



# Lustre & Kerberos: in theory and in practice

**Sebastien Buisson**  
Parallel File Systems  
BDS R&D Data Operations  
sebastien.buisson@atos.net

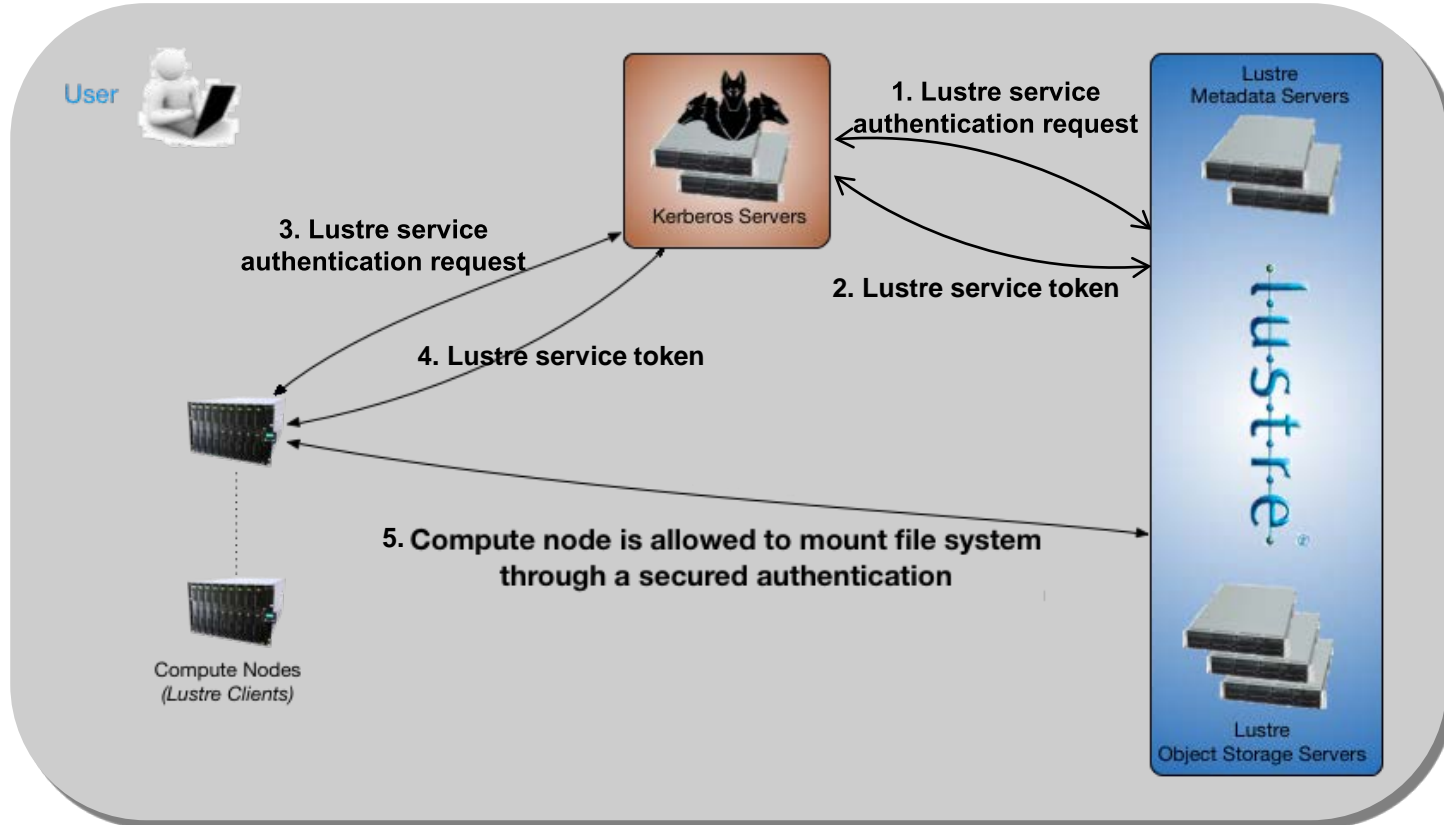
- ▶ Purpose of Kerberos
- ▶ Kerberos support in Lustre: from past to present
- ▶ Kerberos configuration in a nutshell
- ▶ Let's play with Kerberos on Lustre

