

TCP ex Machina: Computer-Generated Congestion Control

Keith Winstein and Hari Balakrishnan

MIT Computer Science and Artificial Intelligence Laboratory

<http://web.mit.edu/remy>

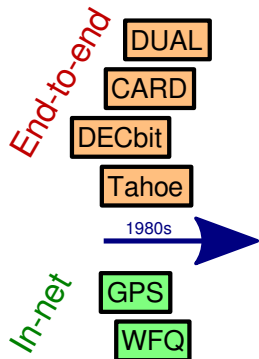
August 14, 2013



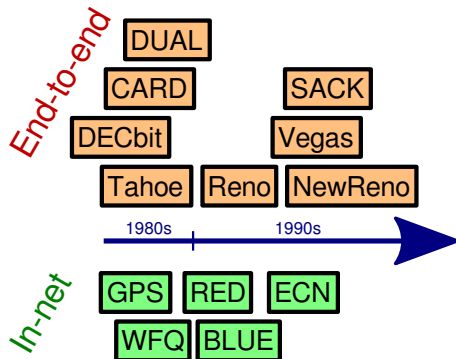
Congestion control!

- ▶ Prevents congestion collapse
- ▶ Allocates network resources among users
- ▶ Can be purely end-to-end or not

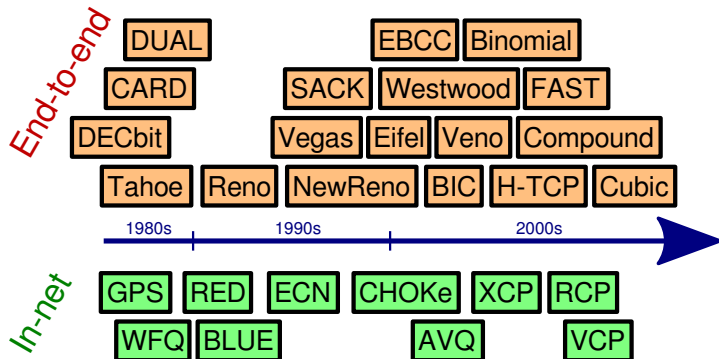
The march of congestion control mechanisms



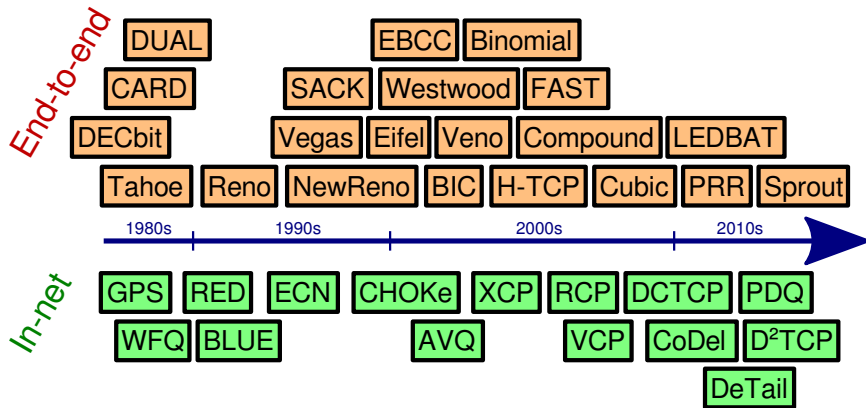
The march of congestion control mechanisms



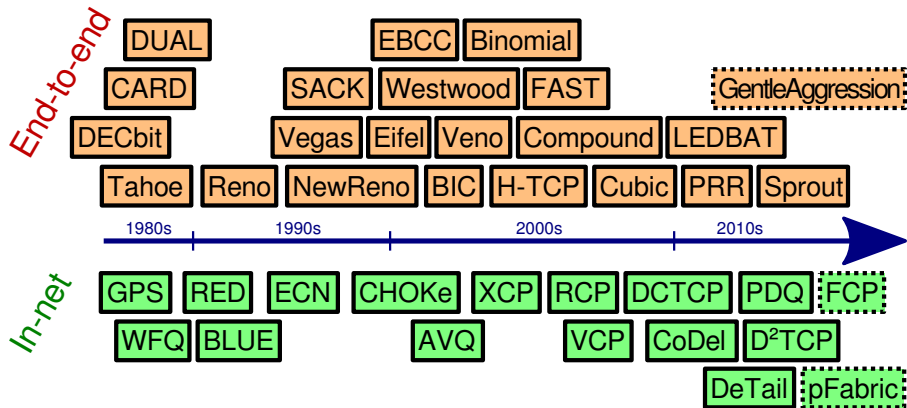
The march of congestion control mechanisms



The march of congestion control mechanisms



The march of congestion control mechanisms



Our work

If congestion control is the answer,
what's the question?

Our work

If congestion control is the answer,
what's the question?

Are there better answers?

Rational choice of scheme is challenging

Cubic vs. **Compound**

- ▶ Different goals?
- ▶ Different assumptions about network?
- ▶ One scheme just plain better?

Networks constrained by a fuzzy idea of TCP's assumptions

- ▶ Mask stochastic loss
- ▶ Bufferbloat
- ▶ Mask out-of-order delivery
- ▶ No parallel/multipath routing

Advice for Internet Subnetwork Designers
(RFC 3819) is 21,000 words!

Apps hack around TCP

- ▶ Open lots of flows
- ▶ Goose slow start
- ▶ Add pacing
- ▶ Give up and do it yourself

Apps hack around TCP

- ▶ Open lots of flows
- ▶ Goose slow start
- ▶ Add pacing
- ▶ Give up and do it yourself



Apps hack around TCP

- ▶ Open lots of flows
- ▶ Goose slow start
- ▶ Add pacing
- ▶ Give up and do it yourself



Google MICROSOFT

Apps hack around TCP

- ▶ Open lots of flows
- ▶ Goose slow start
- ▶ Add pacing
- ▶ Give up and do it yourself



Google MICROSOFT

YouTube

Apps hack around TCP

- ▶ Open lots of flows
- ▶ Goose slow start
- ▶ Add pacing
- ▶ Give up and do it yourself



Google MICROSOFT

YouTube

Chrome (QUIC)

BitTorrent (μ TP)

Mosh (SSP)

Aspera (fasp)

Better: free the network to evolve

Transport layer should adapt to **whatever**:

- ▶ network does
- ▶ application wants

What we built

Remy: a program that generates congestion-control schemes offline

Input:

- ▶ Prior assumptions (what network may do)
- ▶ Goal (what app wants)

Output: CC algorithm for a TCP sender (RemyCC)

Time: a few hours

Cost: \$5–\$10 on Amazon EC²

The basic question of congestion control

At this moment, do I:

- ▶ send a packet
- ▶ not send a packet?

Objectives of congestion control

Maximize

- ▶ $\sum_i \log [\text{throughput}_i]$ (proportionally fair throughput)

Objectives of congestion control

Maximize

- ▶ $\sum_i \log [\text{throughput}_i]$ (proportionally fair throughput)

- ▶ $\sum_i \log \left[\frac{\text{throughput}_i}{\text{delay}_i} \right]$ (proportionally fair throughput/delay)

Objectives of congestion control

Maximize

- ▶ $\sum_i \log [\text{throughput}_i]$ (proportionally fair throughput)

- ▶ $\sum_i \log \left[\frac{\text{throughput}_i}{(\text{delay}_i)^\delta} \right]$ (proportionally fair throughput/delay)

Objectives of congestion control

Maximize

- ▶ $\sum_i \log [\text{throughput}_i]$ (proportionally fair throughput)

- ▶ $\sum_i \log \left[\frac{\text{throughput}_i}{(\text{delay}_i)^\delta} \right]$ (proportionally fair throughput/delay)

- ▶ $\min_i \text{throughput}_i$ (max-min throughput)

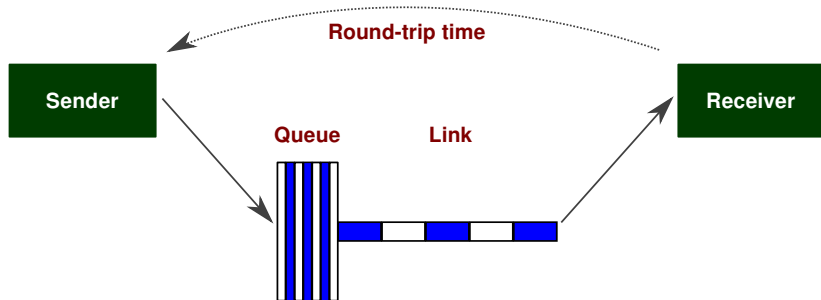
Minimize

- ▶ average flow completion time
- ▶ page load time
- ▶ tail completion time

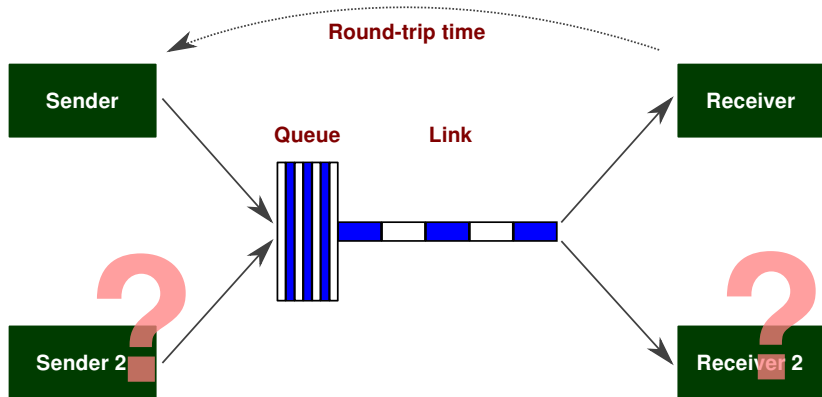
Prior assumptions

- ▶ Model of network uncertainty
 - ▶ Link speed distribution
 - ▶ Delay distribution
- ▶ Traffic model
 - ▶ Web browsing, MapReduce, videoconferencing

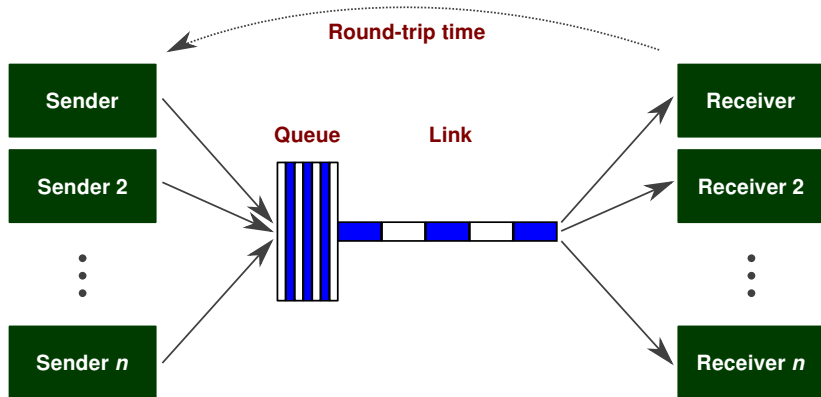
Dumbbell network



Dumbbell network



Dumbbell network



Superrational congestion control

At this moment, * do I:

- ▶ send a packet
- ▶ not send a packet?

* Assuming every node is running the same algorithm.

Remy: search for superrationality

- ▶ Remy searches for the best congestion-control algorithm
- ▶ Optimizes expected objective over prior assumptions
- ▶ Makes tractable by **limiting available state**

A RemyCC tracks three congestion signals

r_ewma: moving average of interval between acks

s_ewma: ... between sender timestamps echoed in acks

rtt_ratio: ratio of last RTT to smallest RTT so far

Why these three congestion signals?

- ▶ Benefit can be measured empirically
 - ▶ In our experiments, little help from adding more
 - ▶ Other networks might find differently
- ▶ More signals increase search time

A RemyCC maps each state to an action

$$\text{RULE}(r_ewma, s_ewma, rtt_ratio) \rightarrow \langle m, b, \tau \rangle$$

m Multiple to congestion window

b Increment to congestion window

τ Minimum interval between two outgoing packets

Runtime for a RemyCC

On ack:

- ▶ $\langle m, b, \tau \rangle \leftarrow \text{RULE}(r_ewma, s_ewma, rtt_ratio)$
- ▶ $wnd \leftarrow m \cdot wnd + b$

Send packet if:

- ▶ $wnd > \text{FlightSize}$, and
- ▶ last packet sent $> \tau$ ago

Remy's job

Find piecewise-continuous $\text{RULE}()$ that optimizes expected value of objective function.

Remy example: 2D state space

On ack:

$$\blacktriangleright \langle m, b, \tau \rangle \leftarrow \text{RULE}(r_ewma, s_ewma, \mathbf{rtt_ratio})$$

Remy example: 2D state space

On ack:

$$\triangleright \langle m, b, \tau \rangle \leftarrow \text{RULE}(r_ewma, s_ewma, \text{[redacted]})$$

Remy example: Prior assumptions

Quantity	Distribution	Units
Link speed	Uniform(10, 20)	Mbps
RTT	Uniform(100, 200)	ms
n	Uniform(1, 16)	
“On” process	$\exp[\mu = 5]$	seconds
“Off” process	same	

Remy example: Objective

$$\sum_i \log \left[\frac{\text{throughput}_i}{\text{delay}_i} \right]$$

One action for all states. Find the best value.

r_{ewma}

$\langle ?, ?, ? \rangle$

s_{ewma}



The best (single) action. Now split it on median.

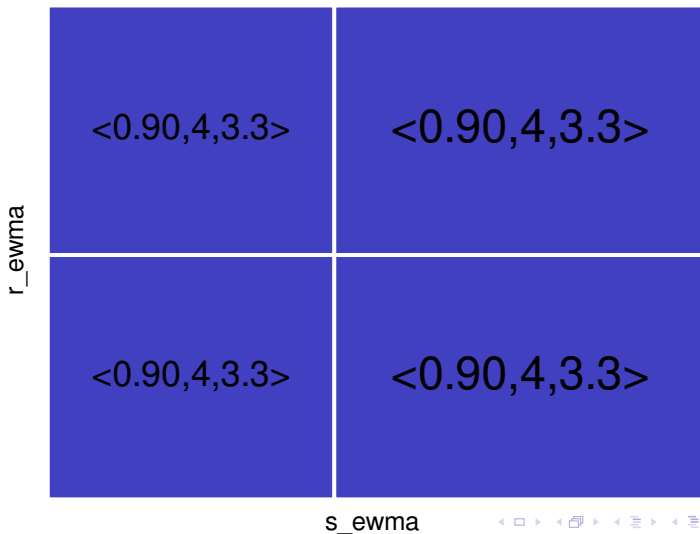
r_ewma

$\langle 0.90, 4, 3.3 \rangle$

s_ewma



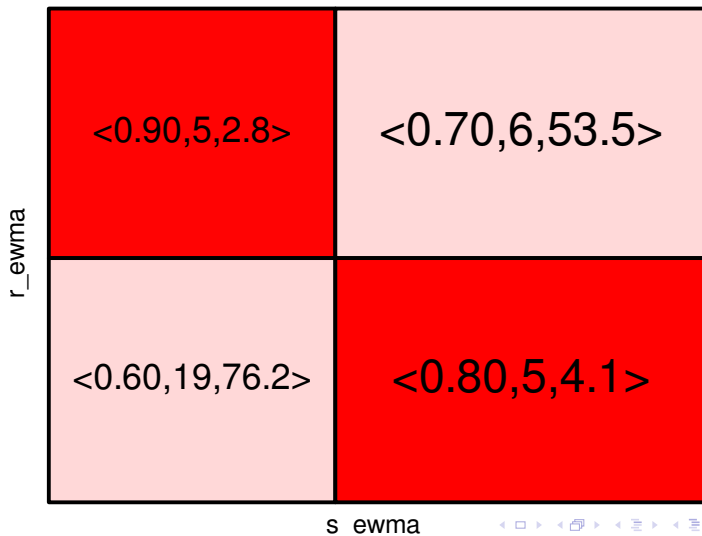
Simulate



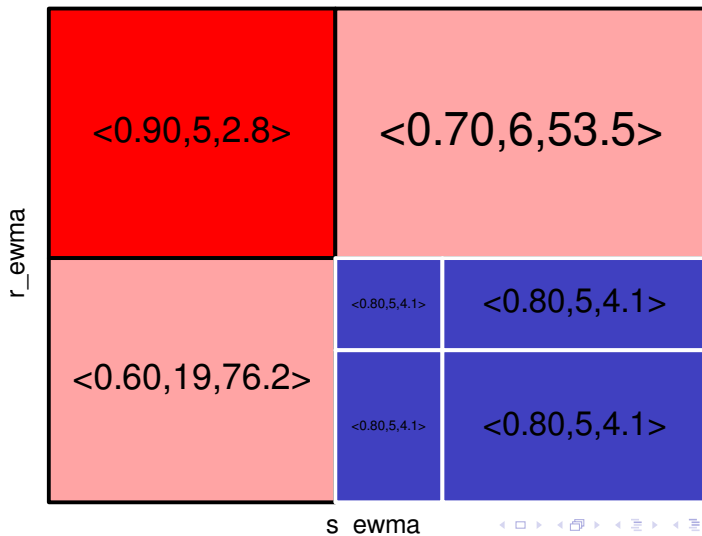
Optimize each of the new actions

r_ewma	$\langle 0.90, 4, 3.3 \rangle$	$\langle 0.90, 4, 3.3 \rangle$
	$\langle 0.90, 4, 3.3 \rangle$	$\langle 0.90, 4, 3.3 \rangle$
	s_ewma	

