neat

Horizon 2020
European Union funding
for Research & Innovation

Congestion Metacontrol to achieve a Deadline Aware Less than Best Effort service

David Hayes

together with David Ros, Andreas Petlund, Iffat Ahmed, Lars Erik Storbukås, and Hugo Wallenburg

ICCRG 13 November 2017

simula

Motivation

- Many bulk transfer applications do not need to send as fast as they can.
  - data-centre synchronisation
  - client to cloud backups
- They can send in a Less-than-Best-Effort (LBE) way
  - Avoid disrupting more quality constrained flows

Motivation

- Many bulk transfer applications do not need to send as fast as they can.
  - data-centre synchronisation
  - client to cloud backups
- They can send in a Less-than-Best-Effort (LBE) way
  - Avoid disrupting more quality constrained flows

Timeliness

- They often have some loose timeliness requirements
- A need for Deadline-Aware LBE (DA-LBE)

Deadline Aware Less than Best Effort (DA-LBE)

Transport qualities

- Keep disruption of concurrent BE interactive services to a minimum
  - **Do Good** – react to network congestion earlier than a BE service
- Have a timeliness constraint
  - **Be pragmatic** – adjust aggressiveness as deadline approaches
  - **Do no harm** – never more aggressive than a BE type service

Deadline Aware Less than Best Effort (DA-LBE)

Transport qualities

- Keep disruption of concurrent BE interactive services to a minimum
  - **Do Good** – react to network congestion earlier than a BE service
- Have a timeliness constraint
  - **Be pragmatic** – adjust aggressiveness as deadline approaches
  - **Do no harm** – never more aggressive than a BE type service

Our approach

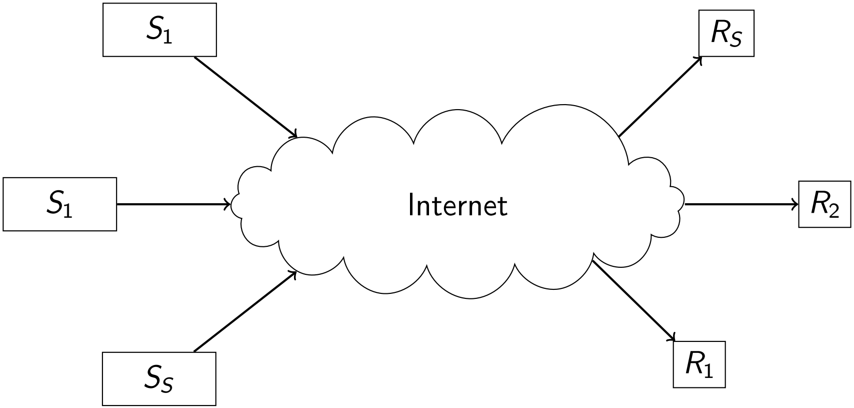
- Model this behaviour and develop a framework for enabling it

Conclusion: In principle the framework allows any E-to-E congestion control to become DA-LBE

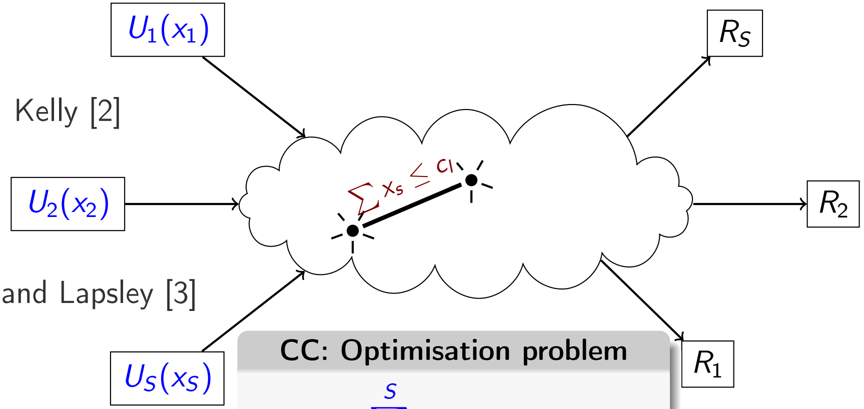
See our first publication:

D. A. Hayes, D. Ros, A. Petlund, and I. Ahmed, "A framework for less than best effort congestion control with soft deadlines", in *Proc. of IFIP Networking*, IFIP, Jun. 2017. [Online]. Available: `http://dl.ifip.org/db/conf/networking/networking2017/1570334752.pdf`

Network Utility Maximization (NUM)



Network Utility Maximization (NUM)



Kelly [2]

Low and Lapsley [3]

