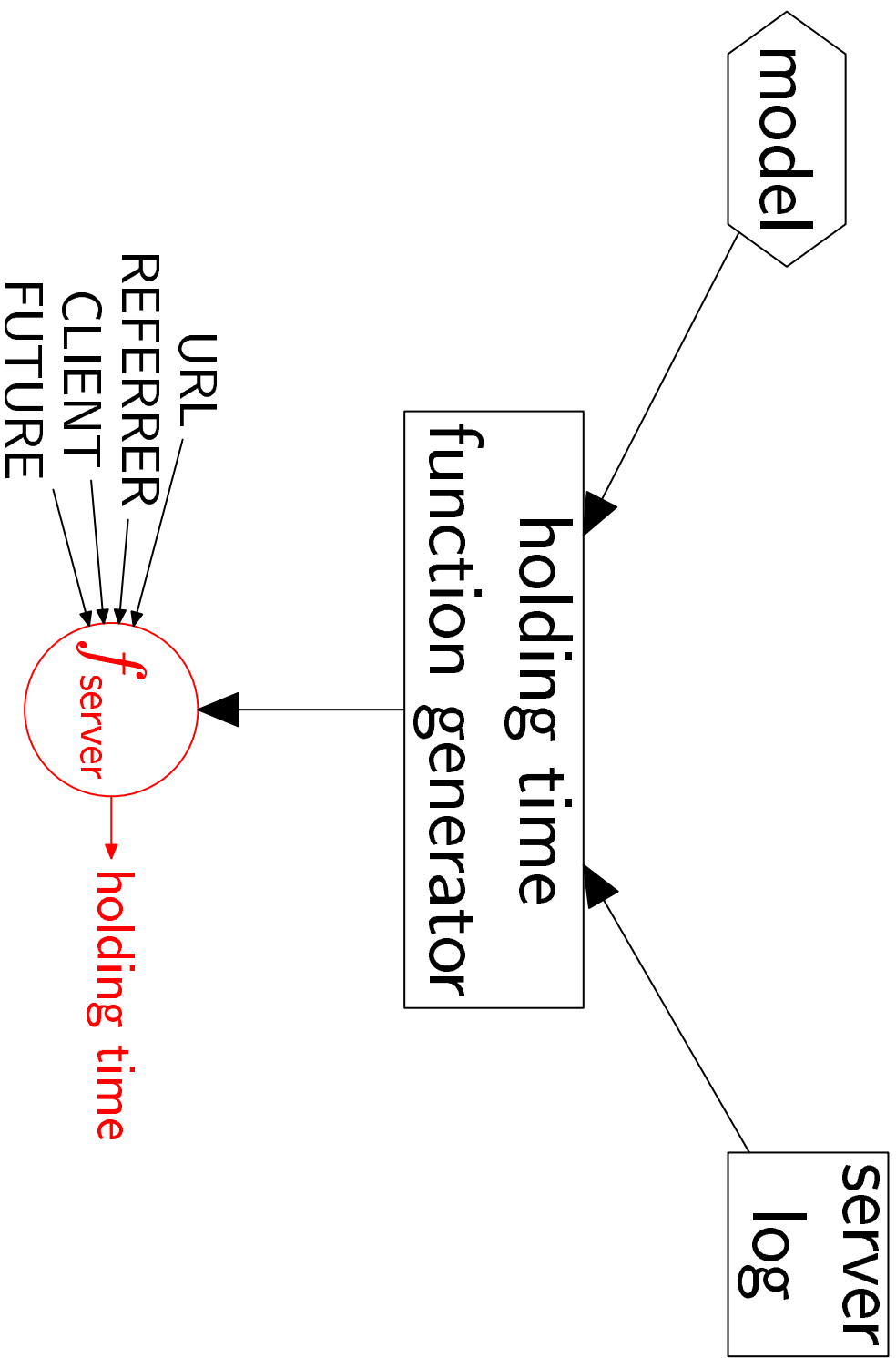
Computing Most Beneficial RESOURCE

$$\text{utility} \approx C + C_o \cdot \text{open-hits} - C_i \cdot \text{idle-time}$$

