Fair Background Data Transfers of Minimal Delay Impact

C. Courcoubetis
A. Dimakis
AUEB
• Propose a framework for the design of protocols for background transfers
• File sizes differ by >10 orders of magnitude, connection rates by few orders of magnitude
• Main concern: how to obtain reasonable throughput with minimal delays on short flows
• Current approaches: TCP-nice, LEDBAT,...
  – behave as second priority traffic (low impact on short flows)
  – no consideration of fairness relative to other long flows
  – no adoption incentives
• Related work: Key, Massoulie, etc.
  – substitution of all long TCPs by on-off senders based on threshold price
  – prove that there is some delay improvement
  – mostly a different traffic model, assumes all flows to convert to new protocol
The competition environment

- No competing long TCP: easy case!! FB: 2\textsuperscript{nd} priority
- 1 long TCP: FB 2\textsuperscript{nd} priority => zero throughput
- existing solutions: > 2\textsuperscript{nd} priority, unspecified throughput

\[ \rho C (1 - \rho)C \]
Adoption incentives vs “niceness”

• Why “long TCP” users adopt FB instead of TCP?
• Which are sensible properties of FBs?
  – When competing with long TCPs for $C(1 - \rho)$:
    – obtain a given fraction $f$ of $C(1 - \rho)$,
    – cause minimum extra delays on short flows
  – Example: obtain same average throughput as TCP $f = \frac{l}{k+l}$
• Achieve all that with reasonable context information
  – public Internet context, competition with non-local flows
Our results

• Obtain optimal BW sharing policy under **complete information**
  – minimize delays on short flows while competing with $k$ long TCPs and obtaining a share $f$ of the leftover capacity

• Implementable approximation: weighted TCP
  – short time scales: use $w$-TCP
  
  $$f = \frac{w}{k+w}$$

  – delay deterioration  $\leq 17.2\%$ for $k = 1$, 
  
  $\downarrow 0$ as $k \to \infty$

• Delay impact $\delta(f)$
Our results (cont.)

• Adopt Kelly’s approach for fairness, but for long-term
  – FBs don’t get a fixed fraction $f$ of leftover capacity
  – max some social welfare function
    • sum of utilities for average throughput of long flows TCPs, FBs
    • add a negative externality term (extra delay to short flows)
  – implementation: using w-TCPs
    • don’t need to know $k, C, \rho, ...$

• Use this framework to design new protocols
  – examples: $y_0 / y_i = 1 + \gamma y_0^{-1}$
  
  $y_0 / y_i = 1 + \gamma y_0$
The model

\[ \rho = \frac{\lambda}{(\mu C)} \]

short TCP flows

\[ k \text{ long TCP flows} \]

\[ l \text{ Fairbat flows} \]

excess capacity = \( C(1 - \rho) \)

\[ C(1 - \rho)f \]

\[ x_n = \text{download bw of TCP flows when } n \text{ short calls are active} \]

\[ n \rightarrow \begin{cases} 
  n+1, & \text{with rate } \lambda, n \geq 0, \\
  n-1, & \text{with rate } \mu nx_n, n \geq 1 
\end{cases} \]

\[ N(k, f, \rho) = \text{average # of short flows} \]

delay impact: \[ \delta = \frac{N(k, f, \rho)}{C \rho} - \frac{1}{C(1 - \rho)} \]

Problem: Find optimal \( x_n^*, n = 0, 1, \ldots \)
The optimization problem:

\[ N_*(k, f, \rho) = \min_{x_n, \pi_n} \sum_{n=0}^{\infty} n\pi_n \]

such that:

\[ \lambda \pi_{n-1} = \mu nx_n \pi_n, n = 1, 2, \ldots \]

\[ \sum_{n=0}^{\infty} \pi_n = 1 \]

\[ x_n \leq \frac{C}{k+n}, n = 0, 1, \ldots \]

\[ \sum_{n=0}^{\infty} x_n \pi_n = \frac{C(1-\rho)(1-f)}{k} \]

over \( x_n \geq 0, \pi_n \geq 0, n = 0, 1, \ldots \)