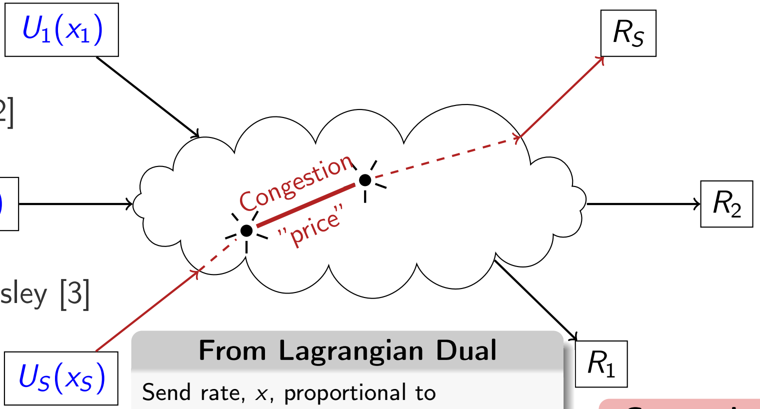
CC: Optimisation problem

$$\max_{x \geq 0} \sum_{s=1}^{S} U_s(x_s)$$

$$\text{s.t.} \sum_{s \in S(l)} x_s \leq c_l \qquad \forall l$$

[2] F. P. Kelly, "Charging and rate control for elastic traffic", *European Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997
[3] S. H. Low and D. E. Lapsley, "Optimization flow control–i: Basic algorithm and convergence", *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999

Network Utility Maximization (NUM)



Kelly [2]

Low and Lapsley [3]

From Lagrangian Dual

Send rate, x, proportional to "price" (congestion) and utility (CC)

$$\Rightarrow x_s = ((U_s)')^{-1}\left(\sum_{l \in L(s)} p_l\right)$$

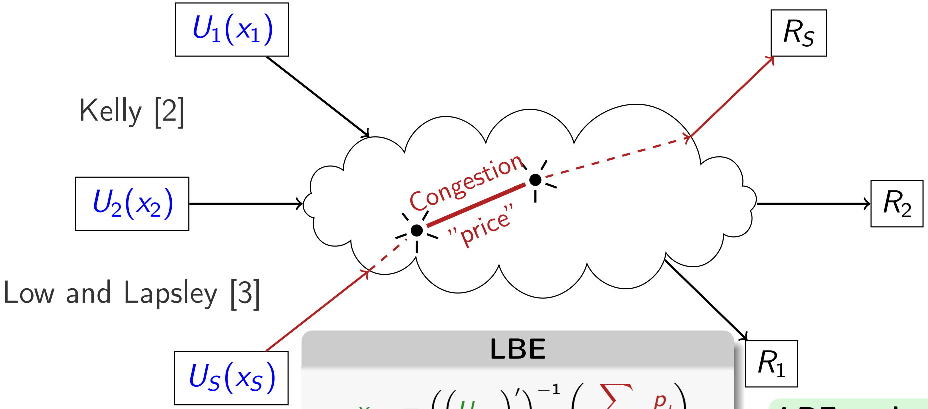
Congestion "price":

- loss, ECN,
- delay, etc

[2] F. P. Kelly, "Charging and rate control for elastic traffic", *European Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997

[3] S. H. Low and D. E. Lapsley, "Optimization flow control–i: Basic algorithm and convergence", *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999

Network Utility Maximization (NUM) for LBE



LBE

$$x_{lbe} = \left(\left(U_{lbe}\right)'\right)^{-1}\left(\sum_{l \in L(lbe)} p_l\right)$$

or

$$x_{lbe} = \left(\left(U\right)'\right)^{-1}\left(w_{lbe}\sum_{l \in L(lbe)} p_l\right)$$

LBE options

- special utility U_{lbe} (ie CC algorithm)
- inflate price – w_{lbe}

[2] F. P. Kelly, "Charging and rate control for elastic traffic", *European Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997
[3] S. H. Low and D. E. Lapsley, "Optimization flow control–i: Basic algorithm and convergence", *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999

DA-LBE with a homogeneous congestion control network

Congestion "price" inflation

$$\hat{q} = \frac{\sum p_l}{w} \quad \text{where } w \in [w_{\min}, 1]$$

- when $w = w_{\min}$, maximum price inflation, maximum "LBEness"
- when $w = 1$, no inflation, BE service.

DA-LBE with a homogeneous congestion control network

Congestion "price" inflation

$$\hat{q} = \frac{\sum p_l}{w} \quad \text{where } w \in [w_{\text{min}}, 1]$$

- when $w = w_{\text{min}}$, maximum price inflation, maximum "LBEness"
- when $w = 1$, no inflation, BE service.