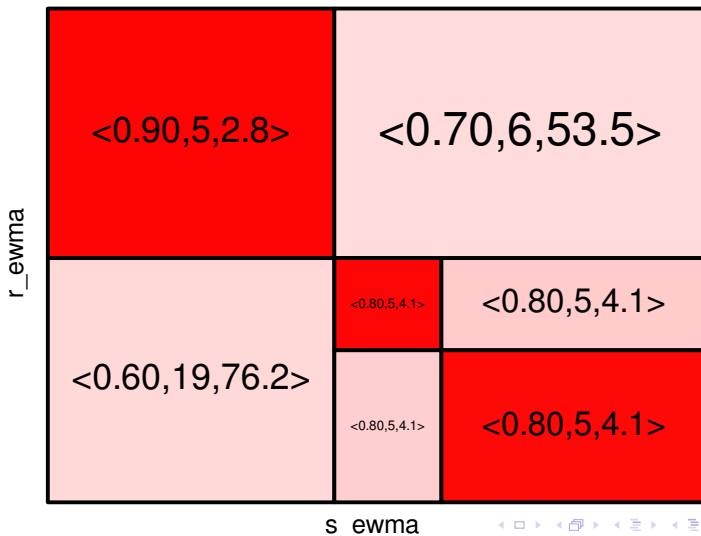
Now split the most-used rule



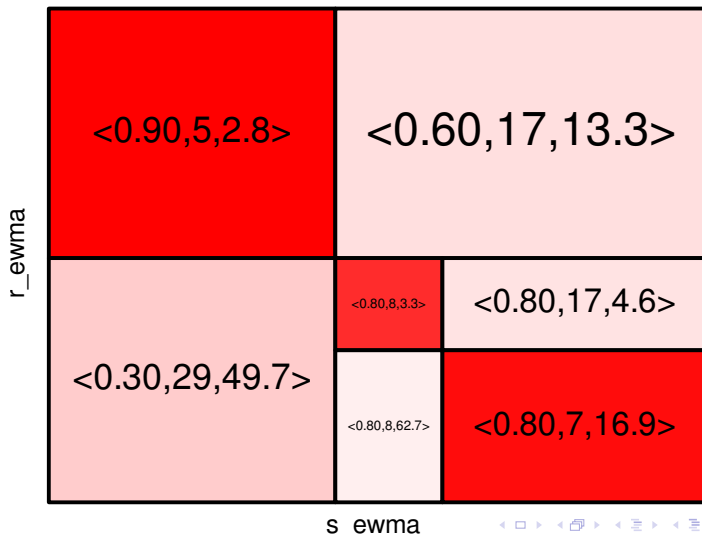
Simulate



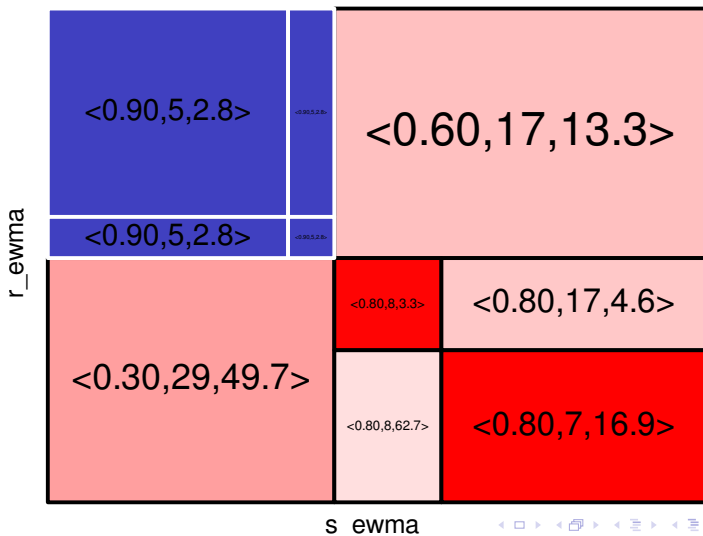
Optimize



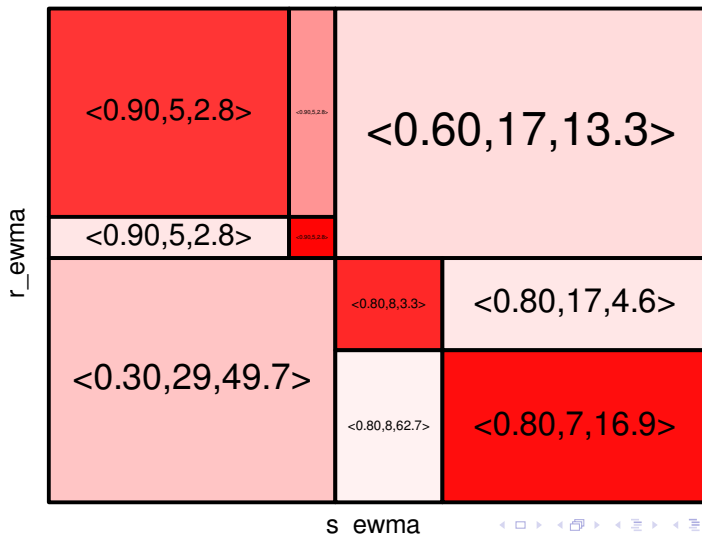
Split



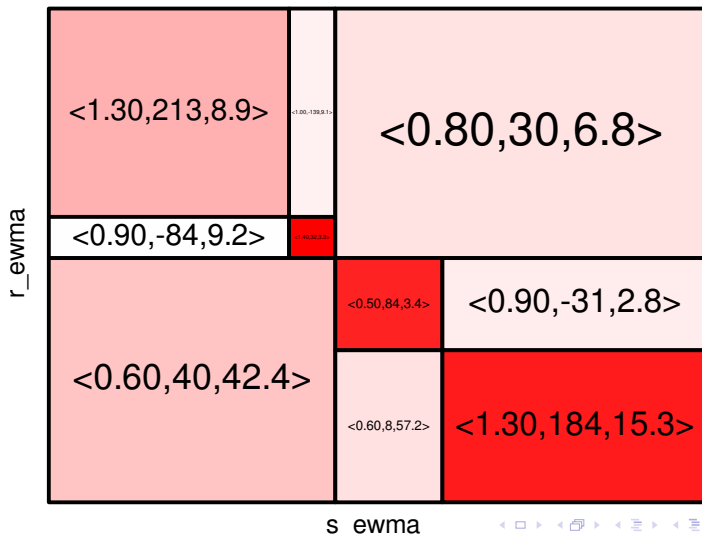
Simulate



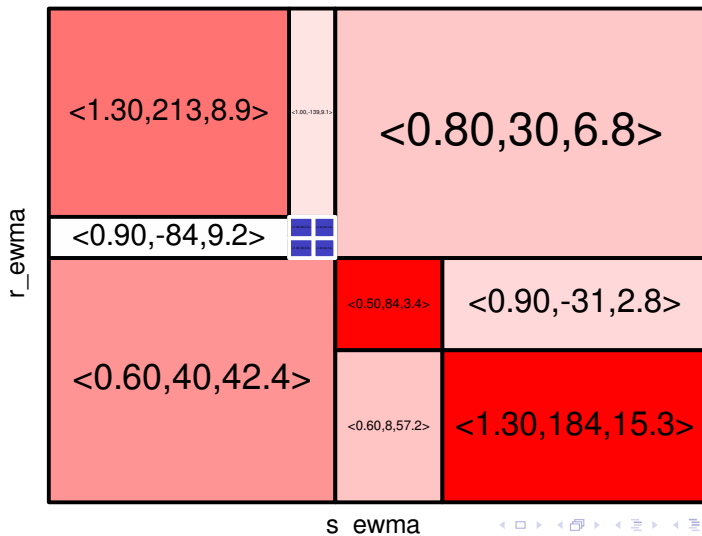
Optimize



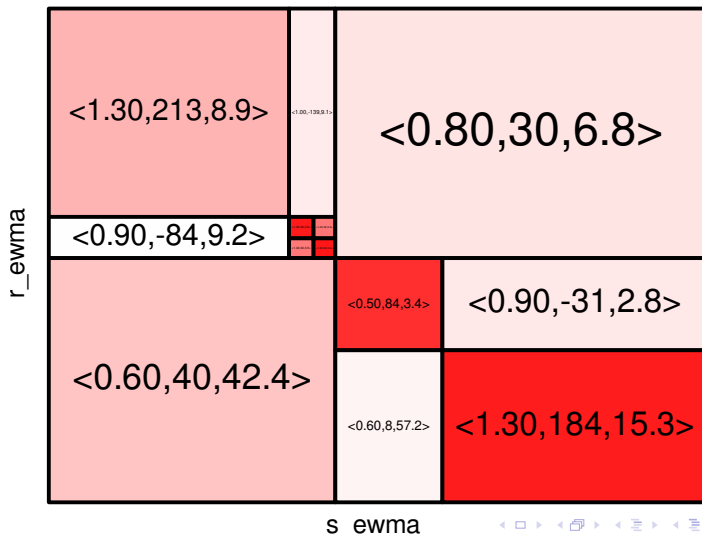
Split



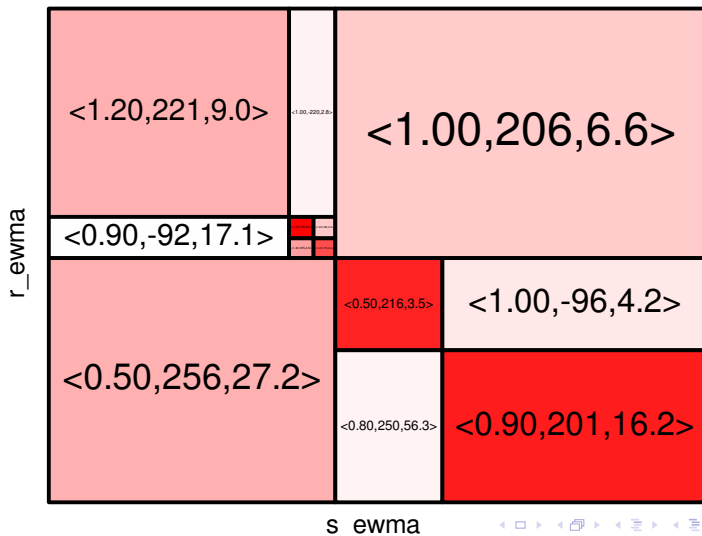
Simulate



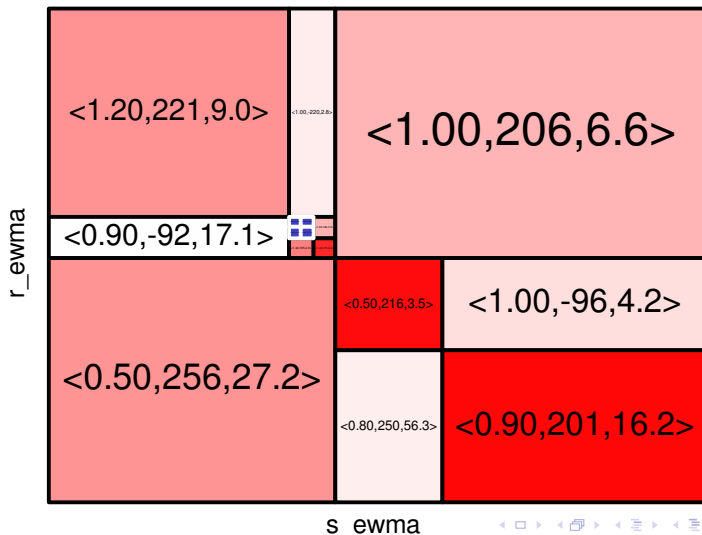
Optimize



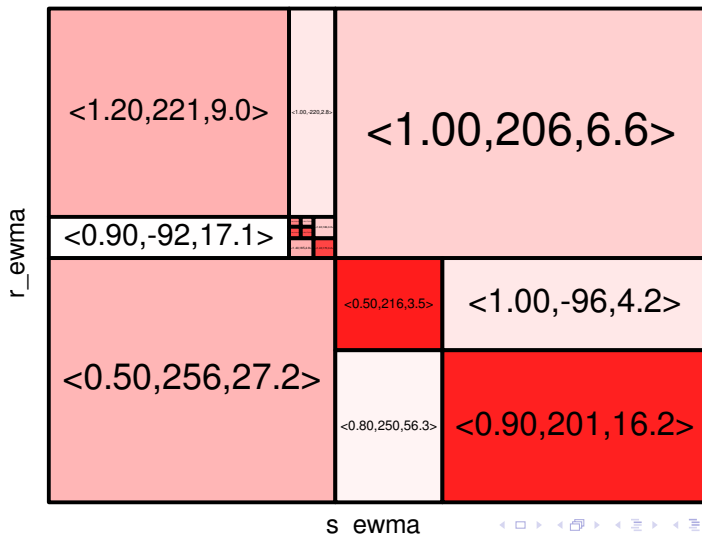
Split



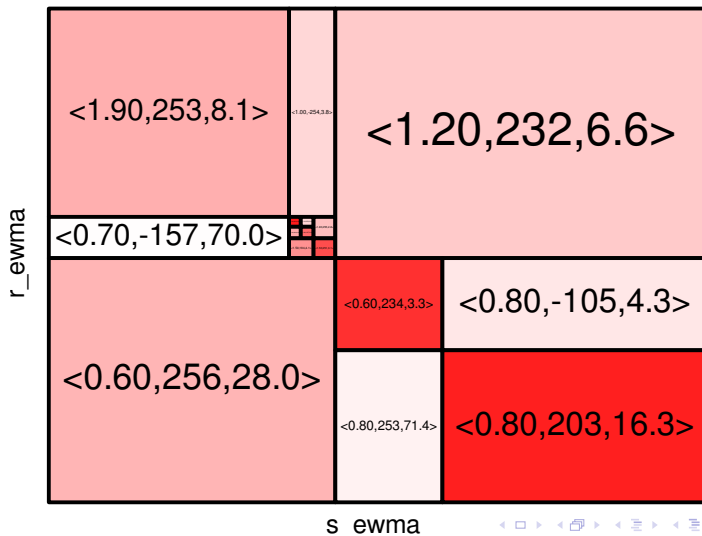
Simulate



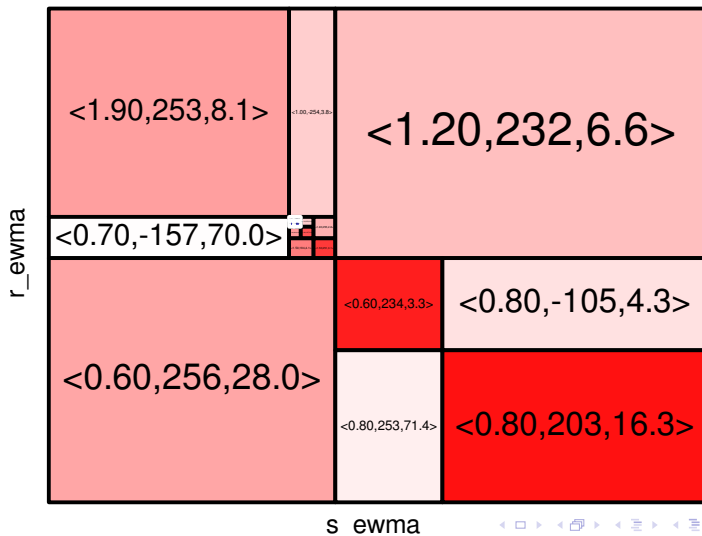
Optimize



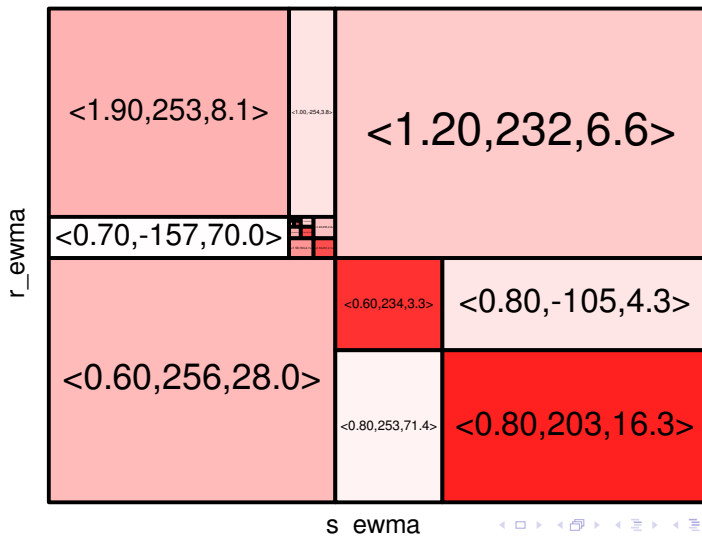
Split



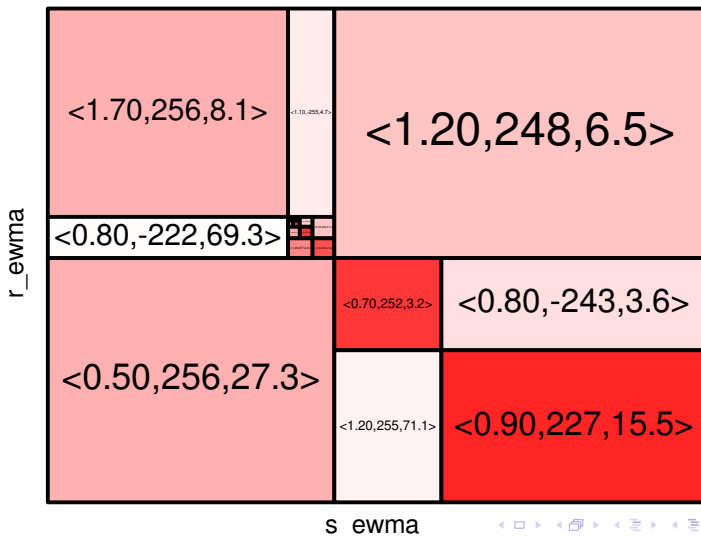