# Purpose of Kerberos

- ▶ Objectives
  - control who can be part of a Lustre file system
  
- ▶ Currently
  - whichever node that
    - is connected to the Interconnect network
    - knows the MGS and file system names
  - can mount Lustre as a client!
  - can format a target and mount Lustre as a server!
  
- ▶ Kerberos is a possible solution
  - authentication of nodes and users

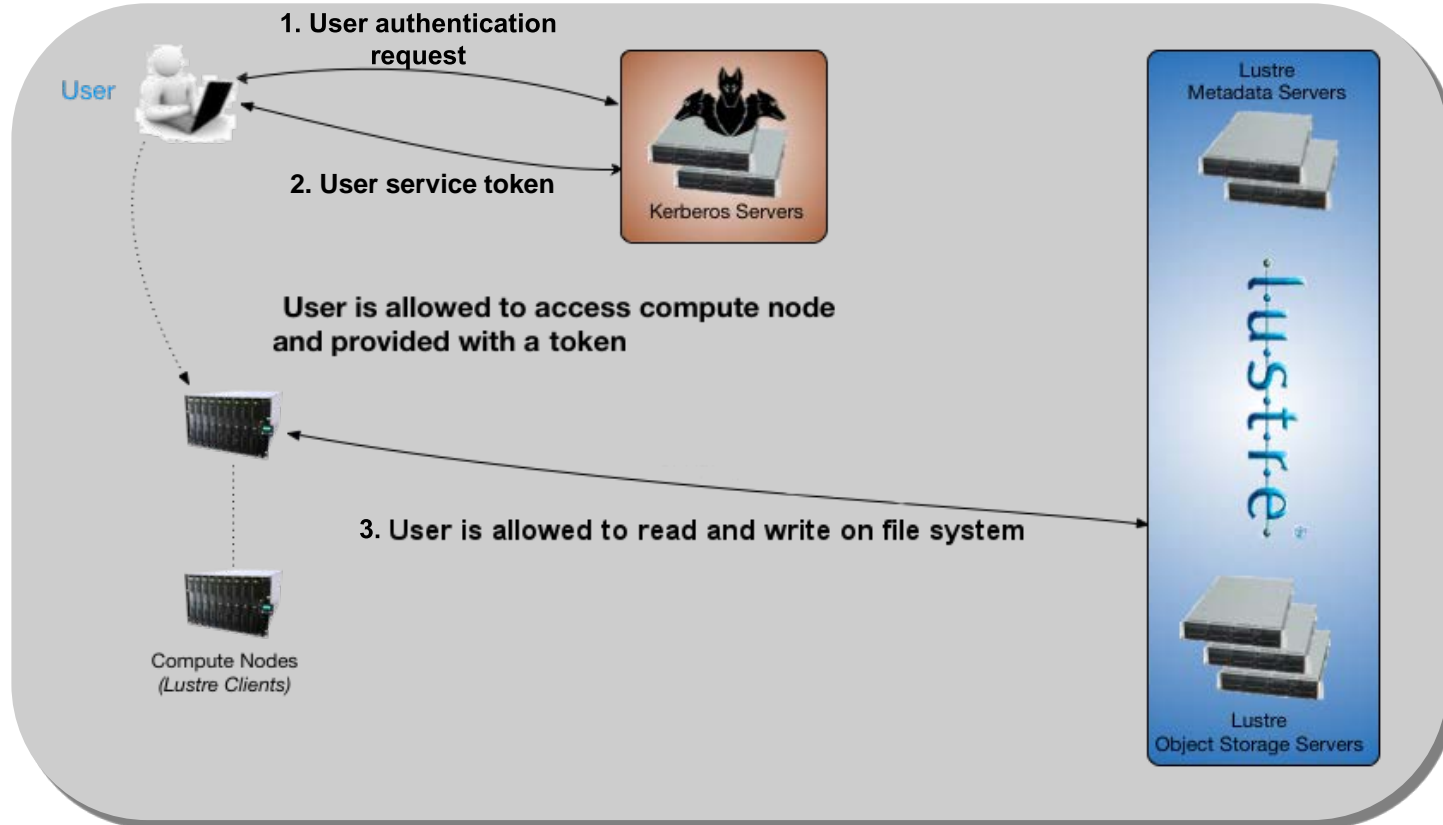
# Purpose of Kerberos

## ► How it works with Lustre: mount



# Purpose of Kerberos

## ► How it works with Lustre: file access



# Purpose of Kerberos

- ▶ Objectives