Controlling price with respect to deadlines

On short packet time scales:

- CC reacts to $\hat{q}$ as normal

On longer time scales adjust w:

- Relative to:
  - recent send rate: $\bar{x}$
  - required send rate: ζ
- PID or Model based control

Applying this to TCP Cubic

For TCP Cubic, price is packet loss or ECN packet marks

Indirect: inflate response

Direct: inflate price

Applying this to TCP Cubic

For TCP Cubic, price is packet loss or ECN packet marks

Indirect: inflate response

Change **cwnd** reduction

- $\text{cwnd}_{\text{new}} = \beta \text{cwnd}$
- Vary β, $\beta \in [\beta_{\text{min}}, \beta_{\text{default}}]$
- β_{min} provides maximum LBEness

Direct: inflate price

Drop additional packets

- lose data
- *causes retransmissions*

Phantom ECN signals

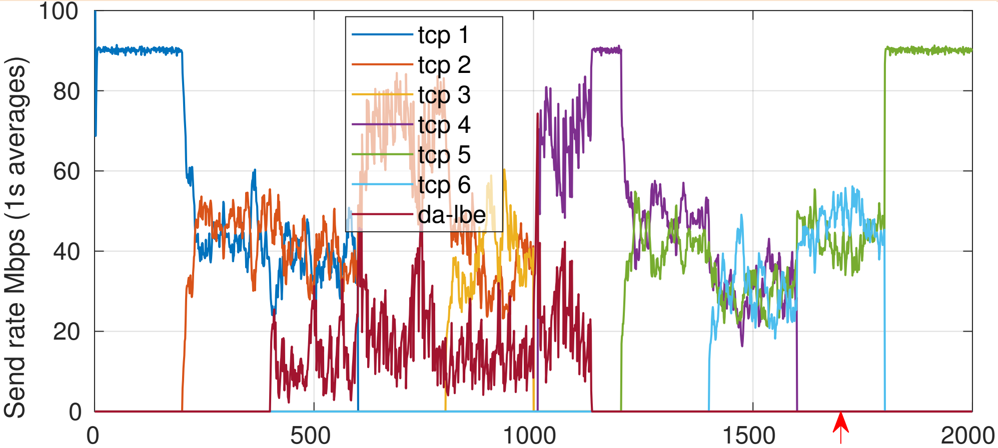
- same congestion response
- *no loss in data*

Simple Scenario Experiments

Scenario

- 6 TCP flows start and stop at different overlapping times
  - No competing TCP flows t=[1000,1010] s
- DA-LBE flow file size equivalent to 10% capacity to deadline
- 10% random background traffic

TCP Cubic with a Cubic based DA-LBE — varying β

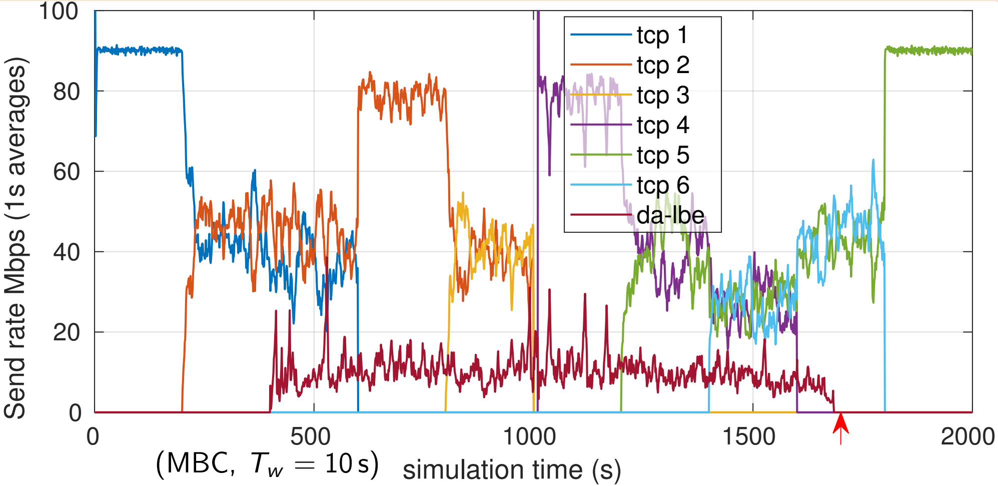


(MBC, $T_w = 10\,\text{s}$)

Lack of LBEness

The loss signal is not frequent enough

TCP Cubic with a Cubic based DA-LBE — Phantom ECN



Unable to take full advantage of available capacity ($t = 1000\,\mathrm{s}$)