Simulate



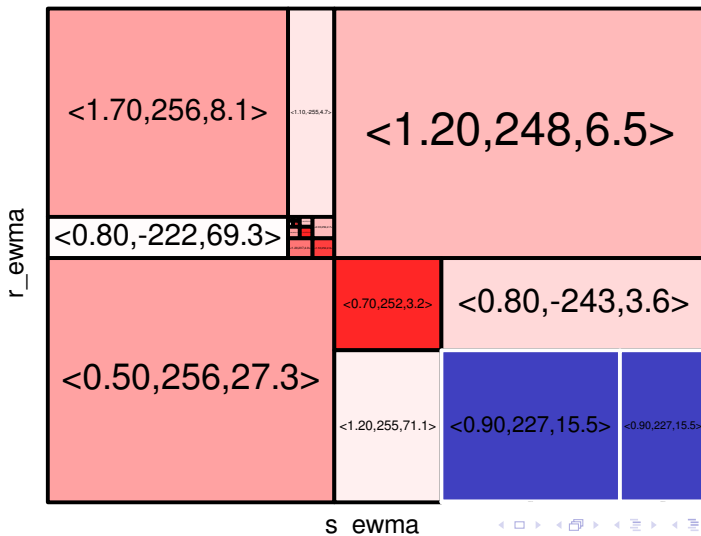
Optimize



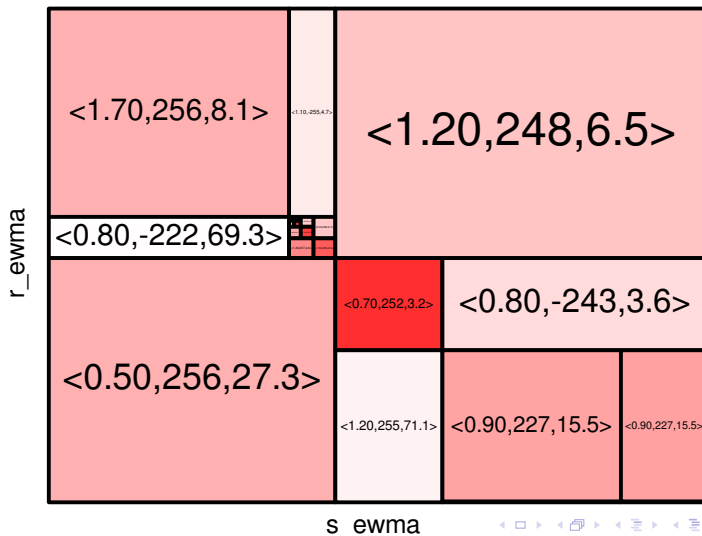
Split



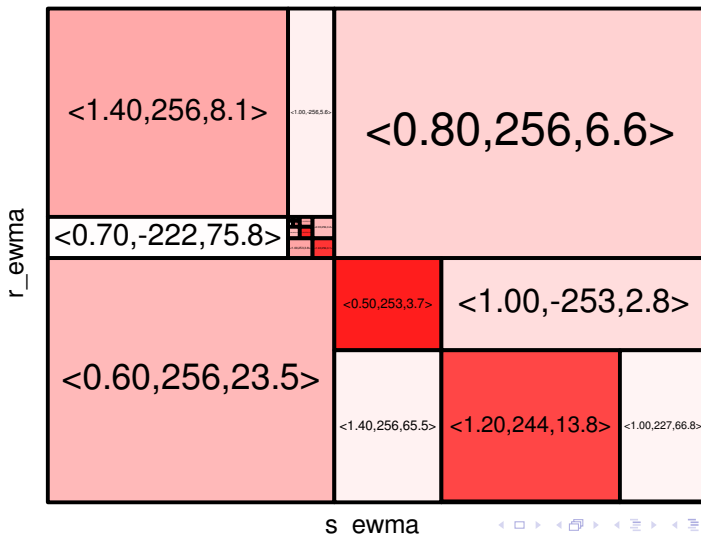
Simulate



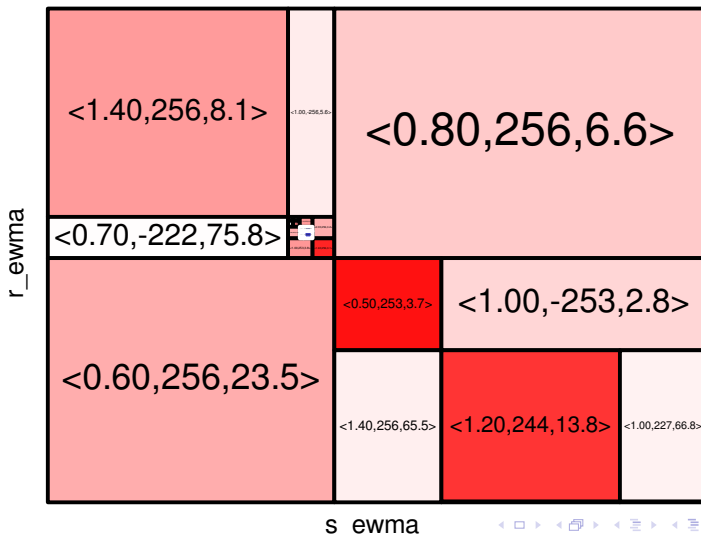
Optimize



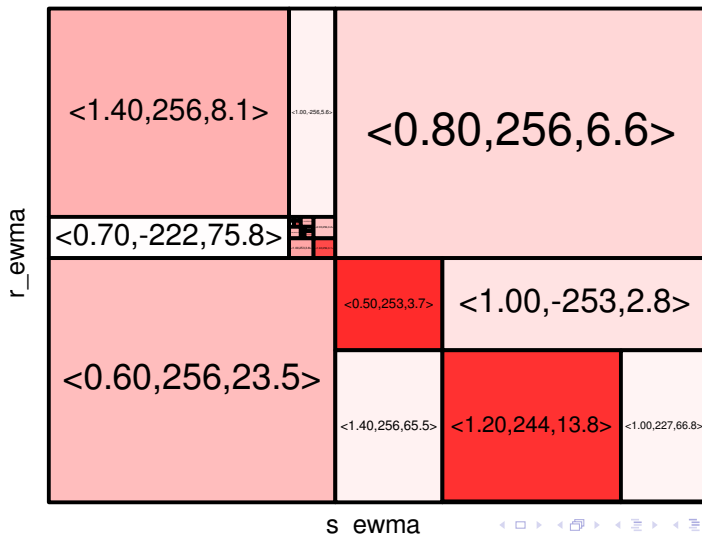
Split



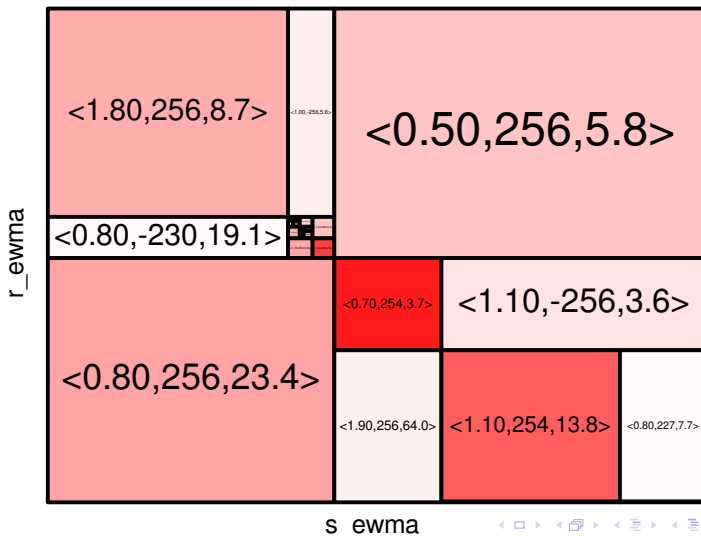
Simulate



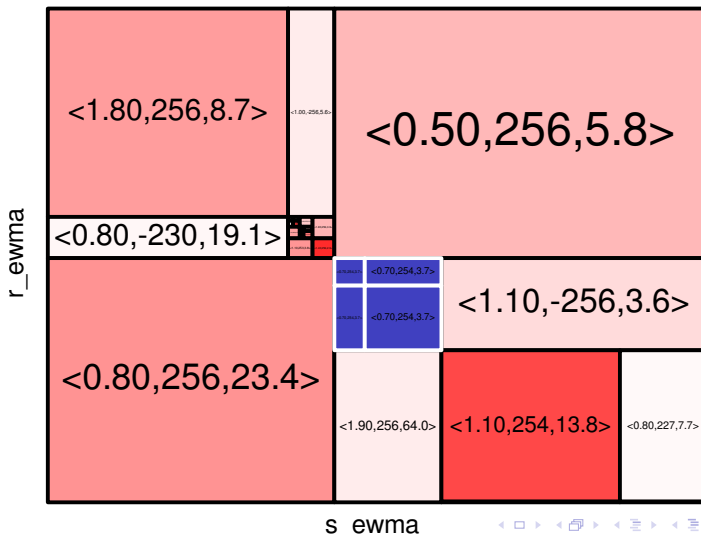
Optimize



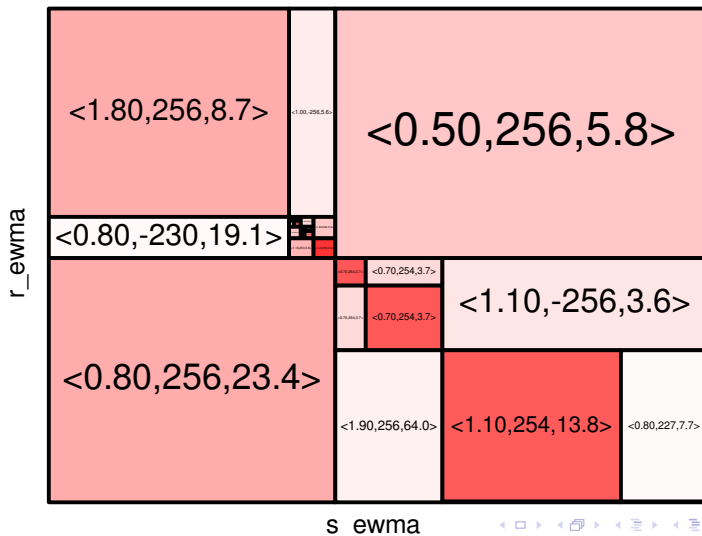
Split



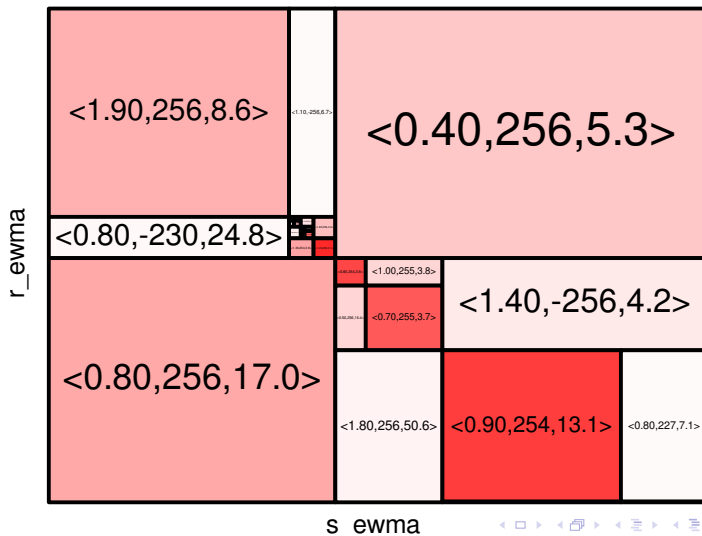
Simulate



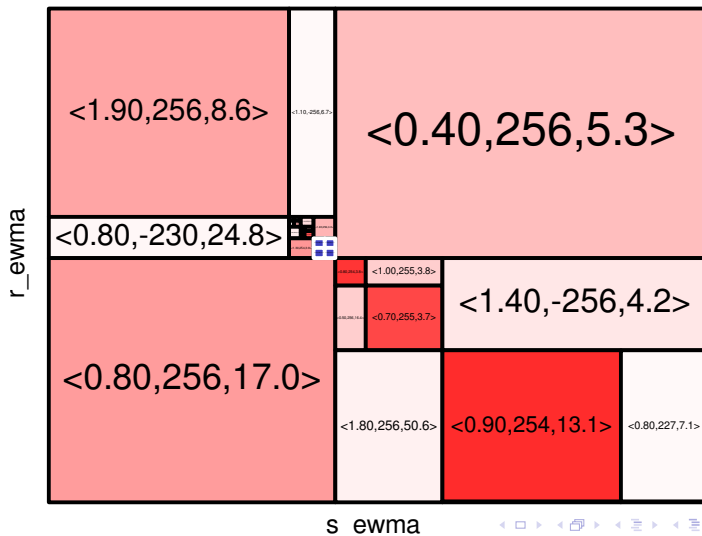
Optimize



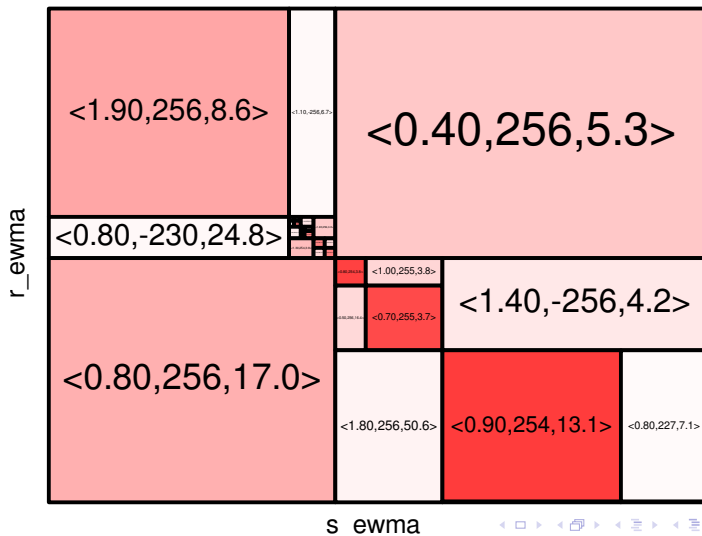
Split



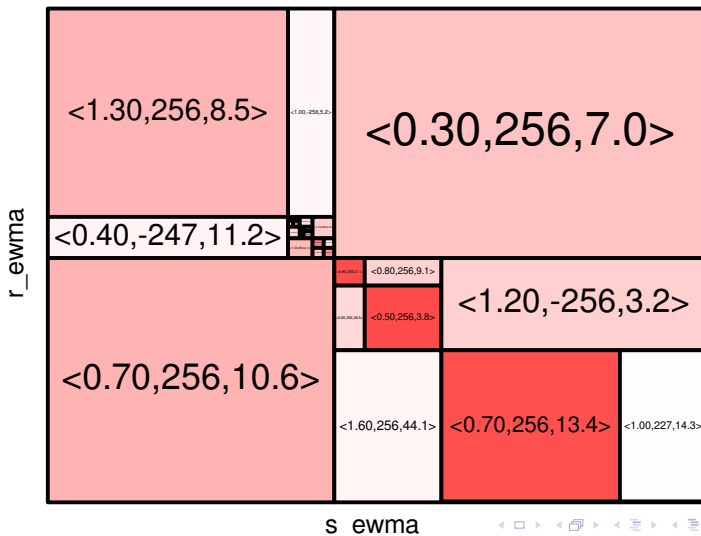
Simulate



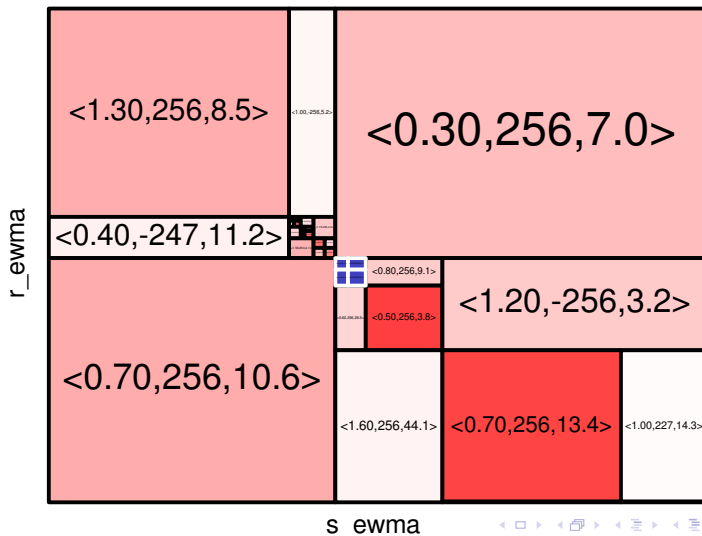
Optimize



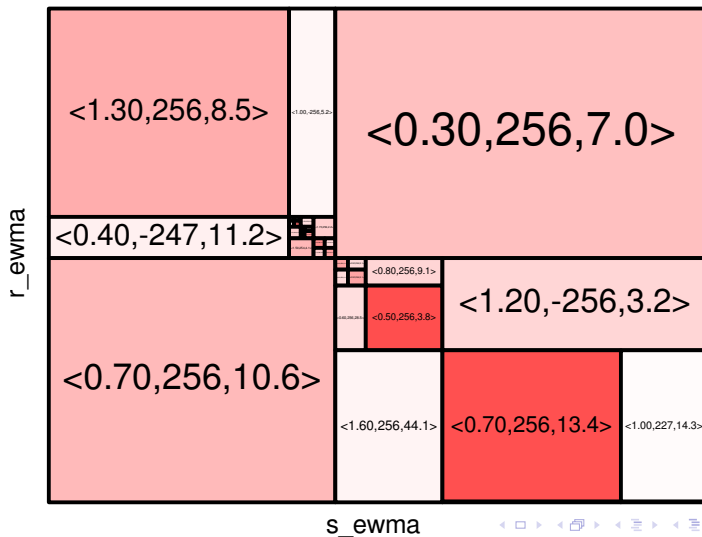