  - protect data transfers between nodes
- ▶ Currently
  - Lustre checksums guard against network data corruption
- ▶ Kerberos is a possible solution
  - integrity and privacy of bulk data and rpc messages

flavor	auth	RPC message protection	Bulk data protection
krb5n	yes	no	checksum
krb5a	yes	headers integrity	checksum
krb5i	yes	integrity	integrity
krb5p	yes	privacy	privacy

# Kerberos support: from past to present



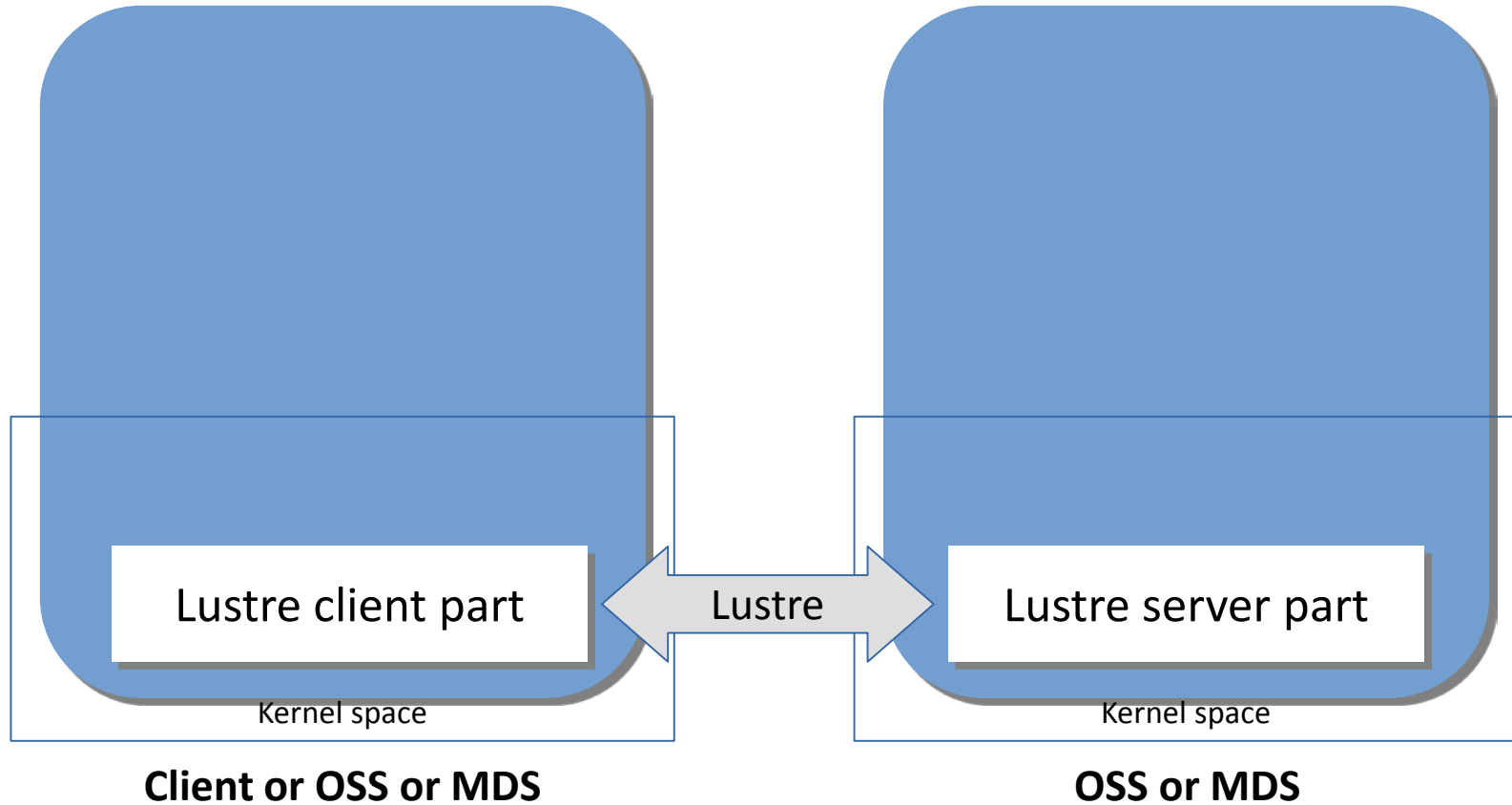
- ▶ Back in **2010**
  - Lustre 2.0 was successfully kerberized on production cluster at Pittsburgh Supercomputing Center
    - « Kerberized Lustre 2.0 over the WAN », Josephine Palencia, PSC, LUG 2010
- ▶ But in **2013**
  - Lustre 2.4 was unable to even start with Kerberos activated
    - « Strong authentication in Lustre & friends », Daniel Kobras, S+C, LAD 2013
- ▶ Bull/Atos R&D experiments with Lustre 2.5
  - '--enable-gss' build broken
  - instant crash when starting Lustre with Kerberos activated
  - ⇒ still a lot of work to do!