Delay-based CC may detect congestion or lack of it more quickly

A network of heterogeneous CCs: Different Prices and Utilities

Different network prices

- Tang, Wei, Low, and Chiang [5] maps prices to a *standard* price (or congestion signal).
  - E.g. mapping a packet delay "price" to a standard packet loss price

Issues

- Tang, Wei, Low, and Chiang required a special factor to make this work.
- More than mapping prices, CCs react differently to congestion signals

[5] A. Tang, X. Wei, S. H. Low, and M. Chiang, "Equilibrium of heterogeneous congestion control: Optimality and stability", *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 844–857, Jun. 2010

A network of heterogeneous CCs: Different Prices and Utilities

Different network prices

- Tang, Wei, Low, and Chiang [5] maps prices to a *standard* price (or congestion signal).
  - E.g. mapping a packet delay "price" to a standard packet loss price

Issues

- Tang, Wei, Low, and Chiang required a special factor to make this work.
- More than mapping prices, CCs react differently to congestion signals

We build on this idea

- Composite congestion signals (delay, loss, and ECN)
- Weight (ϕ) composite congestion signals by CC reaction
- We use a weighted $\mathbb{P}$[cong_ind] to compare "prices"

[5] A. Tang, X. Wei, S. H. Low, and M. Chiang, "Equilibrium of heterogeneous congestion control: Optimality and stability", *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 844–857, Jun. 2010

Applying this to TCP Vegas

Delay based part

- Congestion signal:
  - Estimate of queueing delay (Q)
- Control:
  - cwnd++ or cwnd−−

Loss based part

- Halve cwnd on packet loss

Applying this to TCP Vegas

Delay based part

- Congestion signal:
  - Estimate of queueing delay (Q)
- Control:
  - cwnd++ or cwnd−−

Loss based part

- Halve cwnd on packet loss

Vegas based DA-LBE

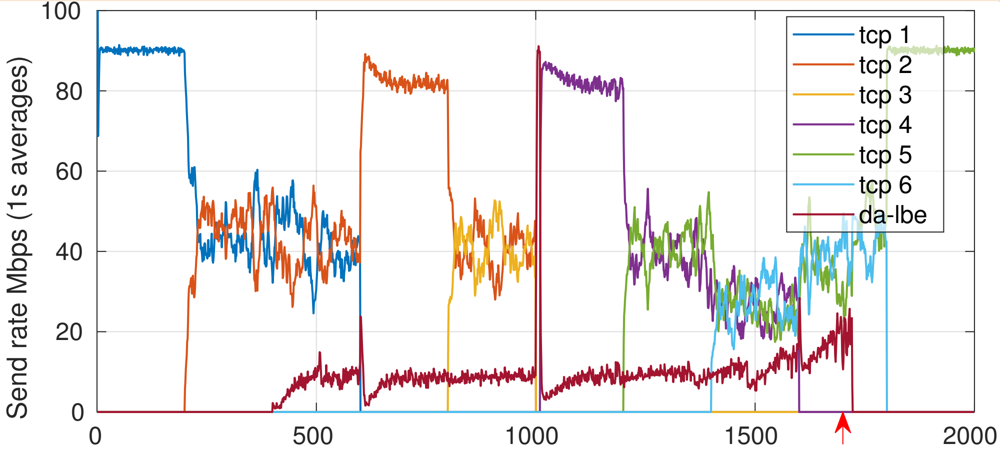
Delay based part

- inflate (or deflate) queueing delay
  - $\hat{q} = \frac{Q}{\phi w}$

Loss based part

- When $w = 1$ and packet loss
  - probabilistically ignore cwnd reduction
  - rand()$< (1 - \frac{1}{w\phi})$

TCP Cubic competing with a Vegas based DA-LBE

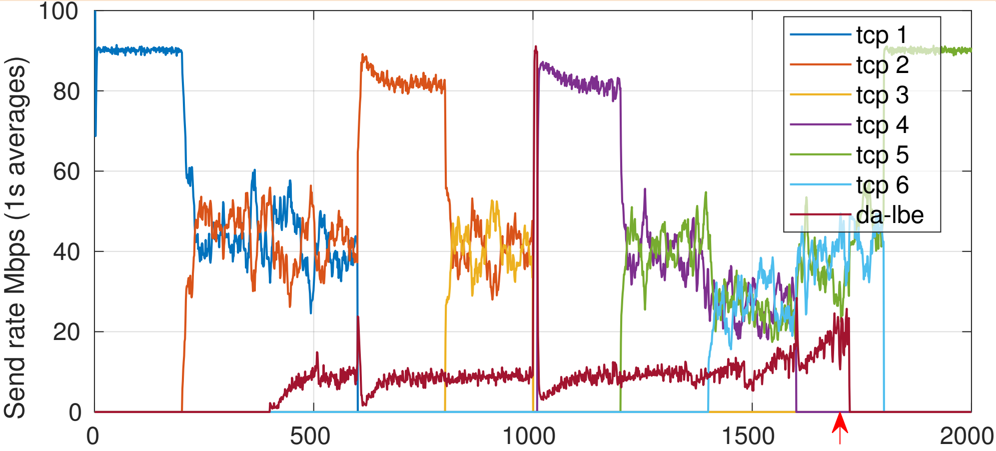


(MBC, $T_w = 10$, $T_\phi = 60\,s$) simulation time (s)

