Participatory Networking

Andrew Ferguson, Arjun Guha, Jordan Place, Rodrigo Fonseca, and Shriram Krishnamurthi
The Problem with Networks
1. in the home

The Problem with Networks
1. in the home
2. in the enterprise
1. in the home
2. in the enterprise
3. in the cloud
1. in the home
2. in the enterprise
3. in the cloud
4. in the datacenter
A problem in the home
Network Working Group
Request for Comment: 2205
Category: Standards Track

Abstract
This memo describes version 1 of RSVP, a resource reservation protocol designed for an integrated services Internet. RSVP provides receiver-initiated setup of resource reservations for multicast or unicast data flows, with good scalability and robustness properties.

Braden, Ed., et. al. Standards Track [Page 1]
RFC 2205 RSVP September 1997
TCP Nice: A Mechanism for Background Transfers

Arun Venkataramani
Ravi Kokkala
Mike Dralil

Laboratory of Advanced Systems Research
Department of Computer Sciences
University of Texas at Austin, Austin, TX 78712
arun, ravi, miked@cs.utexas.edu

Abstract
Many distributed applications can make use of large background transfers - transfers of data that humans are not waiting for to improve availability, reliability, latency or convenience. However, given the rapid fluctuations of available network bandwidth and changing resource costs due to technology trends, hardwiring the aggressiveness of background transfers (1) simplifying applications, (2) being too aggressive and saturating the networking layer, and (3) not getting the benefits of background transfers. Our goal is for the operating system to manage network resources in order to provide a simple abstraction of near-zero cost background transfers. Our system, TCP Nice, can provably bound the interference inflicted by background flows on foreground flows in a constrained network model. And our microbenchmark and case study applications suggest that TCP Nice's interference is tolerable. TCP Nice's performance and efficiency is enabled by its use of standard network operations.

I Introduction

Many distributed applications can make use of large background transfers - transfers of data that humans are not waiting for to improve availability, reliability, latency or convenience. For example, a broad range of applications and services such as data backup [25], prefetching [31], enterprise data distribution [28], Internet content distribution [2] and peer-to-peer storage [10, 43] can benefit from TCP Nice.

TCP Nice provides a simple abstraction of large background transfers. We call it a "background transfer" if the background transfer is on a separate network connection or if the background transfer is not a single large file transfer. TCP Nice is designed to improve the performance and efficiency of background transfers while maintaining the quality of service for foreground applications. The goal of TCP Nice is to provide a simple abstraction of large background transfers. TCP Nice provides a simple abstraction of large background transfers and can be used in conjunction with other background transfers.

TCP Nice: A Mechanism for Background Transfers

Arun Venkataramani
Ravi Kokkala
Mike Dralil

Laboratory of Advanced Systems Research
Department of Computer Sciences
University of Texas at Austin, Austin, TX 78712
arun, ravi, miked@cs.utexas.edu

Abstract
Many distributed applications can make use of large background transfers - transfers of data that humans are not waiting for to improve availability, reliability, latency or convenience. However, given the rapid fluctuations of available network bandwidth and changing resource costs due to technology trends, hardwiring the aggressiveness of background transfers (1) simplifying applications, (2) being too aggressive and saturating the networking layer, and (3) not getting the benefits of background transfers. Our goal is for the operating system to manage network resources in order to provide a simple abstraction of near-zero cost background transfers. Our system, TCP Nice, can provably bound the interference inflicted by background flows on foreground flows in a constrained network model. And our microbenchmark and case study applications suggest that TCP Nice's interference is tolerable. TCP Nice's performance and efficiency is enabled by its use of standard network operations.

I Introduction

Many distributed applications can make use of large background transfers - transfers of data that humans are not waiting for to improve availability, reliability, latency or convenience. For example, a broad range of applications and services such as data backup [25], prefetching [31], enterprise data distribution [28], Internet content distribution [2] and peer-to-peer storage [10, 43] can benefit from TCP Nice.

TCP Nice provides a simple abstraction of large background transfers. We call it a "background transfer" if the background transfer is on a separate network connection or if the background transfer is not a single large file transfer. TCP Nice is designed to improve the performance and efficiency of background transfers while maintaining the quality of service for foreground applications. The goal of TCP Nice is to provide a simple abstraction of large background transfers. TCP Nice provides a simple abstraction of large background transfers and can be used in conjunction with other background transfers.
A problem in the enterprise
A problem in the cloud
Production Platform

Based on “Delusional Boot: Securing Cloud Hypervisors without Massive Re-Engineering” (EuroSys 2012)
Production Platform
Production Platform