- ▶ In current master: GSSAPI/Kerberos related patches
  - build/new kernel support
    - LU-4085, LU-4012, LU-4372: **landed**
  - LWP/OSP support at GSSAPI level
    - LU-3778: **in progress**
  - Bug fixes for GSS/Kerberos
    - LU-4113: **landed**
    - LU-6020 (multiple patches): **landed**
    - LU-6356 (multiple patches): **landed and in progress**

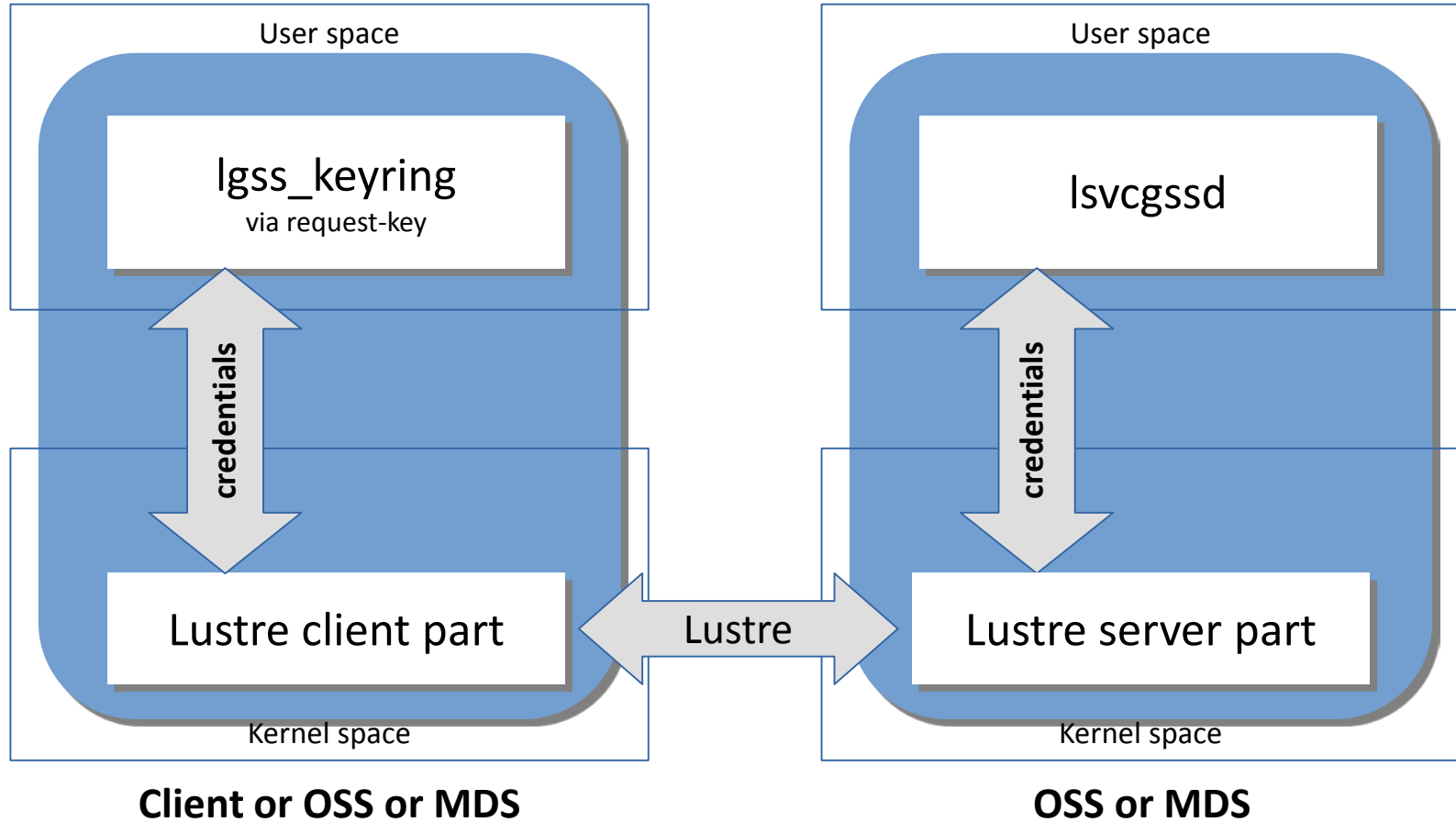
# Kerberos configuration in a nutshell

- ▶ Every file system access needs to be authenticated with Kerberos credentials, named **principals**:
  - MGS
    - **lustre\_mgs/<mgt network address>.DOMAIN**
  - MDS
    - **lustre\_mds/<mds hostname>.DOMAIN**
    - *for each mdt network address:*  
**lustre\_mds/<mdt network address>.DOMAIN**
  - OSS
    - *for each ost network address:*  
**lustre\_oss/<ost network address>.DOMAIN**
  - Client
    - **lustre\_root/<network address>.DOMAIN**
  - normal users need their own principal