Able to take full advantage of available capacity ($t = 1000\,s$)

- Delay-based CC detects congestion *or lack of it* more quickly

TCP Cubic competing with a Vegas based DA-LBE

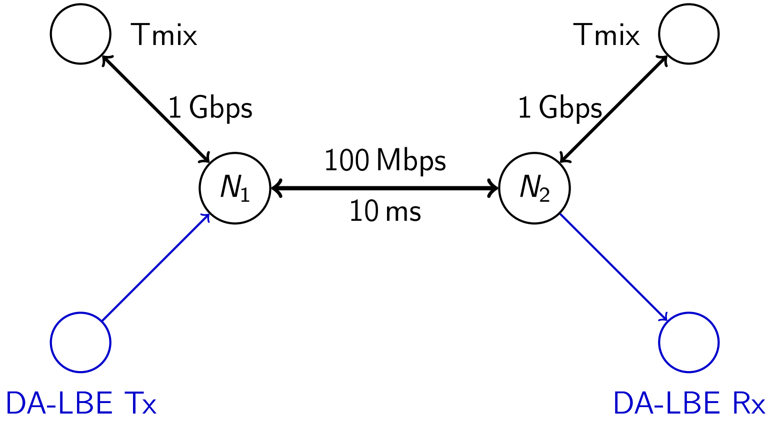


(MBC, $T_w = 10$, $T_\phi = 60\,\mathrm{s}$)

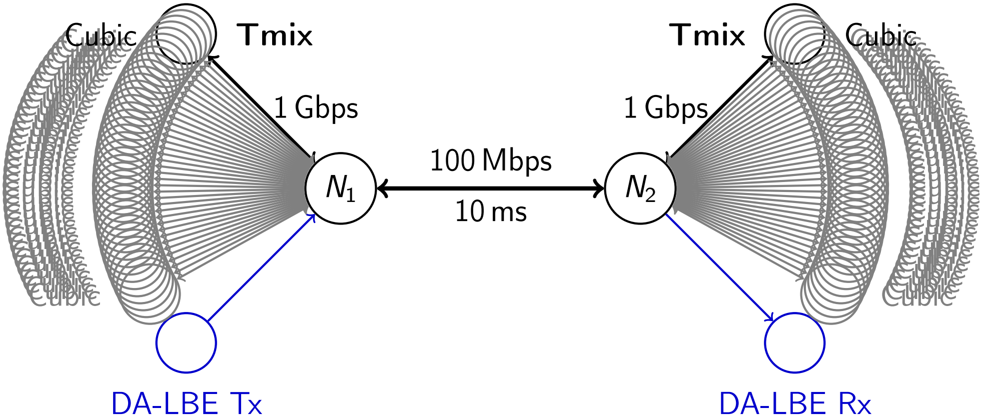
Does not quite meet deadline in this scenario

- But does no harm
- LBEness relatively good

Completion time experimental setup



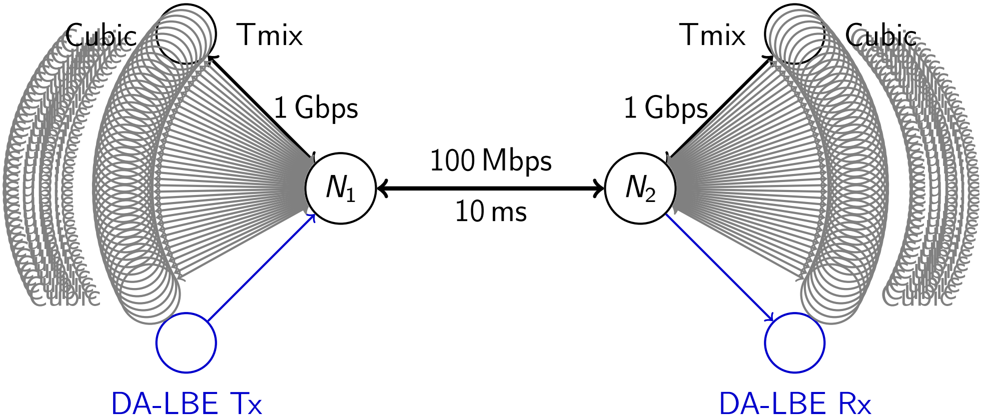
Completion time experimental setup



Tmix

- Based on real measurements
- Models application interaction
- TCP connections come and go
  - hundreds of concurrent connections