Split



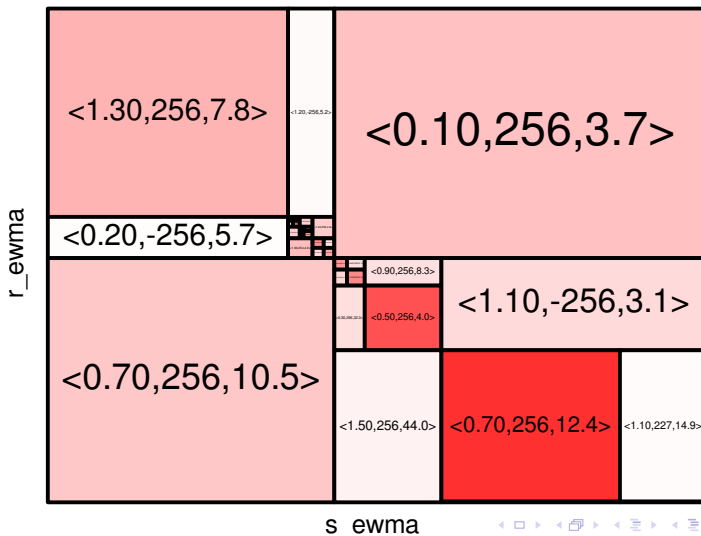
Simulate



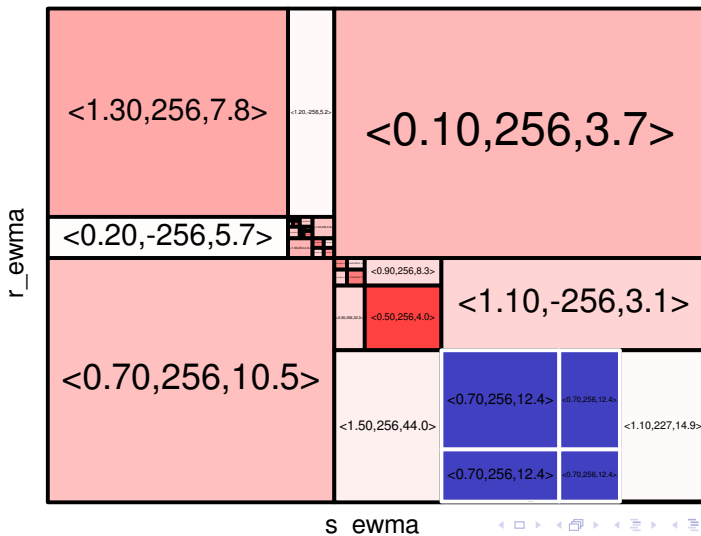
Optimize



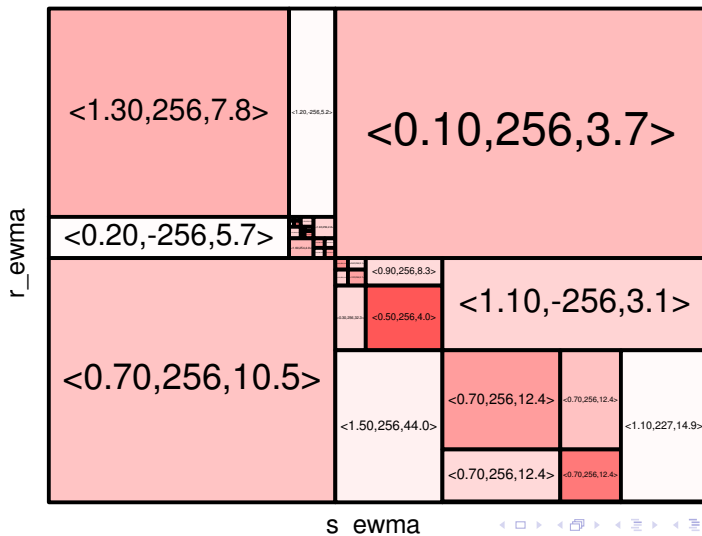
Split



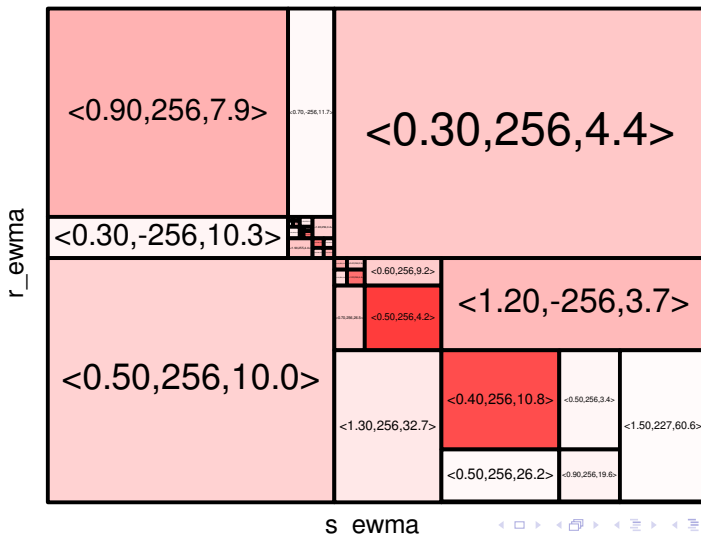
Simulate



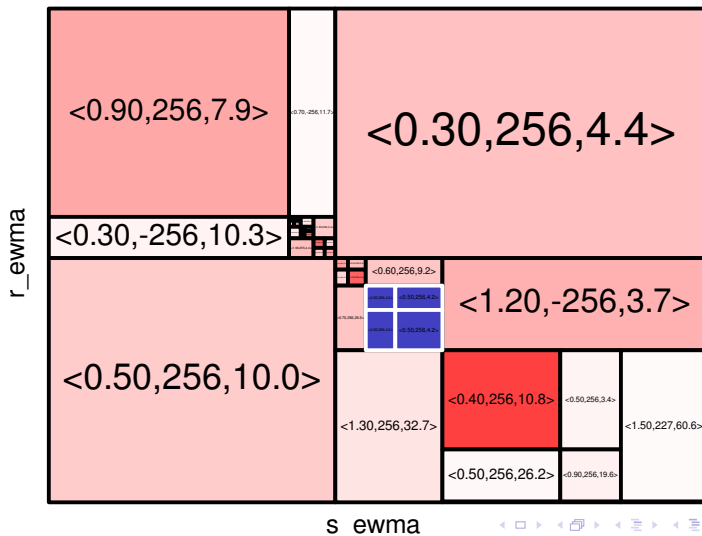
Optimize



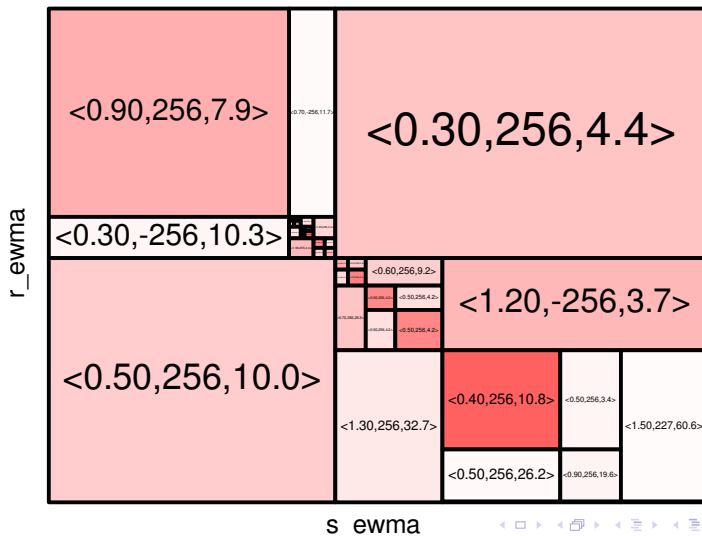
Split



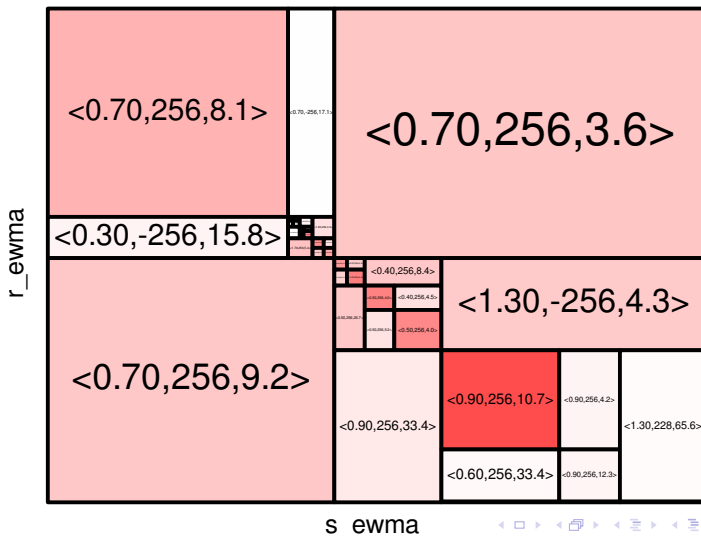
Simulate



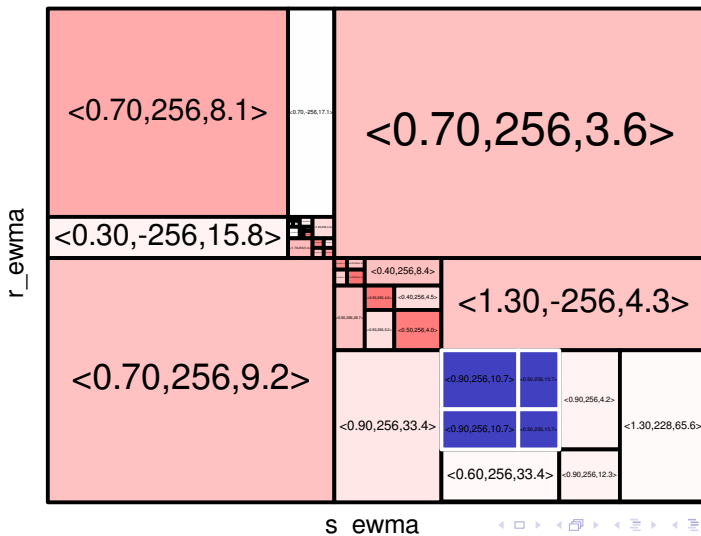
Optimize



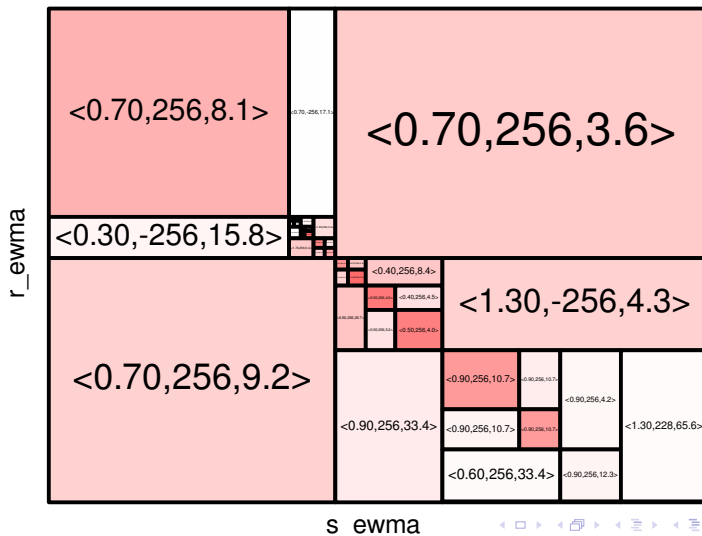
Split



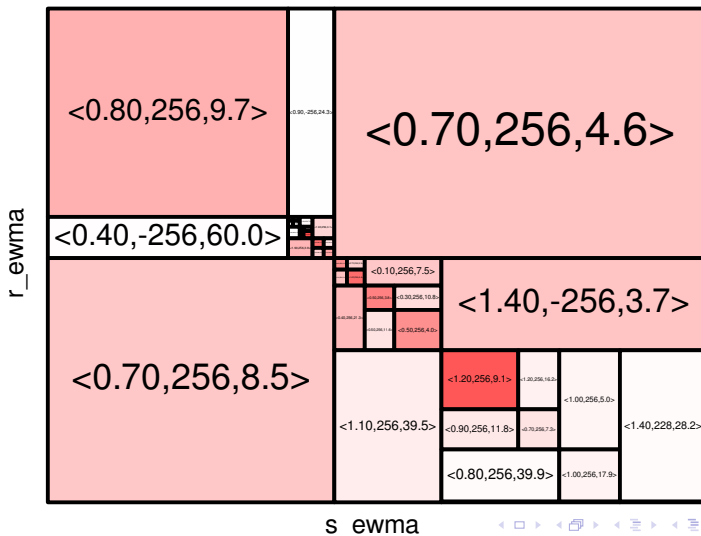
Simulate



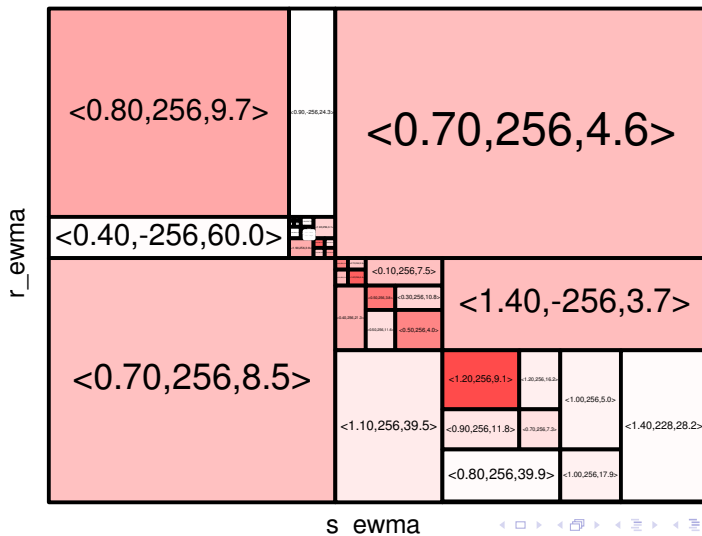
Optimize



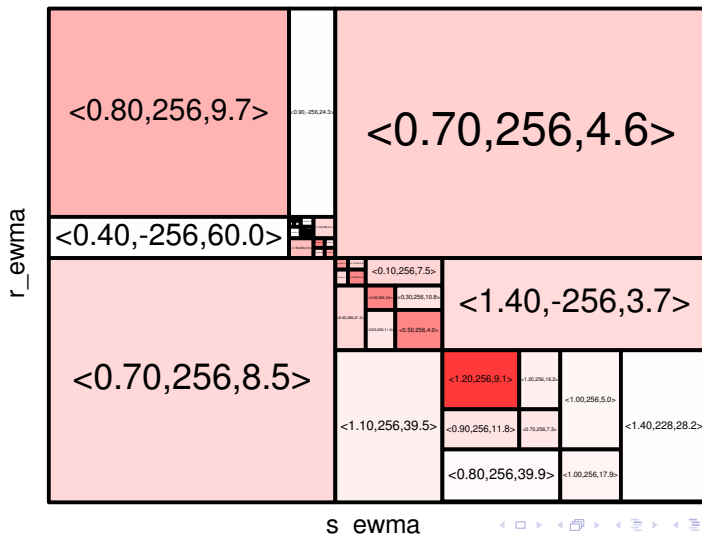
Split



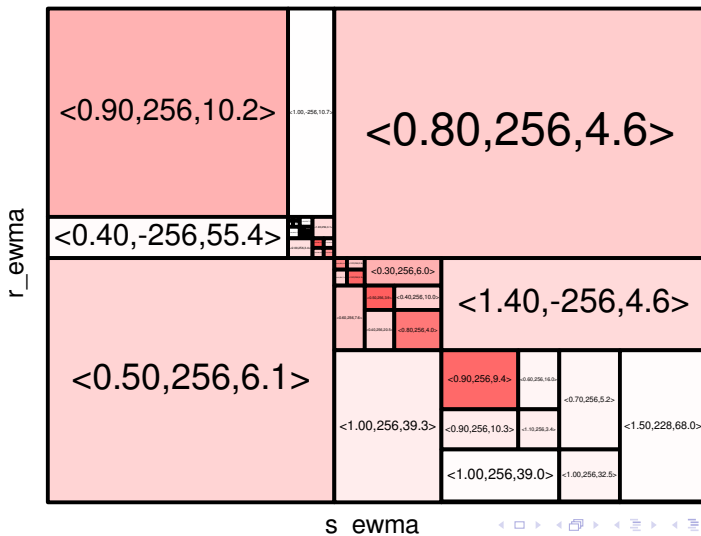
Simulate



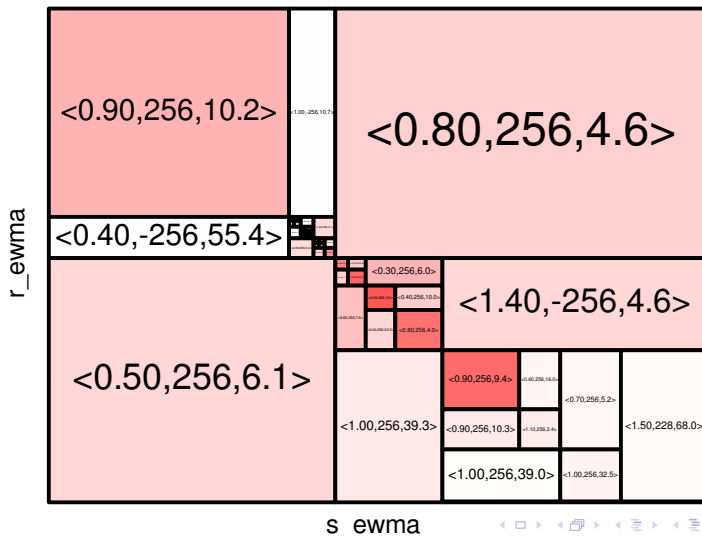