Optimal policy:

If \((1-\rho)^k \leq f\), \(x_n = \begin{cases} 
 0 & \text{for each } n \leq n_* , \\
 \frac{C}{k + n_* + 1 + \varepsilon} \geq 0 & \text{for each } n = n_* + 1 , \\
 \frac{C}{k + n} & \text{for each } n \geq n_* + 2 
\end{cases} \)
Computing optimal thresholds:

\[ E\left(n_* + k, k, \frac{1 - \rho}{\rho}\right) \leq f < E\left(n_* + 1 + k, k, \frac{1 - \rho}{\rho}\right), \quad E(m, q, r) = \frac{\binom{m}{q} r^q}{\sum_{i=0}^{q} \binom{m}{i} r^i}, m \geq q \]

\[ N_* = \frac{(k+1)\rho}{1 - \rho} + n_* E\left(n_* + k, k, \frac{1 - \rho}{\rho}\right) \]

Approximations as \( \rho \to 1 \):

\[ n_*(1 - \rho) \to a_f, \quad B(k, a_f) = f, \]

\[ N_*(k, f, \rho)(1 - \rho) \to k + 1 + a_f f \]
Using w-TCP

- How bad is using w-TCP compared to optimal?
- The relative error vanishes as $k \uparrow$, uniformly over $f$

\[ N_w(k, f, \rho) = \frac{\rho}{1 - \rho} \left( k + 1 + \frac{k f}{1 - f} \right), \quad f = \frac{W}{k + W} \]

For $0 \leq f \leq 1$

\[
\lim_{\rho \to 1} \frac{N_w(k, f, \rho) - N^*(k, f, \rho)}{N_w(k, f, \rho)} \leq B(k - 1, a_f) - B(k, a_f)
\]

\[ \sup_{0 \leq f \leq 1} \left[ B(k - 1, a_f) - B(k, a_f) \right] \downarrow 0 \text{ as } k \to \infty \]

max error $\leq 17.2\%$ for $k = 1$
A corollary

- If we substitute any subset of long TCP flows by “equivalent” optimal FBs, the max improvement of the delay of short flows is less than 17.2%
- The best improvement is achieved when there is competition of 1FB and 1 long TCP flow
- A negative result?
- The incentive compatibility constraint (obtain same average throughput as TCP) in larger systems implies small optimal delay improvements
- To get significant delay improvement we need to relax the IC condition (how?)
The general fairness framework

- Problem: “fair” share of excess capacity
- Express fairness on long-term rates “à la Kelly”
- Take into account delay spillovers to short flows
  - remember the tradeoff \( f \leftrightarrow \text{delay} \)
- Engineering: translate into flow control algorithms
  - decompose controls for short and long timescales
  - make reasonable assumptions on what is known locally
- Reverse engineering: translate existing algorithms into this model

\[
z = C(1 - \rho)f
\]

FBs

TCPs

\( C(1 - \rho) \)

\( \delta(z) \)

\( z \leftarrow 0 \)

\( f \leftarrow 0 \)

\( z \leftarrow C(1 - \rho) \)

\( f \leftarrow 1 \)
The optimization problem

\[
\max \quad ku_0(y_0) + \sum_{i=1}^{l} u_i(y_i) - \int_{0}^{\sum_{i=1}^{l} y_i} \delta(\sum_{i=1}^{n} z_i)^2 F(\delta(z)) dz
\]

such that \( ky_0 + \sum_{i=1}^{l} y_i = C(1 - \rho) \),

over \( y_0, \ldots, y_l \geq 0 \)

Assume that a proportional sharing policy is used in the short t.s. Then magic!!!

\[
\delta_w \left( \sum_{i=1}^{l} y_i \right) = \frac{1}{y_0}
\]