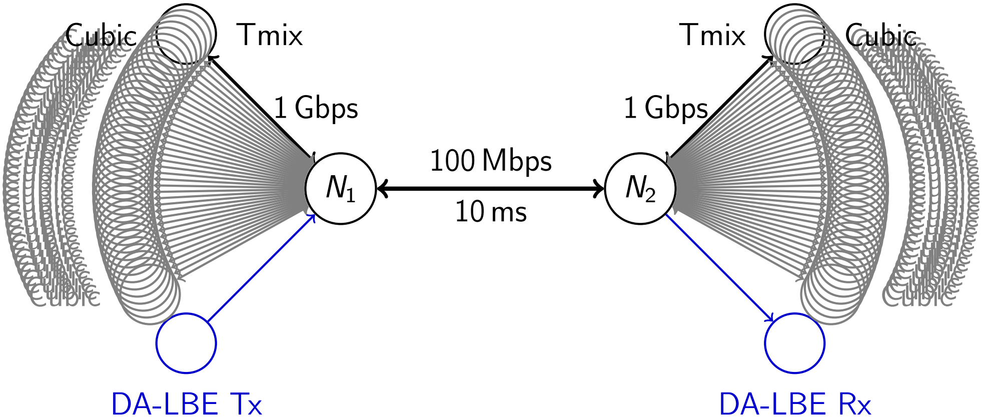
Completion time experimental setup



Tmix in Experiment

- Shuffle traces to remove non-stationarity
  - Application session start times
- Scale application session arrival time to achieve a particular average *offered* load.

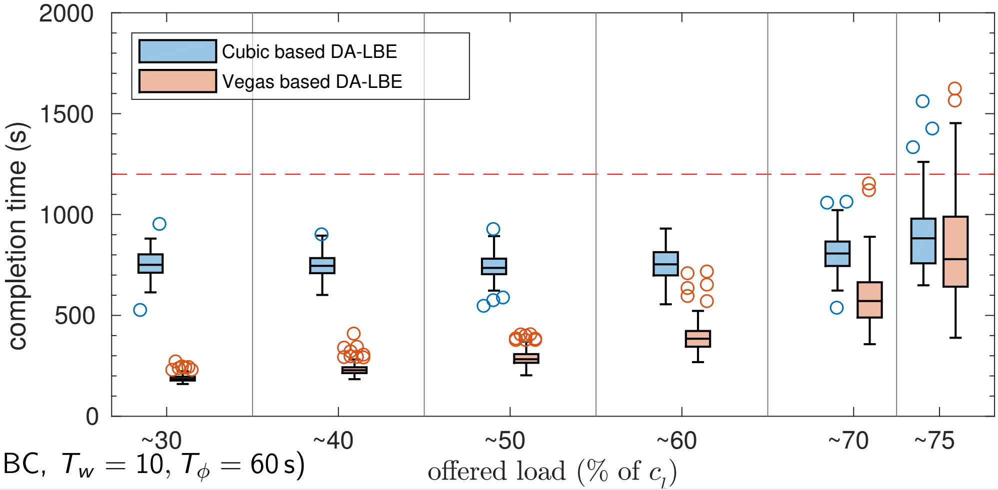
Completion time experimental setup



Experiment comparing Cubic and Vegas based DA-LBE

- NS2
- $> 60\,000$ s
- Deadline 1200 s
- DA-LBE flows restart 10 s after completing
- File size equivalent to 10% of Capacity over 1200 s