Optimize



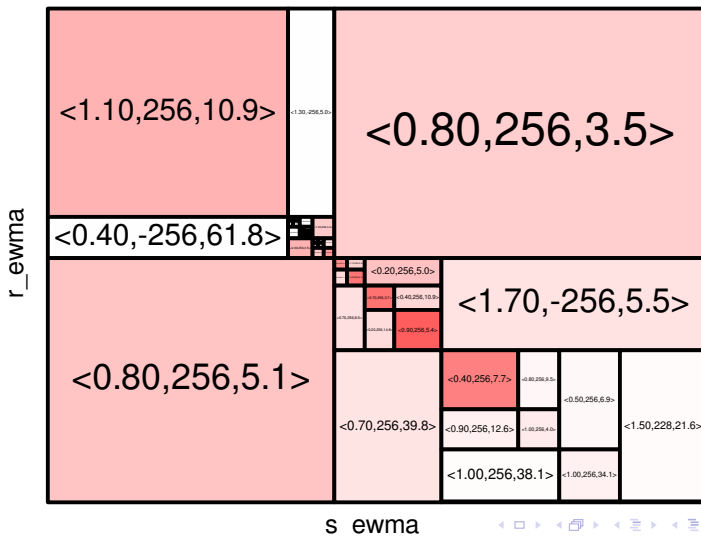
Split



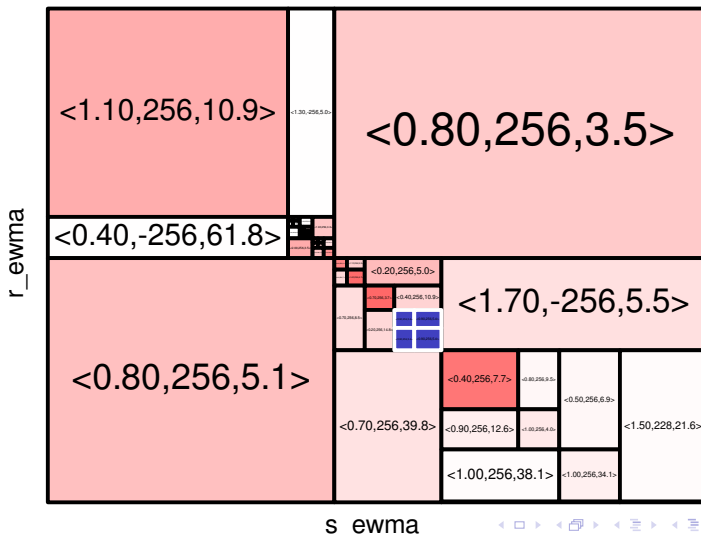
Optimize



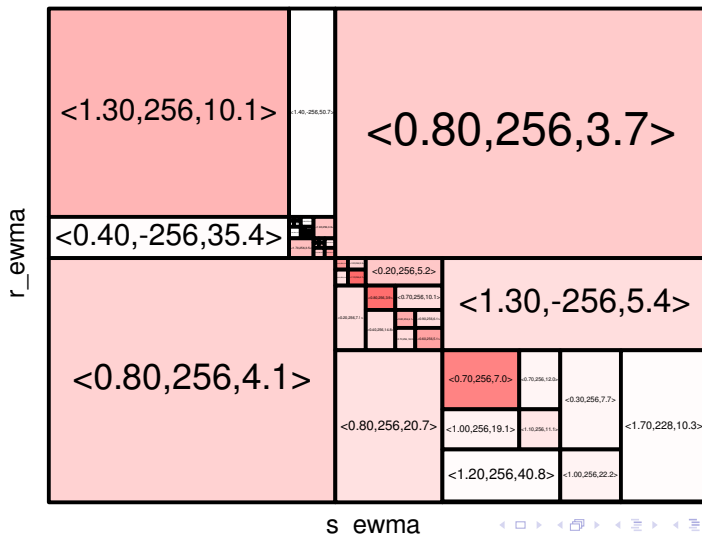
Split



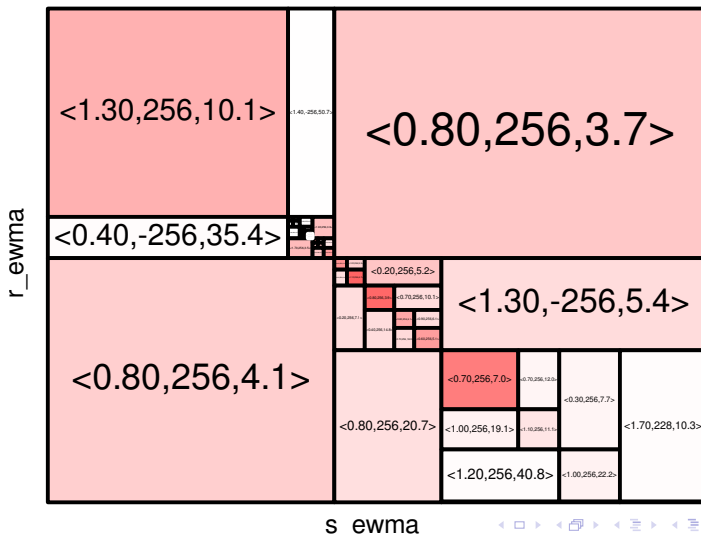
Simulate



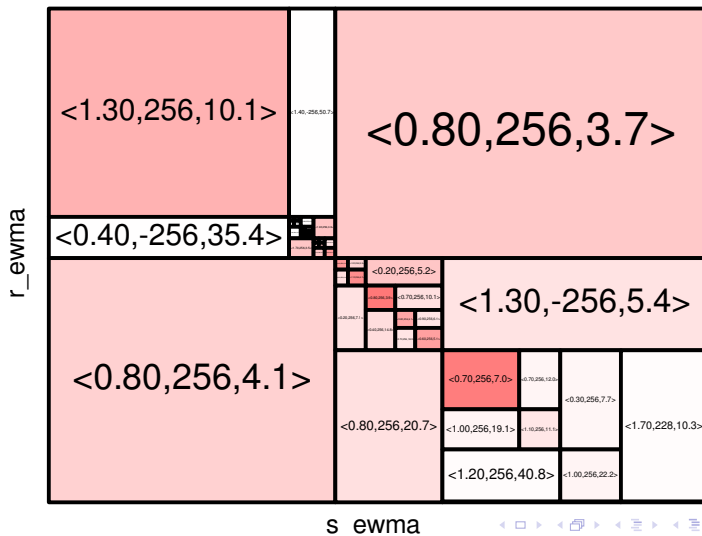
Split



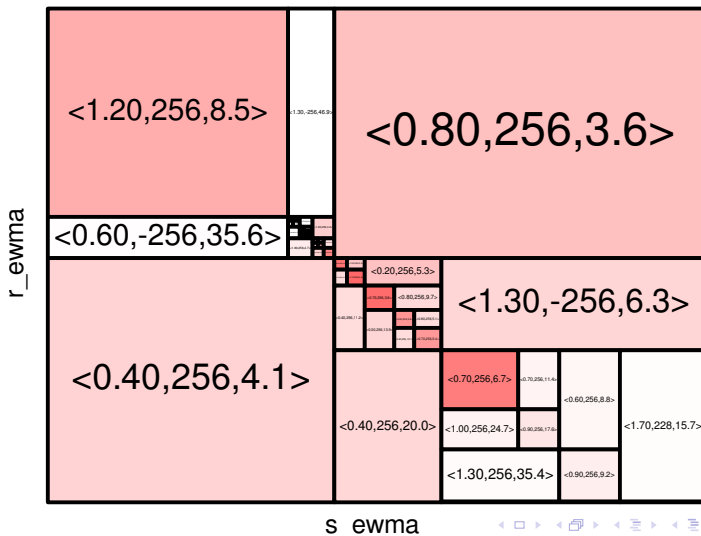
Simulate



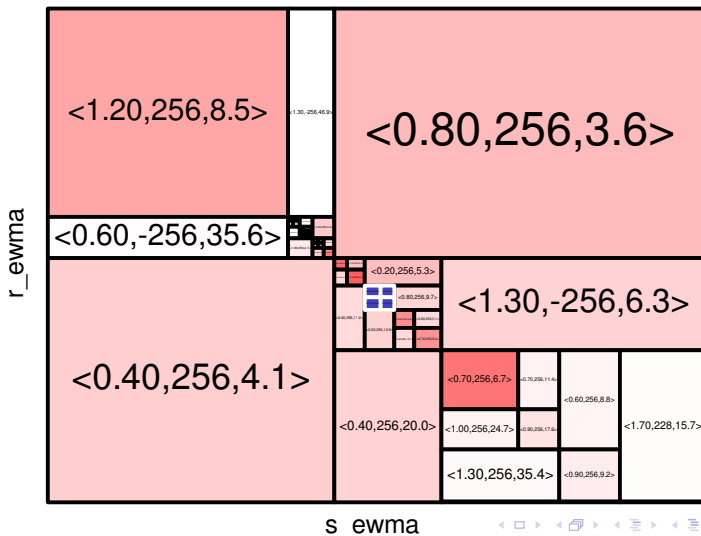
Optimize



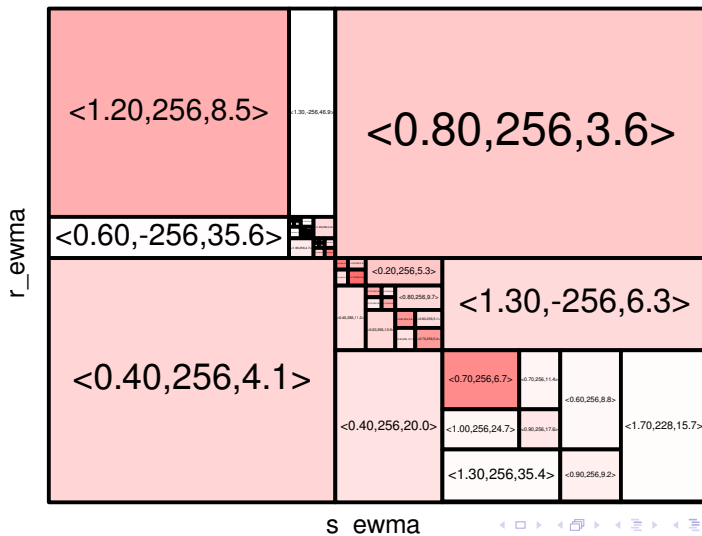
Split



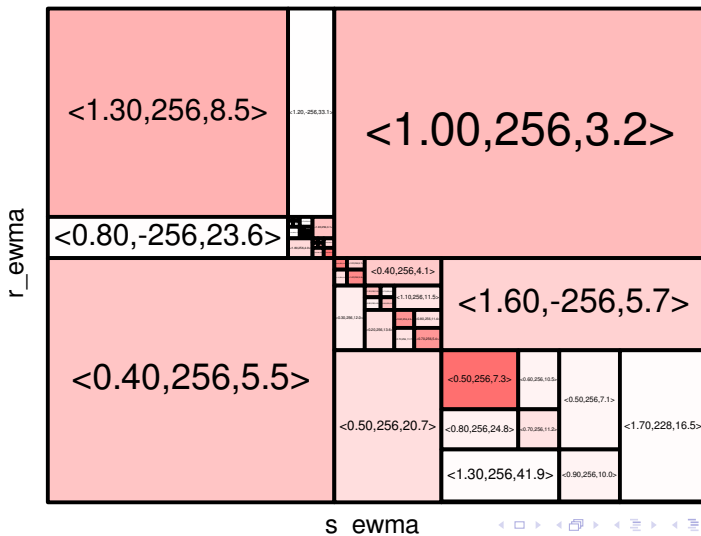
Simulate



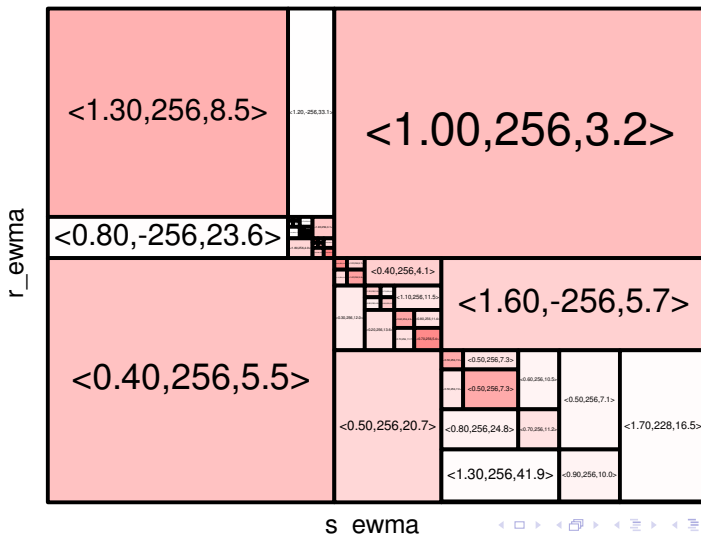
Optimize



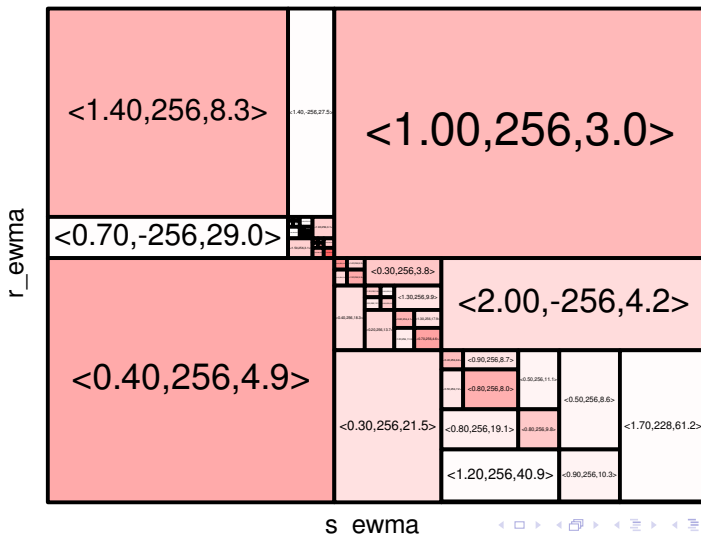
Split



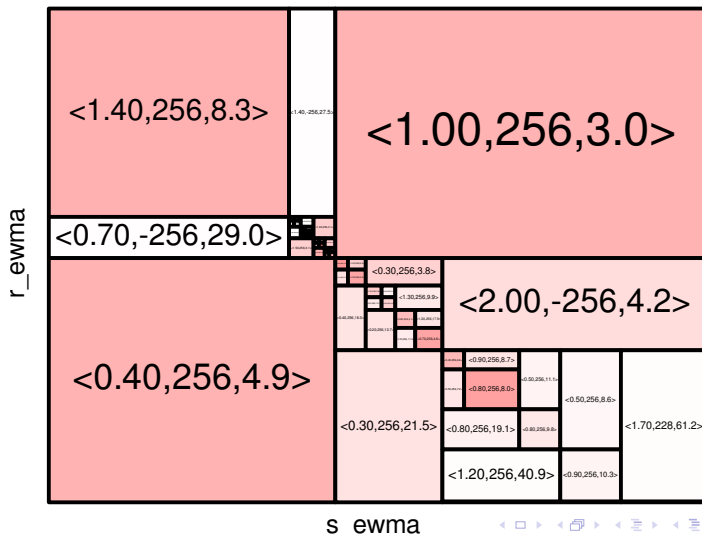
Optimize



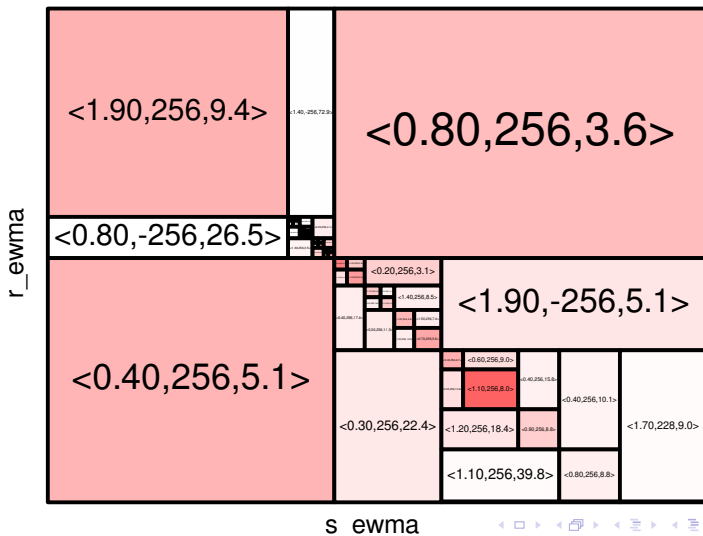
Split



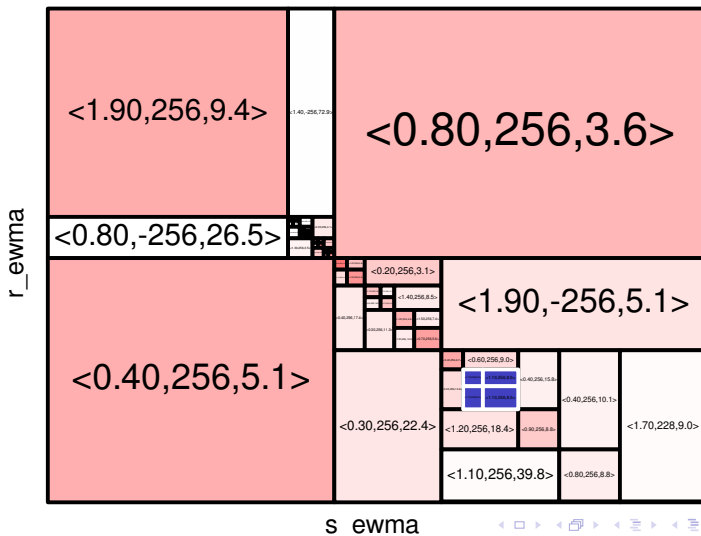
Optimize