Based on “Delusional Boot: Securing Cloud Hypervisors without Massive Re-Engineering” (EuroSys 2012)
Based on “Delusional Boot: Securing Cloud Hypervisors without Massive Re-Engineering” (EuroSys 2012)
Based on “Delusional Boot: Securing Cloud Hypervisors without Massive Re-Engineering” (EuroSys 2012)
Production Platform
Based on “Delusional Boot: Securing Cloud Hypervisors without Massive Re-Engineering” (EuroSys 2012)
A problem in the datacenter
Participatory Networking
Participatory Networking
Participatory Networking
Participatory Networking

1. Requests

PANE
Participatory Networking

1. Requests
2. Hints

PANE
Participatory Networking

1. Requests
2. Hints
3. Queries

PANE
Participatory Networking
Participatory Networking

Safe?
Participatory Networking

Safe?  Secure?
Participatory Networking

Safe?  Secure?  Fair?
Participatory Networking

Safe?    Secure?    Fair?    Practical?
Participatory Networking

Safe?  Secure?  Fair?
Practical?  Efficient?
Participatory Networking
Participatory Networking

- End-user API for SDNs
Participatory Networking

- End-user API for SDNs
- Exposes existing mechanisms
Participatory Networking

- End-user API for SDNs
- Exposes existing mechanisms
- No effect on unmodified applications
The PANE prototype
1. semantics

The PANE prototype
1. semantics
2. protocol

The PANE prototype
1. semantics
2. protocol
3. controller
1. semantics
2. protocol
3. controller
Semantics
Flowgroup
<table>
<thead>
<tr>
<th>Flowgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>src=128.12/16</td>
</tr>
<tr>
<td>Flowgroup</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>src=128.12/16 ∧ dst.port ≤1024</td>
</tr>
</tbody>
</table>

...
<table>
<thead>
<tr>
<th>Flowgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{src}=128.12/16 \land \text{dst.port} \leq 1024$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowgroup</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>src=128.12/16 ∧ dst.port ≤1024</td>
</tr>
<tr>
<td>Privileges</td>
</tr>
<tr>
<td>deny, allow</td>
</tr>
<tr>
<td>Flowgroup</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>src=128.12/16 ∧ dst.port ≤1024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>deny, allow bandwidth: 5Mb/s</td>
</tr>
<tr>
<td>limit: 10Mb/s</td>
</tr>
<tr>
<td>Flowgroup</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>src=128.12/16 ∧ dst.port ≤1024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s hint query</td>
</tr>
<tr>
<td>Flowgroup</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td><code>src=128.12/16 ∧ dst.port ≤ 1024</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s hint query</td>
</tr>
</tbody>
</table>
Flowgroup

\[ \text{src} = 128.12/16 \land \text{dst.port} \leq 1024 \]

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s hint query</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
</tr>
</tbody>
</table>
bandwidth
50Mbps
<table>
<thead>
<tr>
<th>root</th>
<th>bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100Mbps</td>
</tr>
<tr>
<td>root</td>
<td>bandwidth</td>
</tr>
<tr>
<td></td>
<td>50Mbps</td>
</tr>
</tbody>
</table>
Root
share

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>root</td>
<td>bandwidth 100Mbps</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>root</td>
<td>bandwidth 50Mbps</td>
</tr>
</tbody>
</table>

x

y

w

z
**Flowgroup**

\[ \text{src=128.12/16} \land \text{dst.port} \leq 1024 \]

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s</td>
</tr>
<tr>
<td>Bob</td>
<td>limit: 10Mb/s</td>
</tr>
<tr>
<td></td>
<td>hint</td>
</tr>
<tr>
<td></td>
<td>query</td>
</tr>
</tbody>
</table>
### Flowgroup

<table>
<thead>
<tr>
<th>src = 128.12/16 ∧ dst. port ≤ 1024</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>Speakers</strong></th>
<th><strong>Privileges</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s <strong>hint</strong> query</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
</tr>
<tr>
<td>Flowgroup</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>src=128.12/16 ∧ dst.port ≤ 1024</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s hint query</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
</tr>
</tbody>
</table>

Reserve 2 Mbps from now to +5min?
## Flowgroup

\[
\text{src}=128.12/16 \land \text{dst.port} \leq 1024
\]

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow</td>
</tr>
<tr>
<td>Bob</td>
<td>bandwidth: 5Mb/s</td>
</tr>
<tr>
<td></td>
<td>limit: 10Mb/s</td>
</tr>
<tr>
<td></td>
<td>\text{hint}</td>
</tr>
<tr>
<td></td>
<td>\text{query}</td>
</tr>
</tbody>
</table>

Yes
### Flowgroup

<table>
<thead>
<tr>
<th>src=128.12/16 ∧ dst.port ≤1024</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speakers</strong></td>
</tr>
<tr>
<td>Alice</td>
</tr>
<tr>
<td>Bob</td>
</tr>
<tr>
<td><strong>Privileges</strong></td>
</tr>
<tr>
<td>deny, allow bandwidth: 5Mb/s</td>
</tr>
<tr>
<td>limit: 10Mb/s</td>
</tr>
<tr>
<td>hint query</td>
</tr>
</tbody>
</table>