Optimality condition for long ts: \( -u'_0(y_0) + u'_i(y_i) - \frac{1}{y_0^2} F \left( \frac{1}{y_0} \right) = 0, i = 1, \ldots, l \)

Short ts: use w-TCP

Long ts: adapt the weights

\[
\dot{w}_i = -u'_0 \left( \frac{y_i}{w_i} \right) + u'_i(y_i) - \left( \frac{w_i}{y_i} \right)^2 F \left( \frac{w_i}{y_i} \right), i = 1, \ldots, l
\]
Engineering new protocols

\[
\max \quad ku_0(y_0) + \sum_{i=1}^{l} u_i(y_i) - \int_{0}^{\sum_{i=1}^{l} y_i} \delta_w(z)^2 F(\delta(z))dz
\]

\[
-u_0'(y_0) + u_i'(y_i) - \frac{1}{y_0^2} F\left(\frac{1}{y_0}\right) = 0, \quad i = 1, \ldots, l
\]

such that \( ky_0 + \sum_{i=1}^{l} y_i = C(1 - \rho) \),

\[
\Leftrightarrow \dot{w}_i = -u_0'(\frac{y_i}{w_i}) + u_i'(y_i) - \left(\frac{w_i}{y_i}\right)^2 F\left(\frac{w_i}{y_i}\right), \quad i = 1, \ldots, l
\]

over \( y_0, \ldots, y_l \geq 0 \), w-TCP short ts controls

**Case A:**

\[
u_i(y) = \log y, \quad i = 0, \ldots, l, \quad F(\delta) = \gamma
\]

\[
y_0 = 1 + \gamma y_0^{-1}, \quad i = 1, \ldots, l
\]

**Case B:**

\[
u_i(y) = \log y, \quad i = 0, \ldots, l, \quad F(\delta) = \gamma \delta^{-2}
\]

\[
y_0 = 1 + \gamma y_0, \quad i = 1, \ldots, l
\]
Algorithm A

\[
\frac{y_0}{y_i} = 1 + \gamma y_0^{-1}, \quad i = 1, \ldots, l
\]

- FBs get similar throughput as TCP when there is enough excess bandwidth, give away when it becomes scarce
- IC condition relaxed when resources are scarce (second priority when sensible)
Algorithm B

\[
\frac{y_0}{y_i} = 1 + \gamma y_0, \quad i = 1, \ldots, l
\]

- FBs get similar throughput as TCP when excess bandwidth is scarce, give away when there is enough bandwidth.
- IC condition relaxed when resources are abundant (risk averse?)
Reverse engineering

\[ \rho \]

average price = \( q \)

\[
q = \frac{\text{average cost}}{\text{average total rate}} = \frac{\beta \cdot \text{average loss rate}}{\text{average total rate}} = \frac{\beta w_i E[p_n x_n]}{w_i E x_n} \approx \frac{\beta 2 \pi}{T^2 y_0^2}
\]

\( FB_i \) achieves: \( u'_i(y_i) = q = \frac{\beta 2 \pi}{T^2 y_0^2} \)

Same conditions as solving \(-u'_0(y_0) + u'_i(y_i) - \frac{1}{y_0^2} F\left(\frac{1}{y_0}\right) = 0\)

for \( u_0(y_0) = -\beta \pi / (T^2 y_0) \), \( F(\delta) = \beta \pi / T^2 \)
Conclusions

• Protocols for background transfers operate in the context of other long and short TCP flows
• TCP is the incumbent protocol, new protocols should compare to TCP
• We derived the optimal short time scale policy for achieving a given share of long term throughput, but has practical implementation issues
• w-TCP seems a reasonable practical alternative, provably small efficiency loss
• We provided a utility-based definition for fair sharing including a negative externality term for delay caused to short flows
• We derived two new interesting protocols for background transfers by relaxing the IC condition for adoption relative to TCP