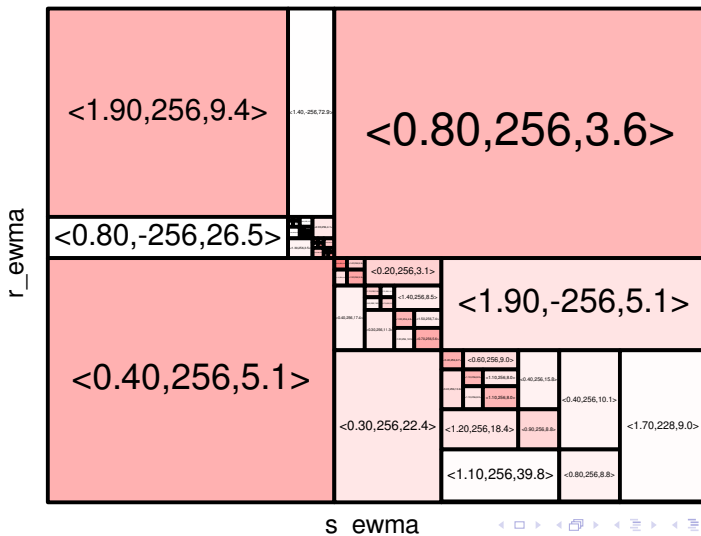
Split



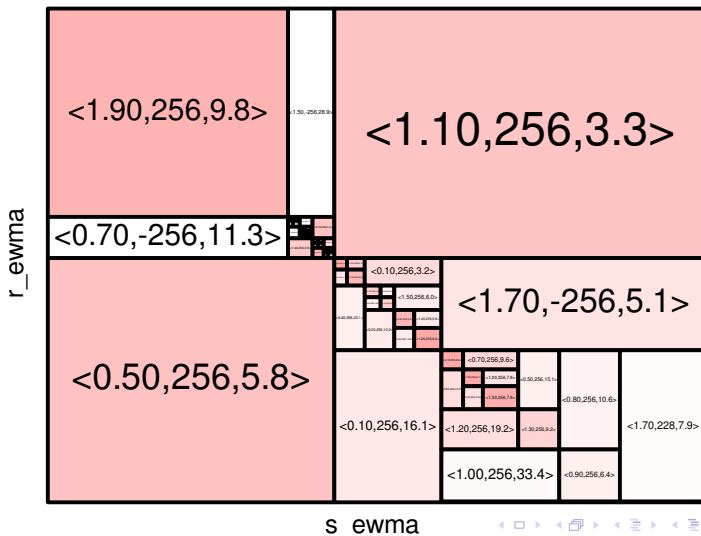
Simulate



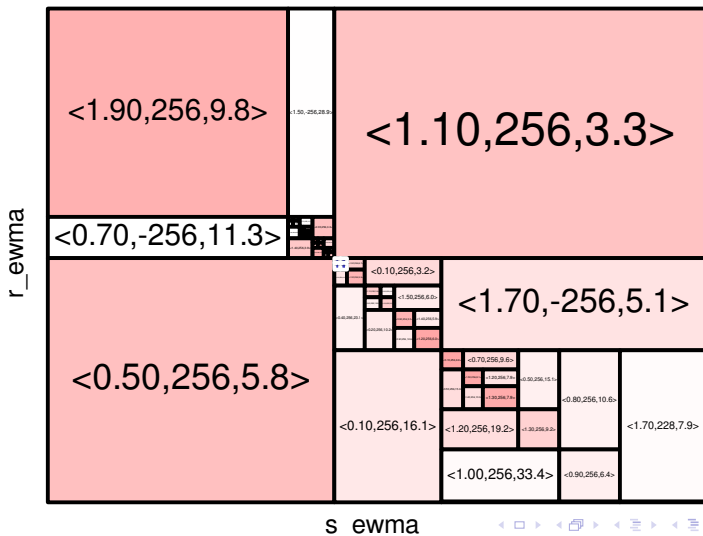
Optimize



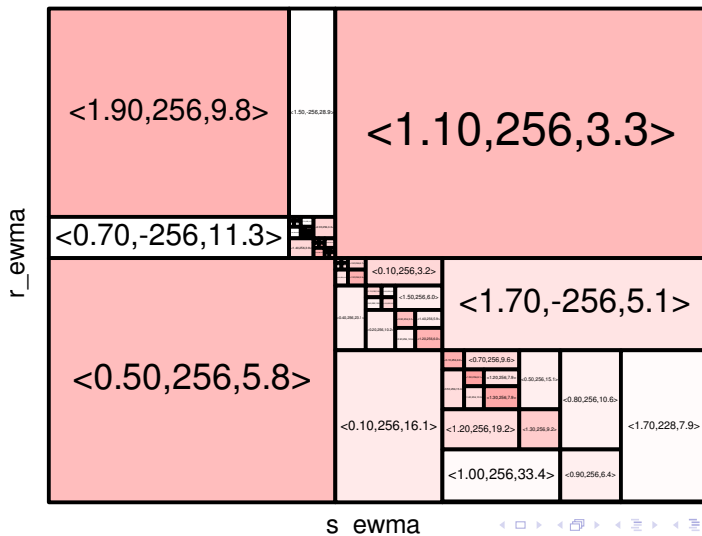
Split



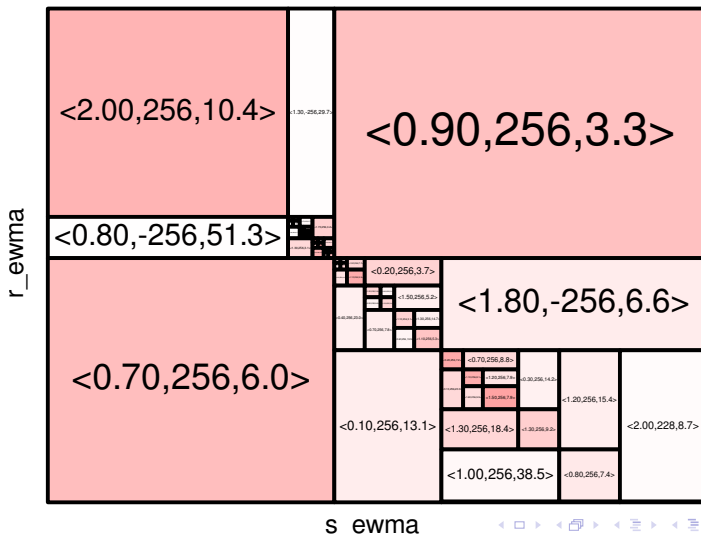
Simulate



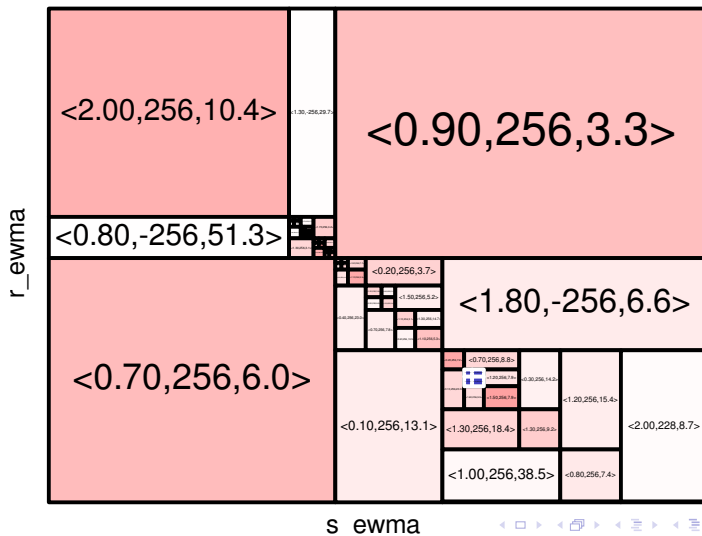
Optimize



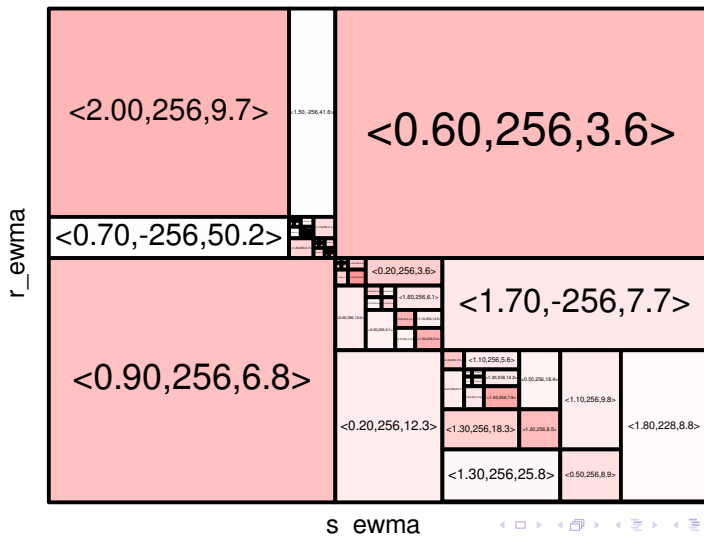
Split



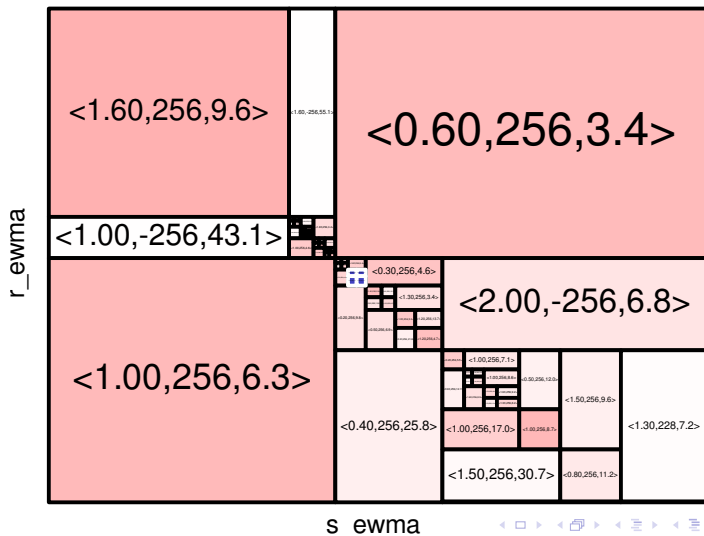
Simulate



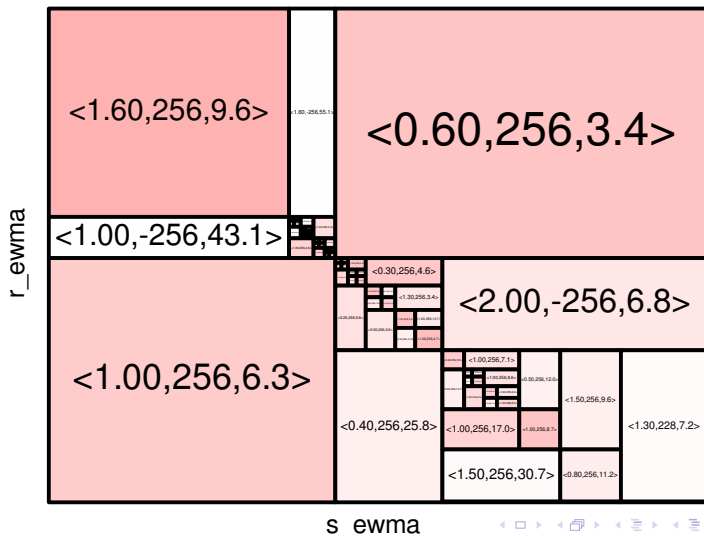
Split



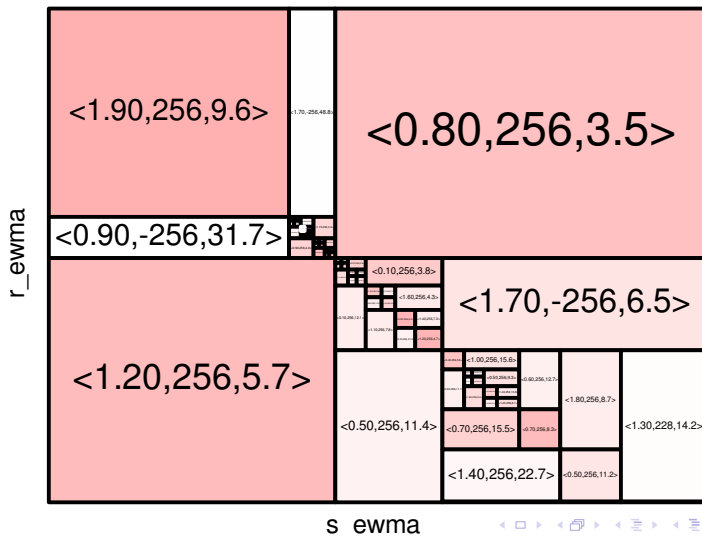
RemyCC



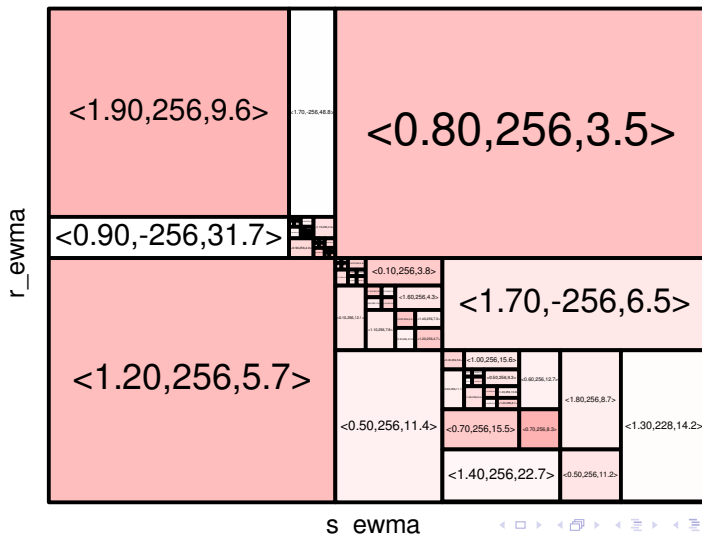
RemyCC



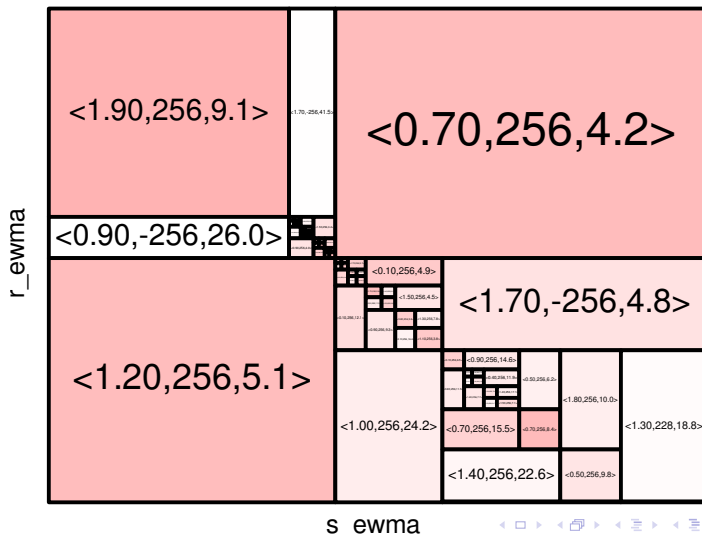
RemyCC



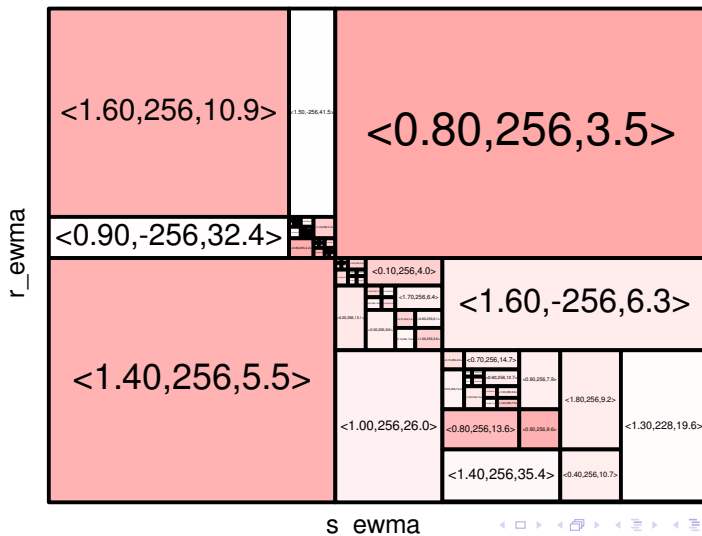
RemyCC



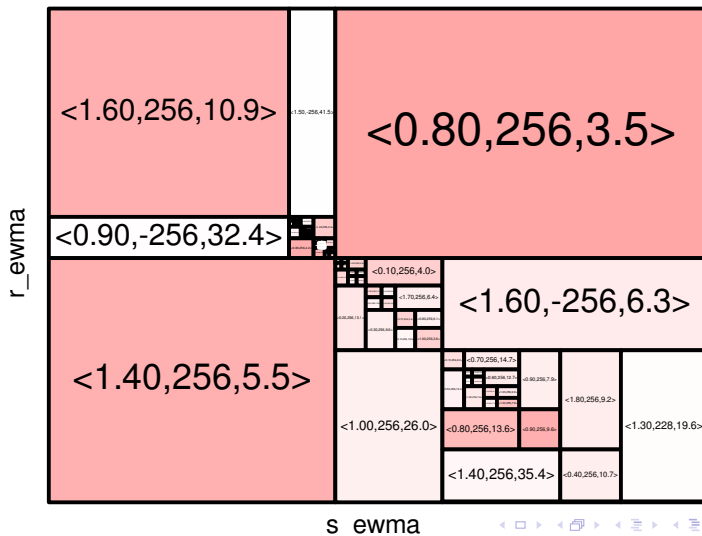
RemyCC