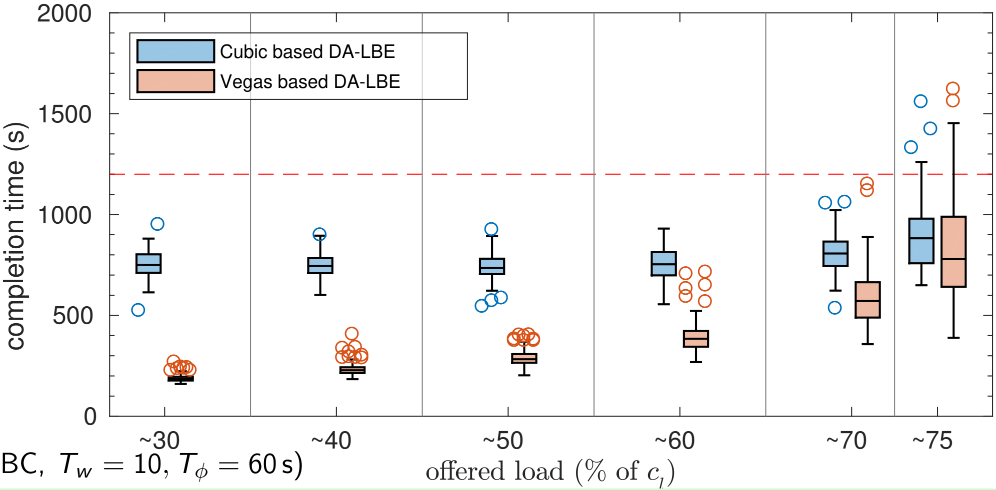
DA-LBE Completion time results



(MBC, $T_w = 10$, $T_\phi = 60\,\text{s}$)

- Cubic based is more consistent, but less able to use available capacity
- 75% offered load is near congestion "knee"
  - Both do not always meet deadlines

DA-LBE Completion time results



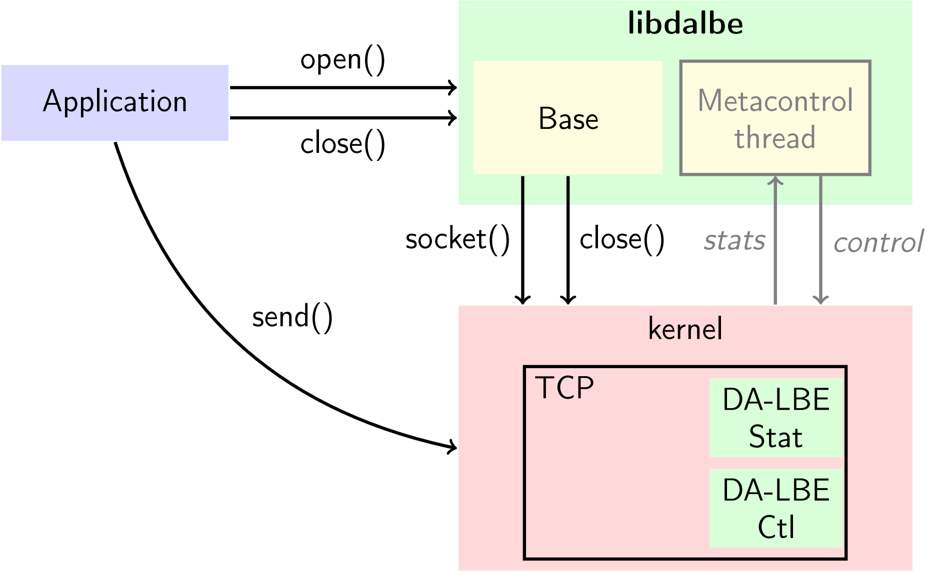
(MBC, $T_w = 10$, $T_\phi = 60\,\mathrm{s}$)

We have also looked at:

Refer to next publication :)

- impact of completion times on other traffic
- relative fairness, etc.

Stand alone DA-LBE in Linux (*mostly* working prototype)



DA-LBE in NEAT

What is NEAT?

- A new transport API (see work in the TAPS WG!)
  - applications request the service they need
  - agnostic to the specific choice of transport protocol underneath
- Allows deployment of new (and better) transports

DA-LBE in NEAT

What is NEAT?

- A new transport API (see work in the TAPS WG!)
  - applications request the service they need
  - agnostic to the specific choice of transport protocol underneath
- Allows deployment of new (and better) transports

Read more about it: `https://www.neat-project.org/`

[6] N. Khademi, D. Ros, M. Welzl, Z. Bozakov, A. Brunstrom, G. Fairhurst, K.-J. Grinnemo, D. Hayes, P. Hurtig, T. Jones, S. Mangiante, M. Tüxen, and F. Weinrank, "NEAT: A Platform- and Protocol-Independent Internet Transport API", *IEEE Commun. Mag.*, Jun. 2017. [Online]. Available: `https://www.neat-project.org/wp-content/uploads/2017/03/commag16-accepted-version.pdf`

DA-LBE in NEAT

What is NEAT?