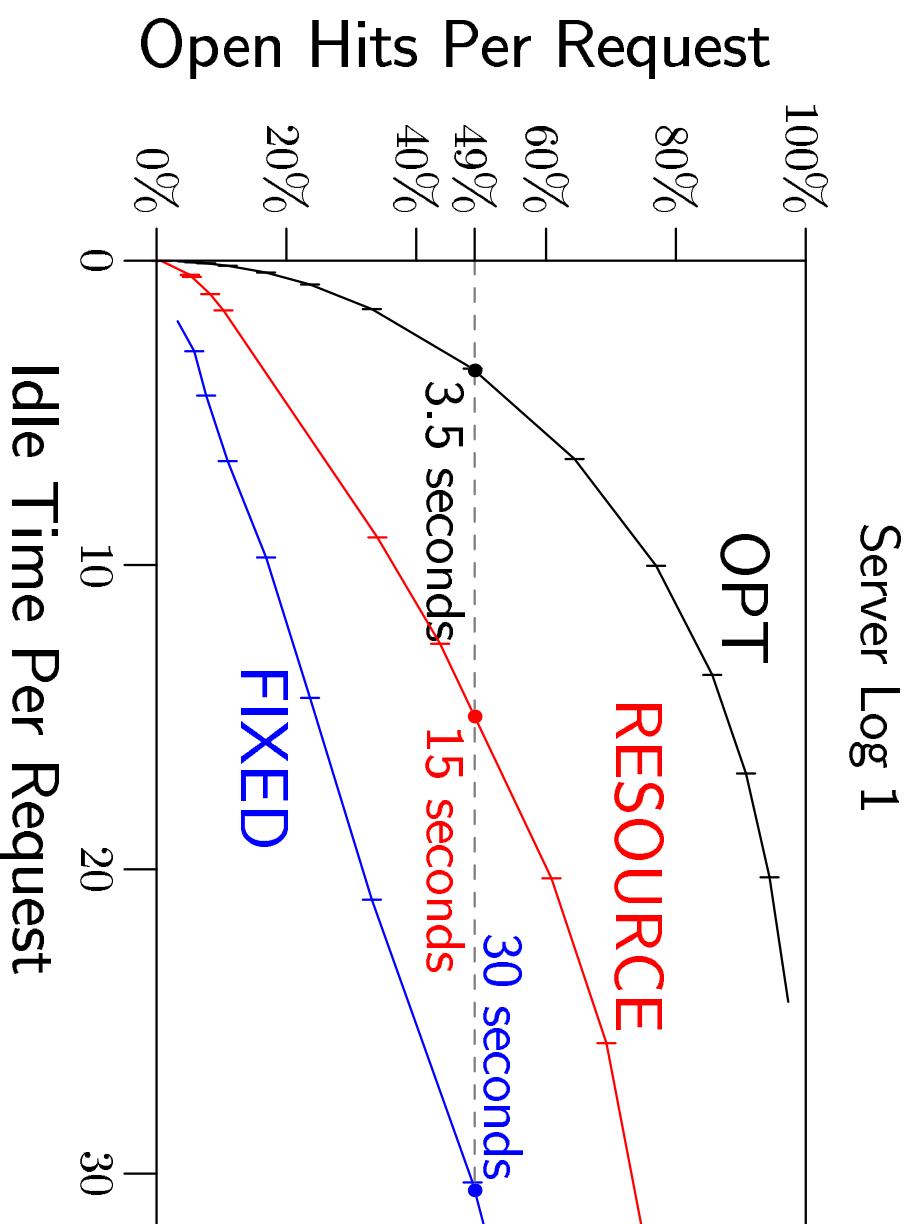
Server logs: future requests may reflect past requests