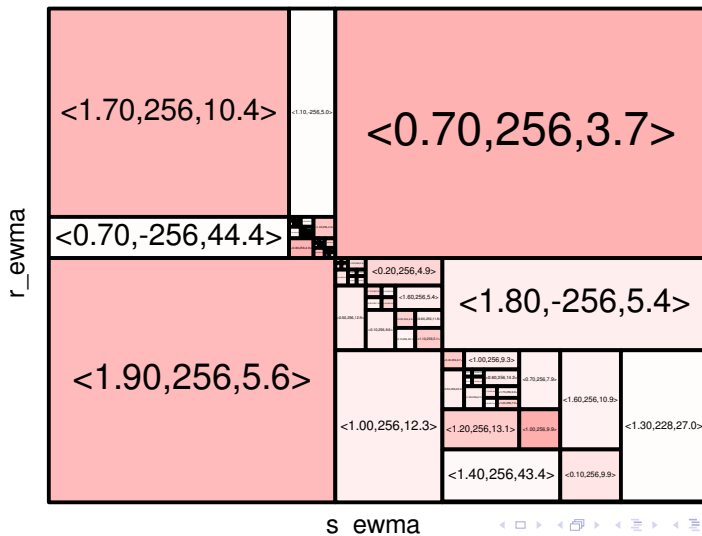
RemyCC



RemyCC

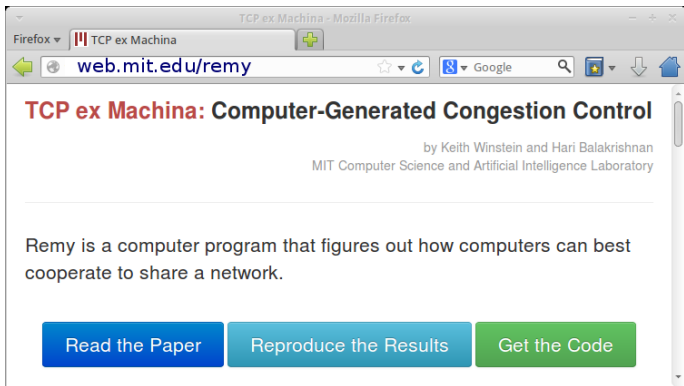


RemyCC

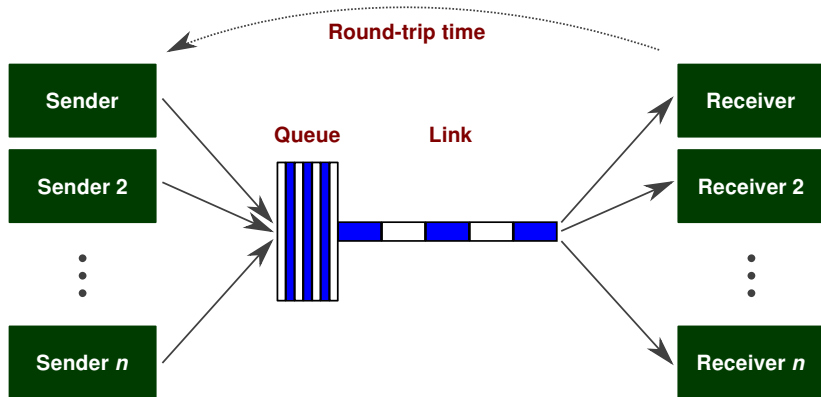


Evaluation in ns-2

- ▶ End-to-end comparators: **NewReno, Cubic, Compound, Vegas**
- ▶ In-net comparators: **Cubic-over-sfqCoDel, XCP**
- ▶ Simulation setup published for replication



Scenario 1: fixed-rate network, homogenous senders



Scenario 1: details

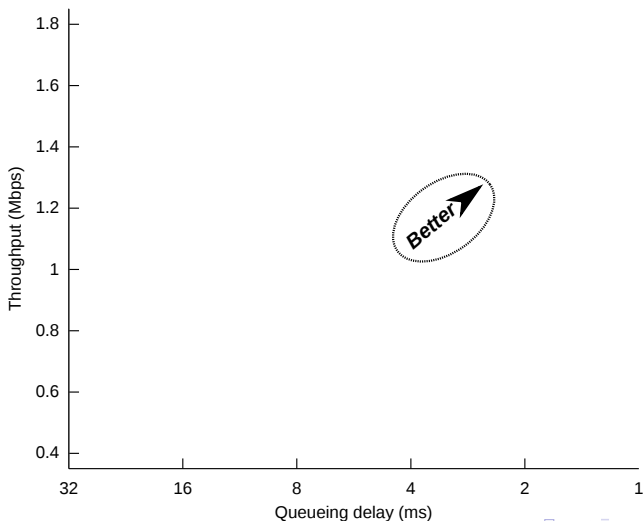
Quantity	Simulation parameter	Remy assumptions
Link speed	15 Mbps	Uniform(10, 20) Mbps
RTT	150 ms	Uniform(100, 200) ms
n	8	Uniform(1, 16)
“On” process	$\exp[\mu = 100]$ kB	$\exp[\mu = 5]$ s
“Off” process	$\exp\left[\mu = \frac{1}{2}\right]$ s	$\exp[\mu = \mathbf{5}]$ s

Remy objective:

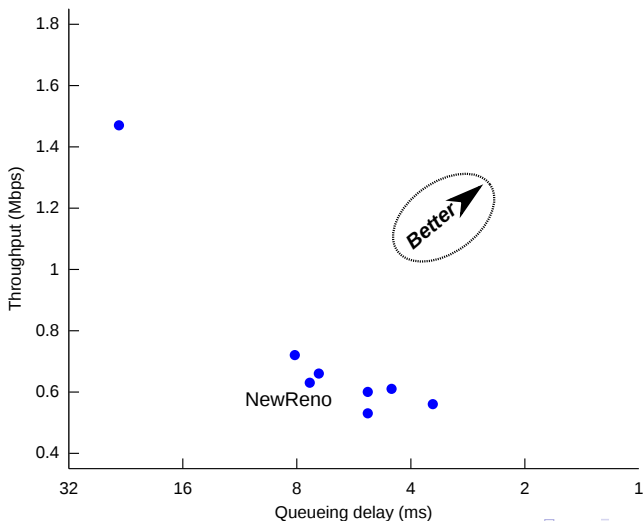
$$\sum_i \log \left[\frac{\text{throughput}_i}{(\text{delay}_i)^\delta} \right]$$

$$\delta \in \left\{ \frac{1}{10}, 1, 10 \right\}$$

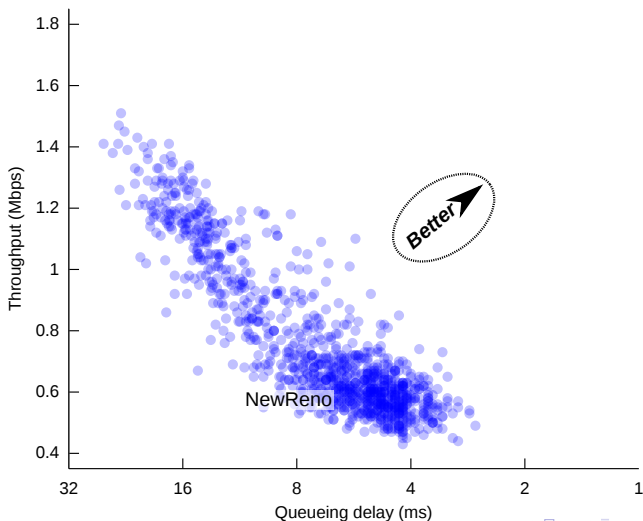
Scenario 1: throughput-delay plot



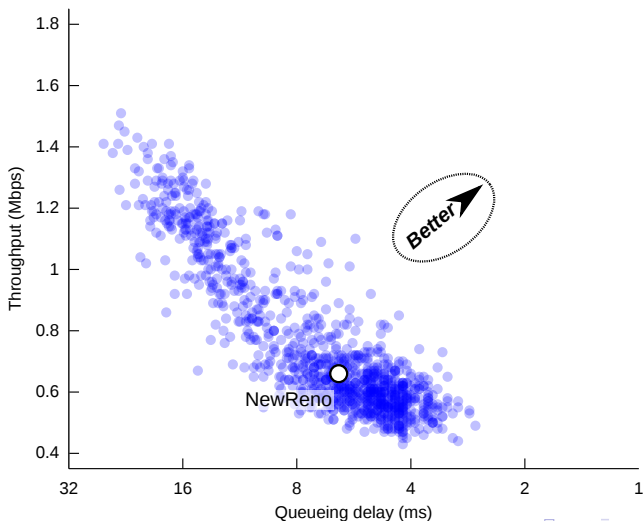
Scenario 1: throughput-delay plot



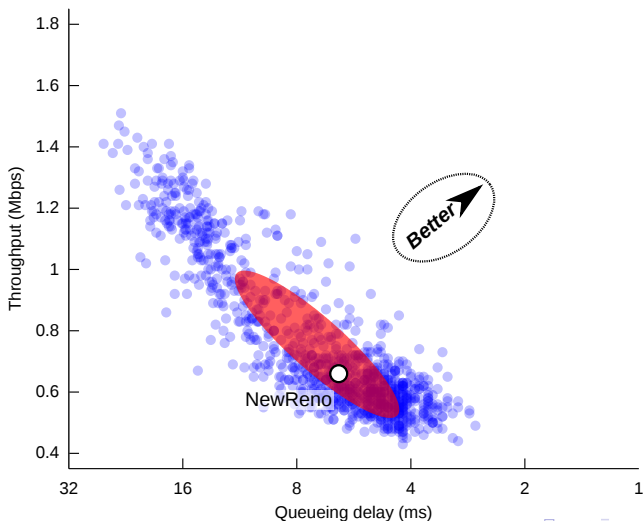
Scenario 1: throughput-delay plot



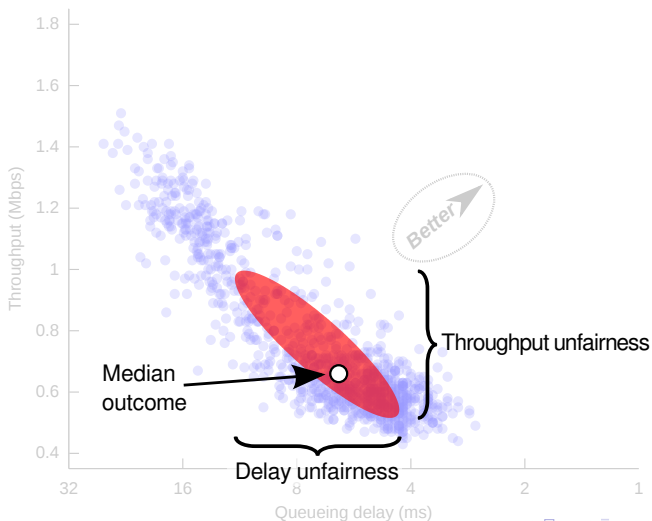
Scenario 1: throughput-delay plot



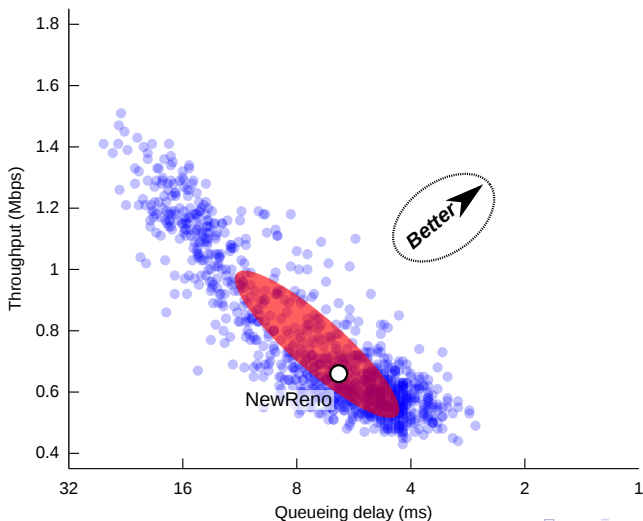
Scenario 1: throughput-delay plot



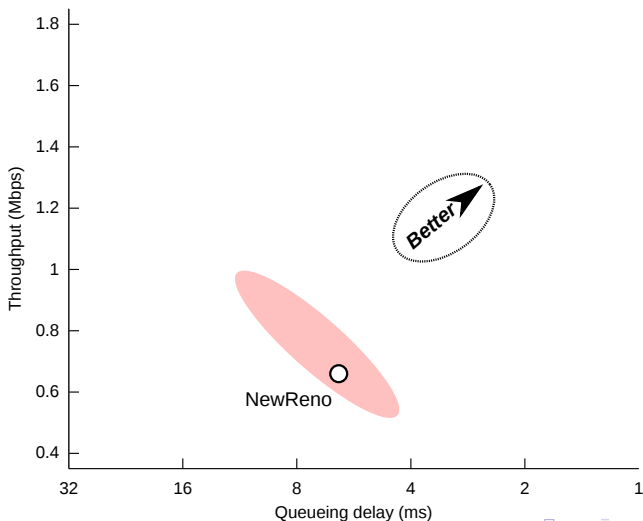
Scenario 1: throughput-delay plot



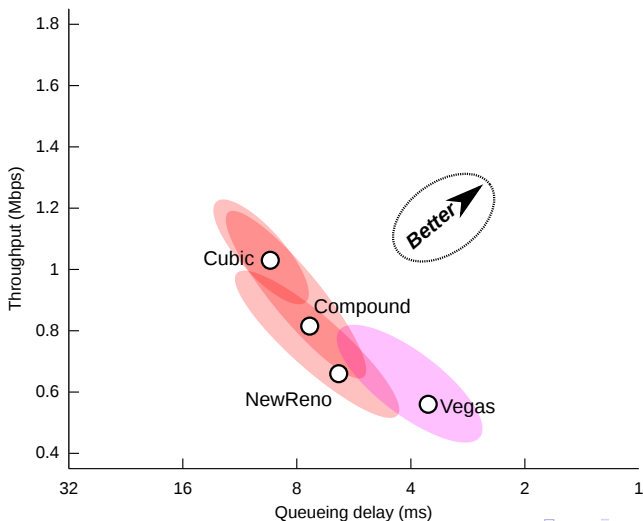
Scenario 1: throughput-delay plot



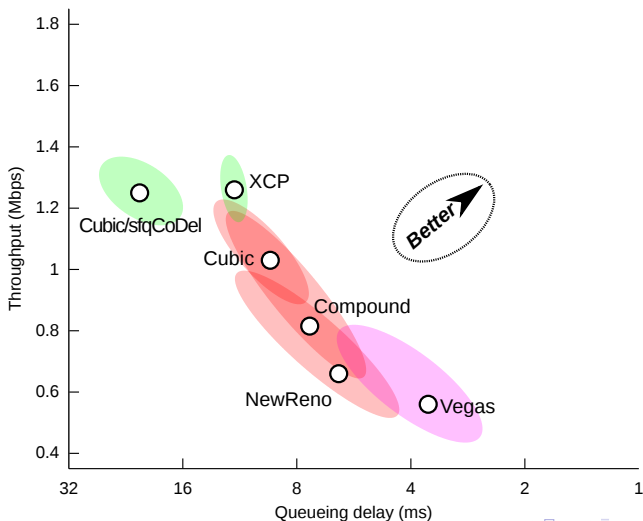
Scenario 1: throughput-delay plot



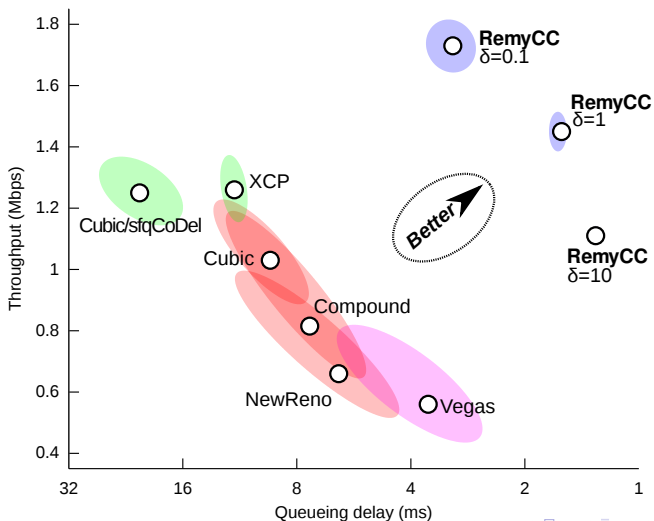
Scenario 1: throughput-delay plot



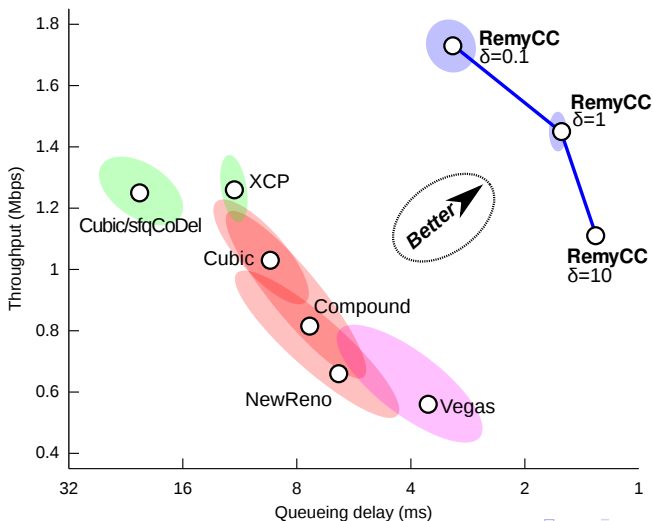
Scenario 1: throughput-delay plot



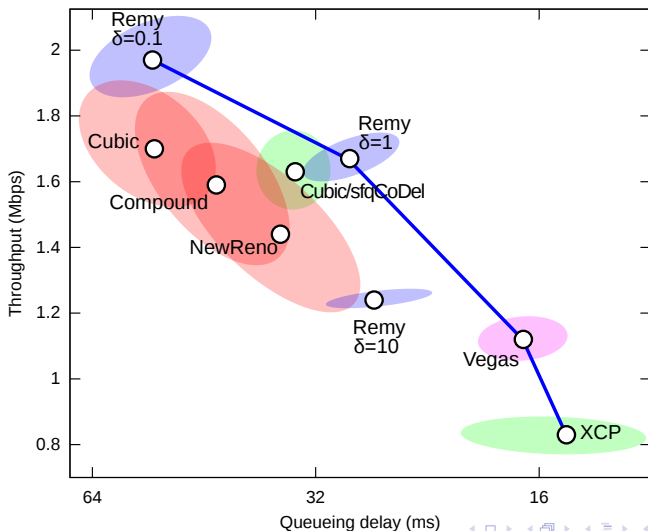
Scenario 1: throughput-delay plot



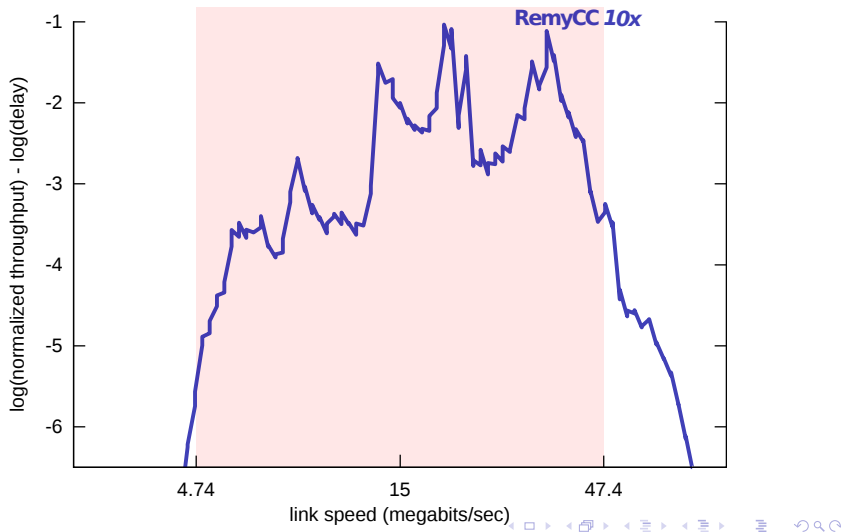
Scenario 1: throughput-delay plot



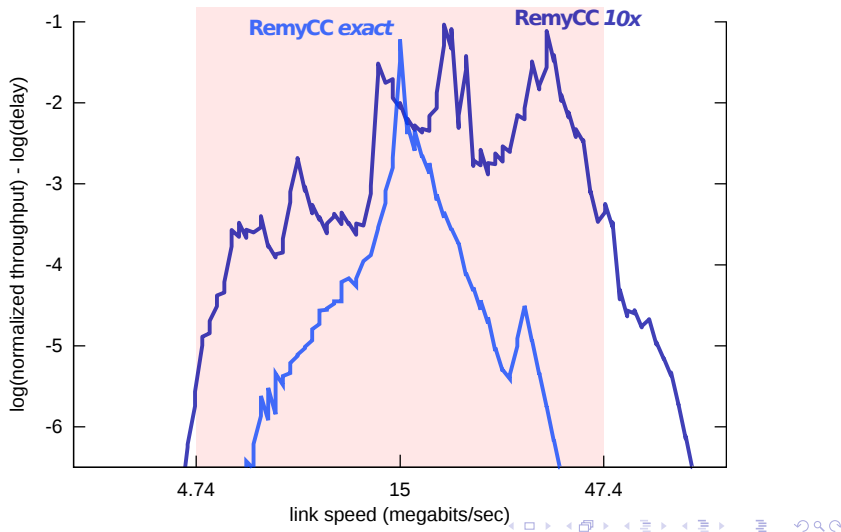
Scenario 2: Verizon LTE, $n = 8$



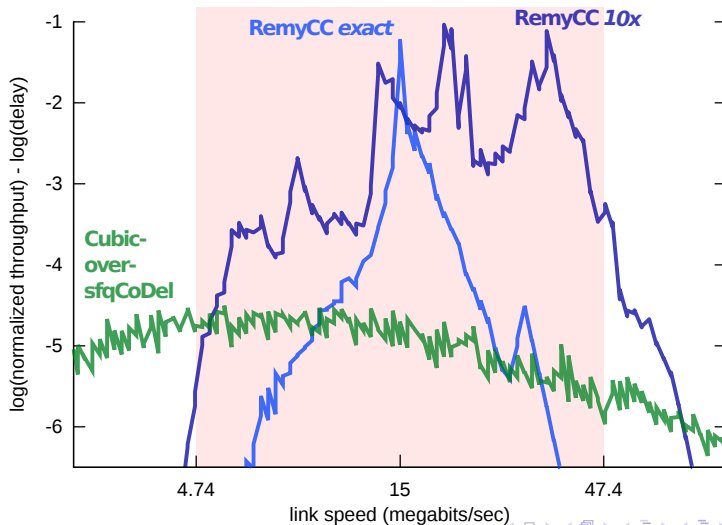
The effect of prior knowledge



The effect of prior knowledge



The effect of prior knowledge



Limitations and unknowns

- ▶ Tested only in simulation so far
- ▶ How to characterize robustness to the unforeseen?
- ▶ Can we make a RemyCC for a 10,000x range of throughputs?
- ▶ Agreeing on assumptions and goal may not be easy
- ▶ End-to-end = hard to handle an overaggressive competitor
- ▶ Not proposing Internet-scale deployment any time soon

Conclusions

- ▶ Traditionally: simple rules, complex behavior
- ▶ With Remy: complex rules, consistent behavior

- ▶ Computer-designed > human-designed
- ▶ End-to-end > in-network

- ▶ The network evolves. Transport should let it!

<http://web.mit.edu/remy>

{keithw, hari}@mit.edu