- A new transport API (see work in the TAPS WG!)
  - applications request the service they need
  - agnostic to the specific choice of transport protocol underneath
- Allows deployment of new (and better) transports

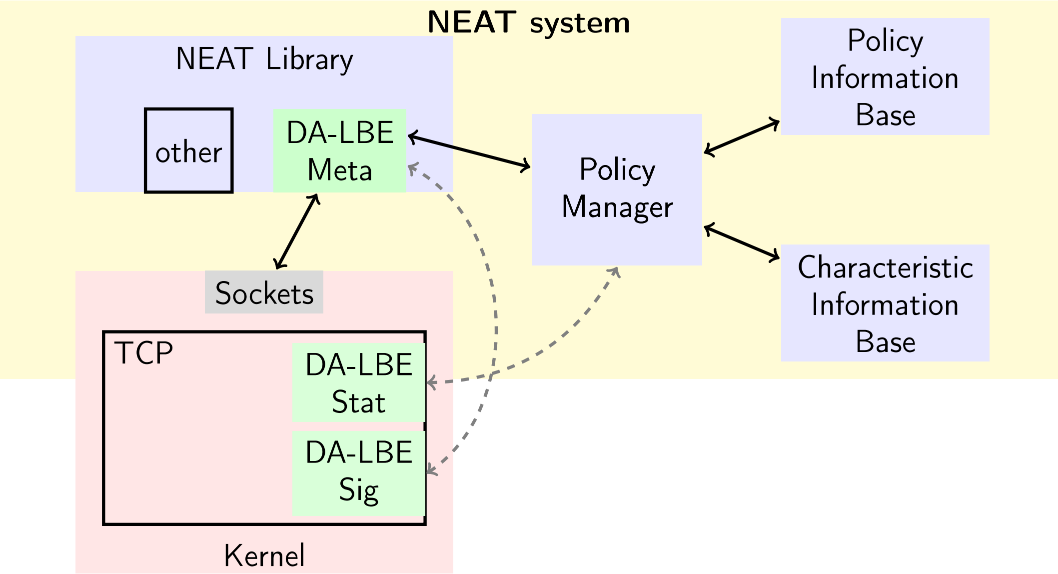
Read more about it: `https://www.neat-project.org/`

[6] N. Khademi, D. Ros, M. Welzl, Z. Bozakov, A. Brunstrom, G. Fairhurst, K.-J. Grinnemo, D. Hayes, P. Hurtig, T. Jones, S. Mangiante, M. Tüxen, and F. Weinrank, "NEAT: A Platform- and Protocol-Independent Internet Transport API", *IEEE Commun. Mag.*, Jun. 2017. [Online]. Available: `https://www.neat-project.org/wp-content/uploads/2017/03/commag16-accepted-version.pdf`

DA-LBE in NEAT

- DA-LBE will be implemented as a meta-protocol in NEAT
- NEAT choosing the best underlying transport to adapt

DA-LBE meta-protocol in NEAT (work in progress)



Conclusions

Deadline-Aware–Less-than-Best-Effort (DA-LBE)

- valuable transport for bulk data transfers
  - soft deadline
  - disruption of other traffic minimised

In principle allows *any* congestion control to become DA-LBE

- Concepts based on NUM
  - *inflate* (or *discount*) network "prices" to achieve goals.
- Tested with TCP Cubic and Vegas
  - Delay based mechanisms generally perform better
  - Immediate ECN would have benefits of delay based mechanisms

Ongoing work

- Integration into NEAT
- Trans-Internet tests
- Modularisation of kernel elements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644334 (NEAT). The views expressed are solely those of the authors.

Bibliography

D. A. Hayes, D. Ros, A. Petlund, and I. Ahmed, "A framework for less than best effort congestion control with soft deadlines", in *Proc. of IFIP Networking*, IFIP, Jun. 2017. [Online]. Available: http://dl.ifip.org/db/conf/networking/networking2017/1570334752.pdf.

F. P. Kelly, "Charging and rate control for elastic traffic", *European Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997.

S. H. Low and D. E. Lapsley, "Optimization flow control–i: Basic algorithm and convergence", *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

N. Trichakis, A. Zymnis, and S. Boyd, "Dynamic network utility maximization with delivery contracts", in *Proc. of IFAC World Congress*, Seoul, South Korea, Jul. 2008, pp. 2907–2912.

A. Tang, X. Wei, S. H. Low, and M. Chiang, "Equilibrium of heterogeneous congestion control: Optimality and stability", *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 844–857, Jun. 2010.

N. Khademi, D. Ros, M. Welzl, Z. Bozakov, A. Brunstrom, G. Fairhurst, K.-J. Grinnemo, D. Hayes, P. Hurtig, T. Jones, S. Mangiante, M. Tüxen, and F. Weinrank, "NEAT: A Platform- and Protocol-Independent Internet Transport API", *IEEE Commun. Mag.*, Jun. 2017. [Online]. Available: https://www.neat-project.org/wp-content/uploads/2017/03/commag16-accepted-version.pdf.