Compute utility: for each resource



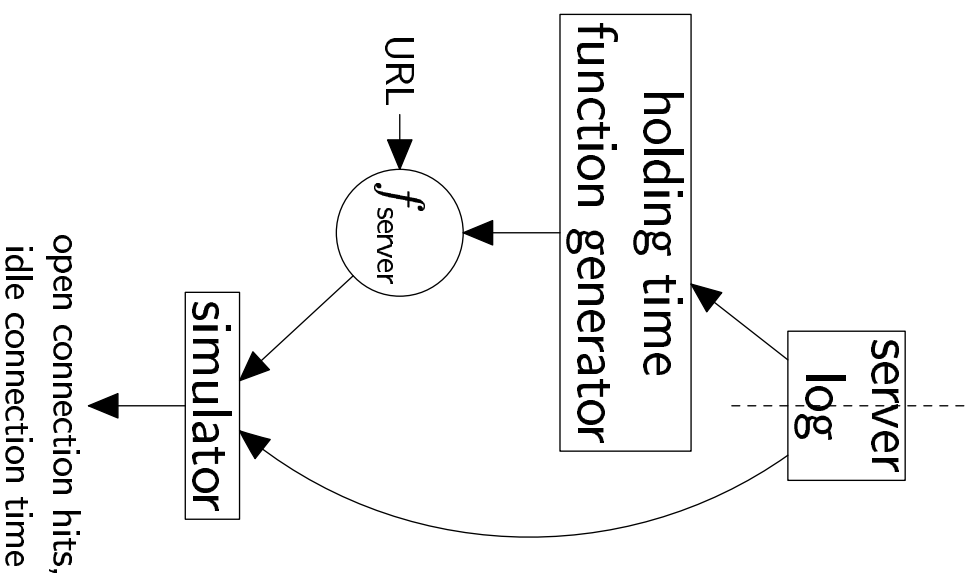


Predicting Using RESOURCE

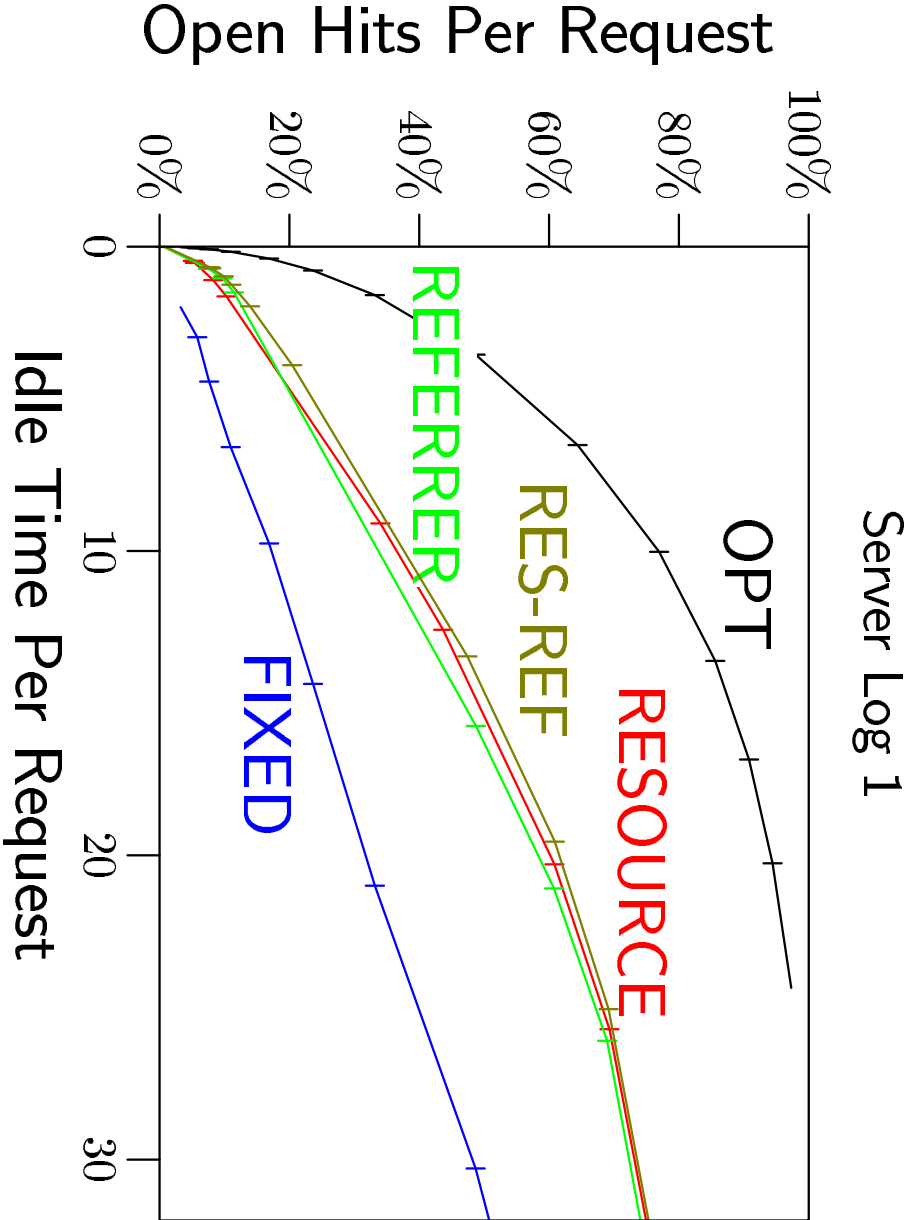


$$\frac{\text{utility}}{\text{connection}} = C_o \frac{\text{open-hits}}{\text{connection}} - C_i \frac{\text{idle-time}}{\text{connection}}$$