This traffic will be short and bursty.
## Flowgroup

\[ \text{src}=128.12/16 \land \text{dst.port} \leq 1024 \]

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s hint query</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
</tr>
</tbody>
</table>
## Flowgroup

\[ \text{src}=128.12/16 \land \text{dst.port} \leq 1024 \]

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s limit: 10Mb/s</td>
</tr>
<tr>
<td>Bob</td>
<td>hint query</td>
</tr>
</tbody>
</table>

How much web traffic in the last hour?
<table>
<thead>
<tr>
<th>Speakers</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>deny, allow bandwidth: 5Mb/s</td>
</tr>
<tr>
<td></td>
<td>limit: 10Mb/s</td>
</tr>
<tr>
<td></td>
<td>hint</td>
</tr>
<tr>
<td>Bob</td>
<td>query</td>
</tr>
</tbody>
</table>

Flowgroup
src=128.12/16 ∧ dst.port ≤1024

67,560 bytes
Current: 0 Mbps

Root share

Current: 0 Mbps

Current: 0 Mbps
Current: 0 Mbps

Root share

Current: 0 Mbps

x

bandwidth 100Mbps

Current: 0 Mbps

y

bandwidth 100Mbps

PANE
Current: 0 Mbps

<table>
<thead>
<tr>
<th>bandwidth 100Mbps</th>
</tr>
</thead>
</table>

Current: 0 Mbps

**Root share**

Current: 0 Mbps

Current: 0 Mbps

**Reserve 80 Mbps?**
Current: 80 Mbps

Root
share

x

Current: 80 Mbps

y

Current: 0 Mbps

Reserve 50 Mbps?
Current: 80 Mbps

Root share

Current: 80 Mbps

Current: 0 Mbps

No
Current: 80 Mbps

Root share

x

bandwidth 100Mbps

y

bandwidth 100Mbps

Current: 80 Mbps

Current: 0 Mbps

OpenFlow

Pane
Protocol
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.

Grant aBW to Alice.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.
Grant aBW to Alice.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.

Grant aBW to Alice.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.

Grant aBW to Alice.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.

Grant aBW to Alice.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
NewShare aBW for (user=Alice) [reserve <= 10Mb] on rootShare.

Grant aBW to Alice.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
Alice

reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from +20min to +30min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from +20min to +30min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from +20min to +30min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from +20min to +30min.
reserve(user=Alice, dstPort=80) = 5Mb on aBW from now to +10min.

reserve(user=Alice, dstPort=80) = 5Mb on aBW from +20min to +30min.
Root

Alice

10.0.0.2

PANE
NewShare aAC for (dstHost=10.0.0.2) [deny = True] on rootShare.
NewShare aAC for (dstHost=10.0.0.2) [deny = True] on rootShare.
NewShare aAC for (dstHost=10.0.0.2) [deny = True] on rootShare.

Grant aAC to Alice.
NewShare aAC for 
(dstHost=10.0.0.2) [deny = True]  
on rootShare.

Grant aAC to Alice.

10.0.0.2
Alice

10.0.0.2

PANE
deny(dstHost=10.0.0.2, srcHost=10.0.0.3) on aAC from now to +5min.
deny(dstHost=10.0.0.2, srcHost=10.0.0.3) on aAC from now to +5min.
deny(dstHost=10.0.0.2, srcHost=10.0.0.3) on aAC from now to +5min.
Current Status
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
---!
OCCUPY EVERYTHING

#OCCUPYWALLST

WE ALREADY KNOW THAT WE OWN EVERYTHING—THE TASK IS TO EXCLUDE THE INTRUSIONS OF CAPITAL AND POWER.
Participatory Networking
Participatory Networking

1. management API
Participatory Networking

1. management API
2. network controller
Participatory Networking

1. management API
2. network controller

Safe
Participatory Networking

1. management API
2. network controller

Safe Secure
Participatory Networking

1. management API
2. network controller

Safe Secure Fair
Questions?

Andrew Ferguson
adf@cs.brown.edu
Co-authors

• Arjun Guha
• Jordan Place
• Rodrigo Fonseca
• Shriram Krishnamurthi

Questions?

Andrew Ferguson
adf@cs.brown.edu
Backup Slides
PANE Implementation

- Policy Tree
- Share Tree
- Network Flow Table
- Forwarding & Queue Configuration
- Network Information Base (NIB)
- OpenFlow Controller
- Switches
- OpenFlow messages
- Valid Configuration
- Linearization

PANE user requests
Packet headers:
[srcIP = 10.0.0.1
dstIP = 10.0.0.2
dstPort = 80]