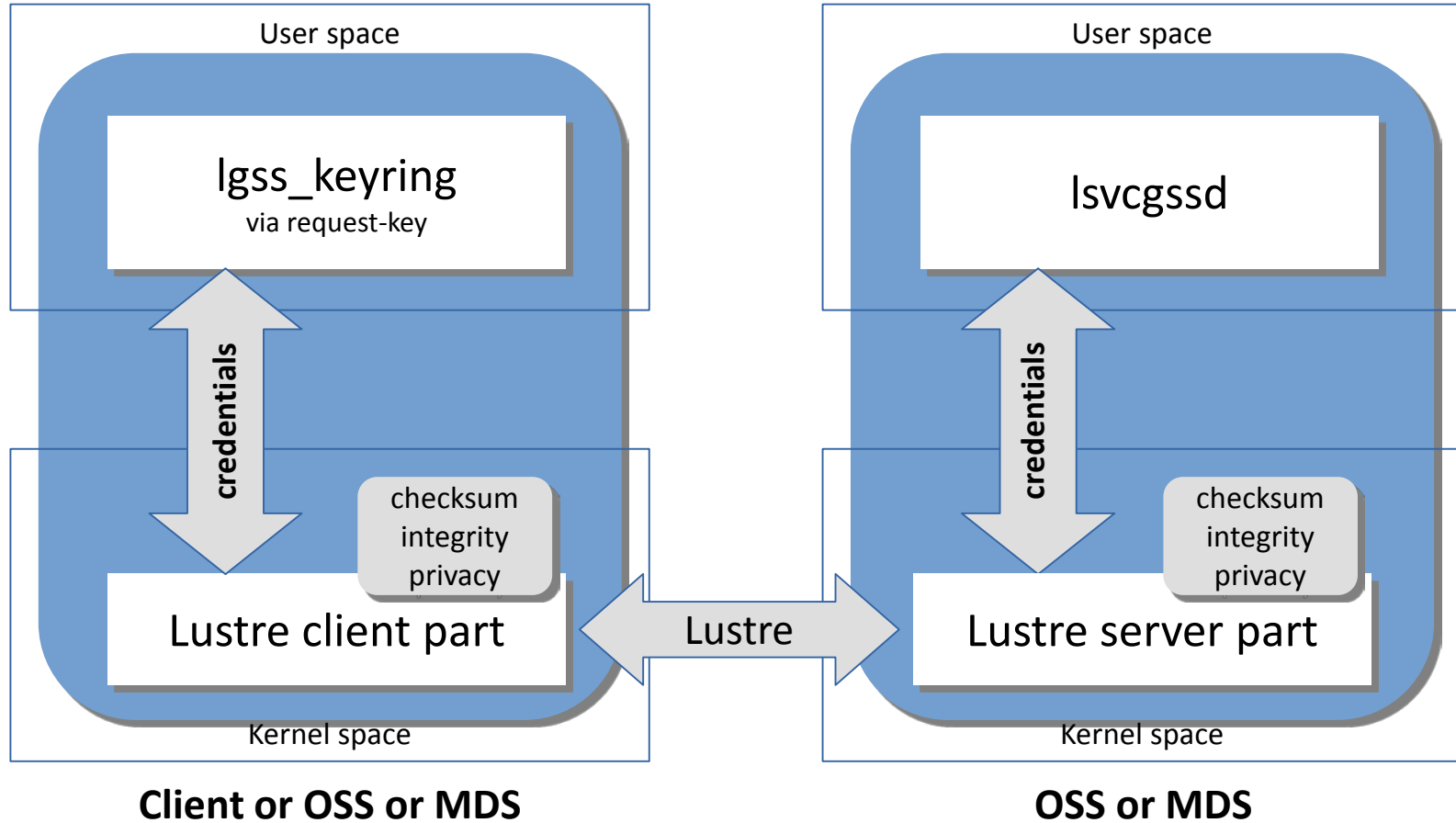
# Configuration in a nutshell



# Configuration in a nutshell



# Configuration in a nutshell



## ► Supported Kerberos flavors

flavor	auth	RPC message protection	Bulk data protection
krb5n	yes	no	checksum
krb5a	yes	headers integrity	checksum
krb5i	yes	integrity	integrity
krb5p	yes	privacy	privacy

## ► Flavors can be refined at various levels:

- `lctl conf_param <fs>.srpc.flavor.default = krb5i`
- `lctl conf_param <fs>.srpc.flavor.o2ib0 = null`
- `lctl conf_param <fs>.srpc.flavor.default.client2ost = krb5p`