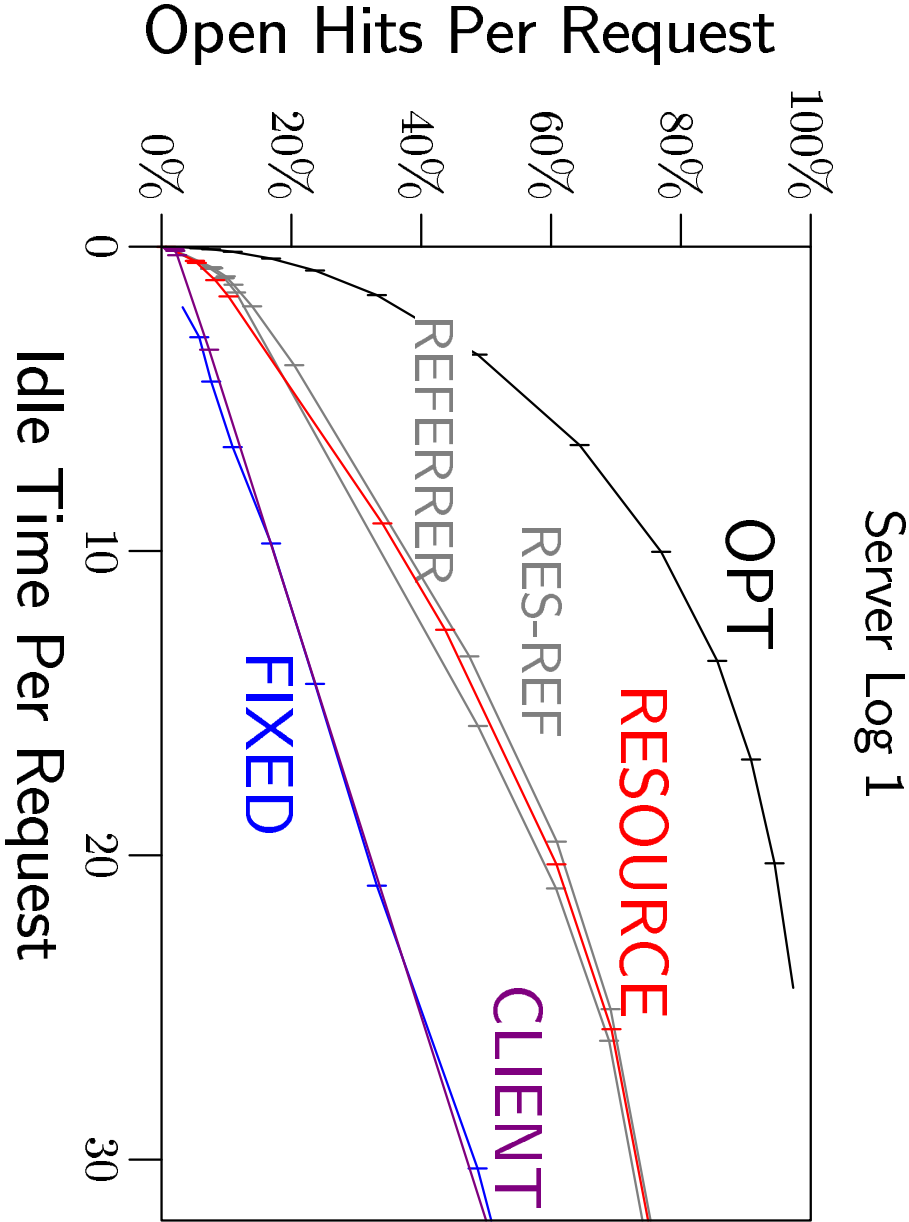
Experimental Method



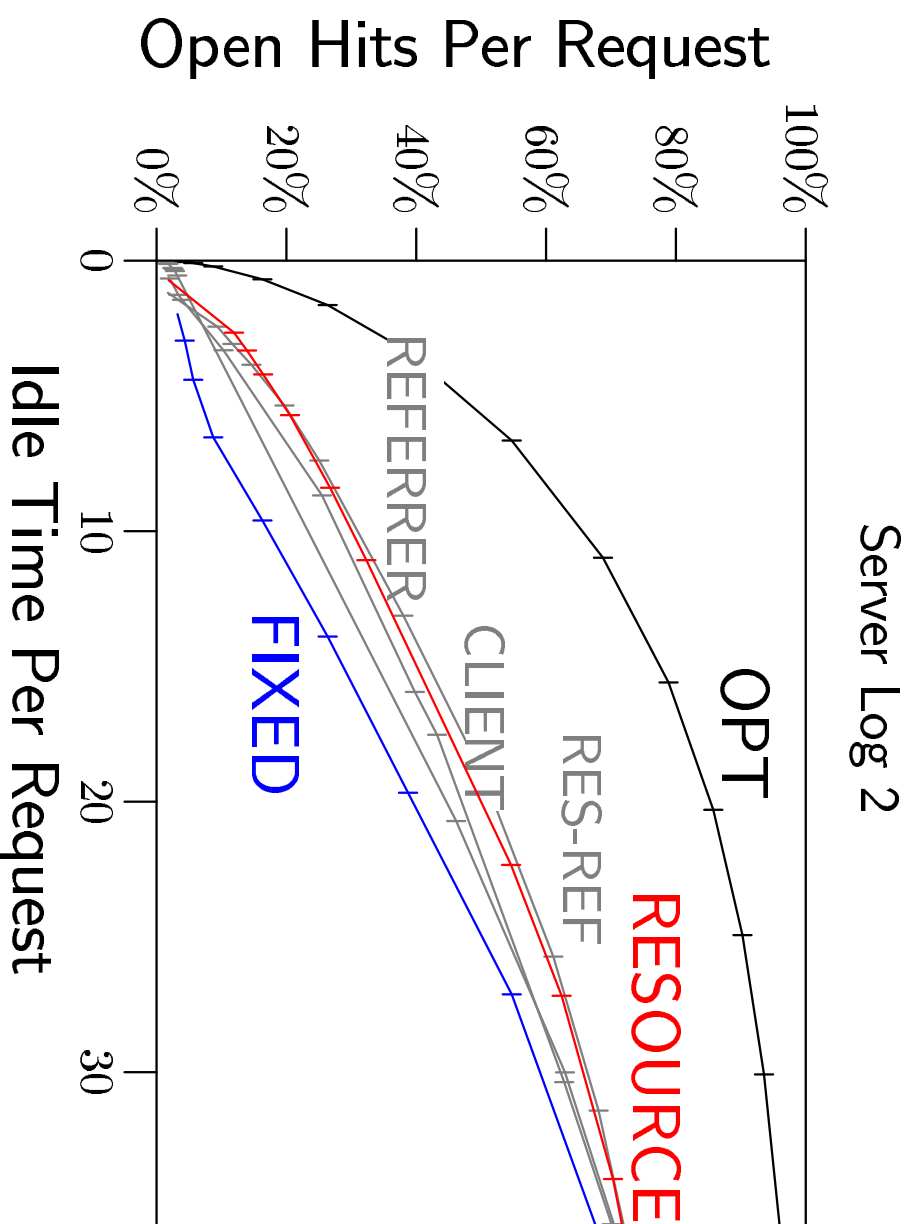
Comparing RESOURCE, REFERRER, RES-REF

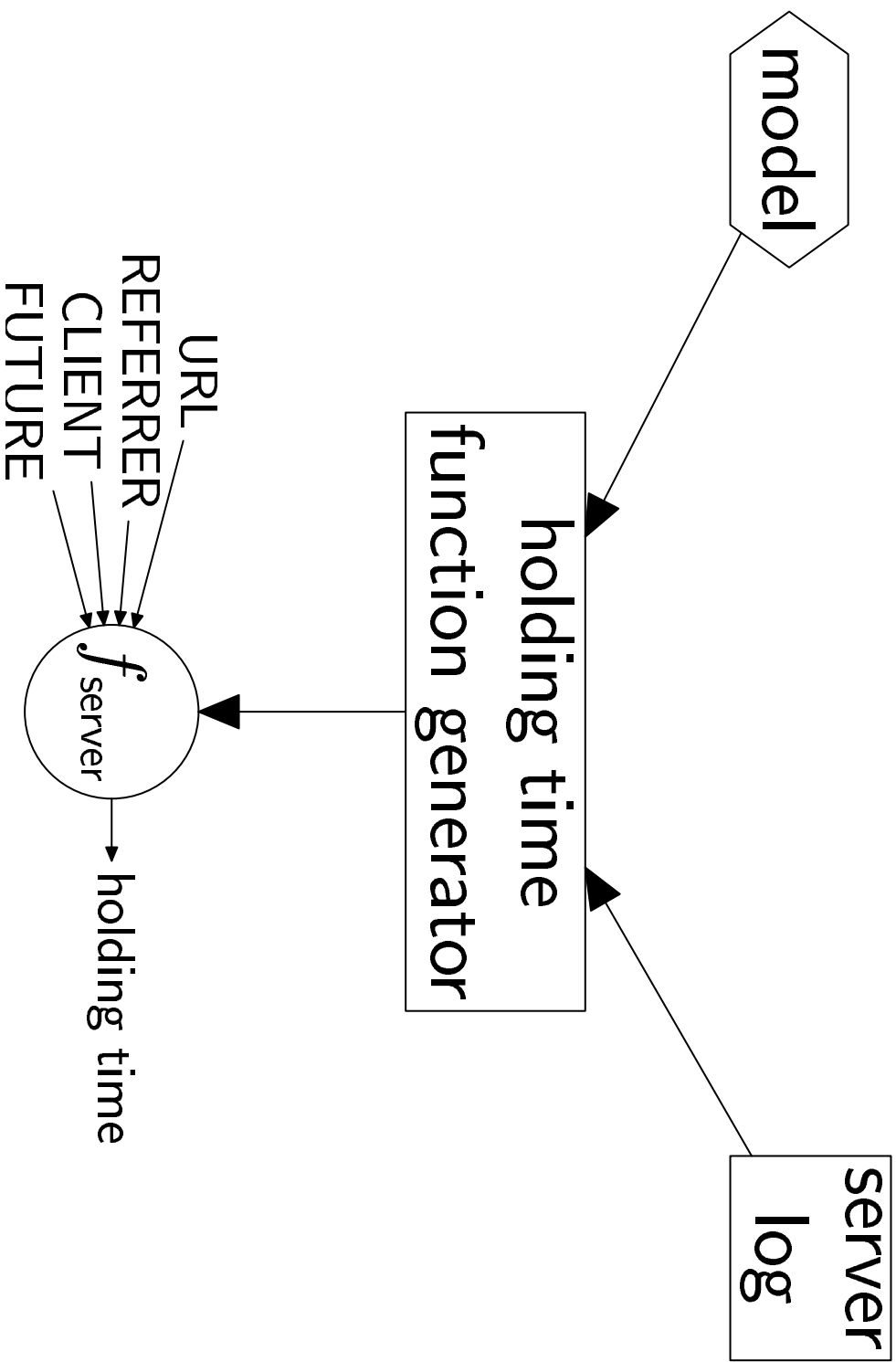


Predicting Using CLIENT



RESOURCE Reliable Attribute





Conclusions

Use talk outline slide.

- Introduced model for evaluating persistent HTTP connections.
- Presented technique to produce beneficial holding time functions.
- Experimentally showed our holding time functions beneficial.
 - reduce idle connection time up to 50%
- Experimentally evaluated relevance of attributes to holding times.

Future Work

- are conversion rates constant?
- holding times' effects on number of hits?
- effects of clients' closing connections?
- connection management functions for proxies