## ► MGS particular case:

- `'mgssec=flavor'` mount option for targets and clients



# Let's play with Kerberos for Lustre

## ▶ R&D testbed

### – software:

- RHEL 6
- Kerberos MIT v5
- Lustre 2.7.0 + patches

### – hardware: 1 node per Lustre role, to 'ease' Kerberos setup

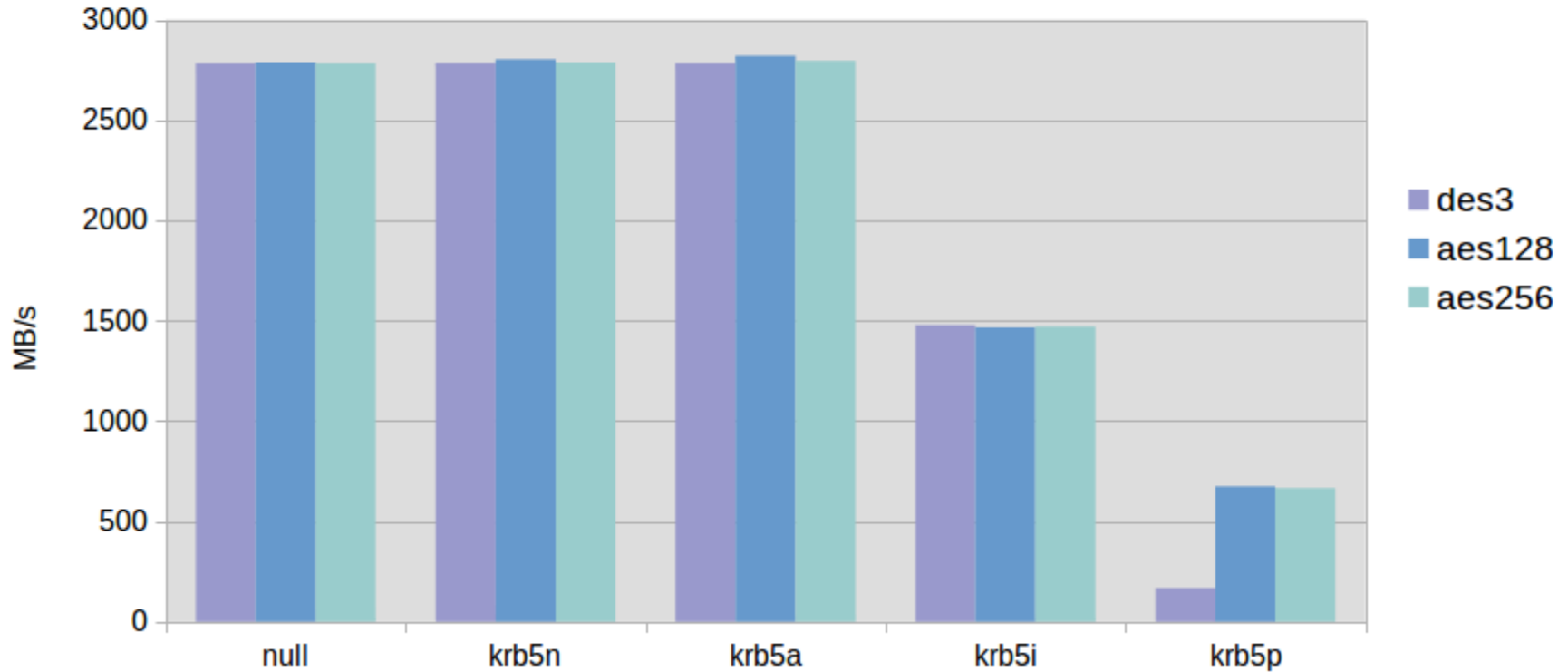
- 1 MGS
  - 1 MDS
  - 1 OSS
- } ramdisk storage
- 1 client
    - 12 cores
    - 24 GB RAM
  - Interconnect: Infiniband QDR

- ▶ With patches *in progress* in LU-3778, LU-6020, LU-6356
  - all flavors are functional
    - krb5n, krb5a, krb5i, krb5p
  - on every communication channel
    - cli2mdt, cli2ost, mdt2mdt, mdt2ost
  - for all parties:
    - MGS, MDS, OSS, Client
  
- ▶ Let's have a look at impact of Kerberos over performance

# Impact over data performance

## IOR - write

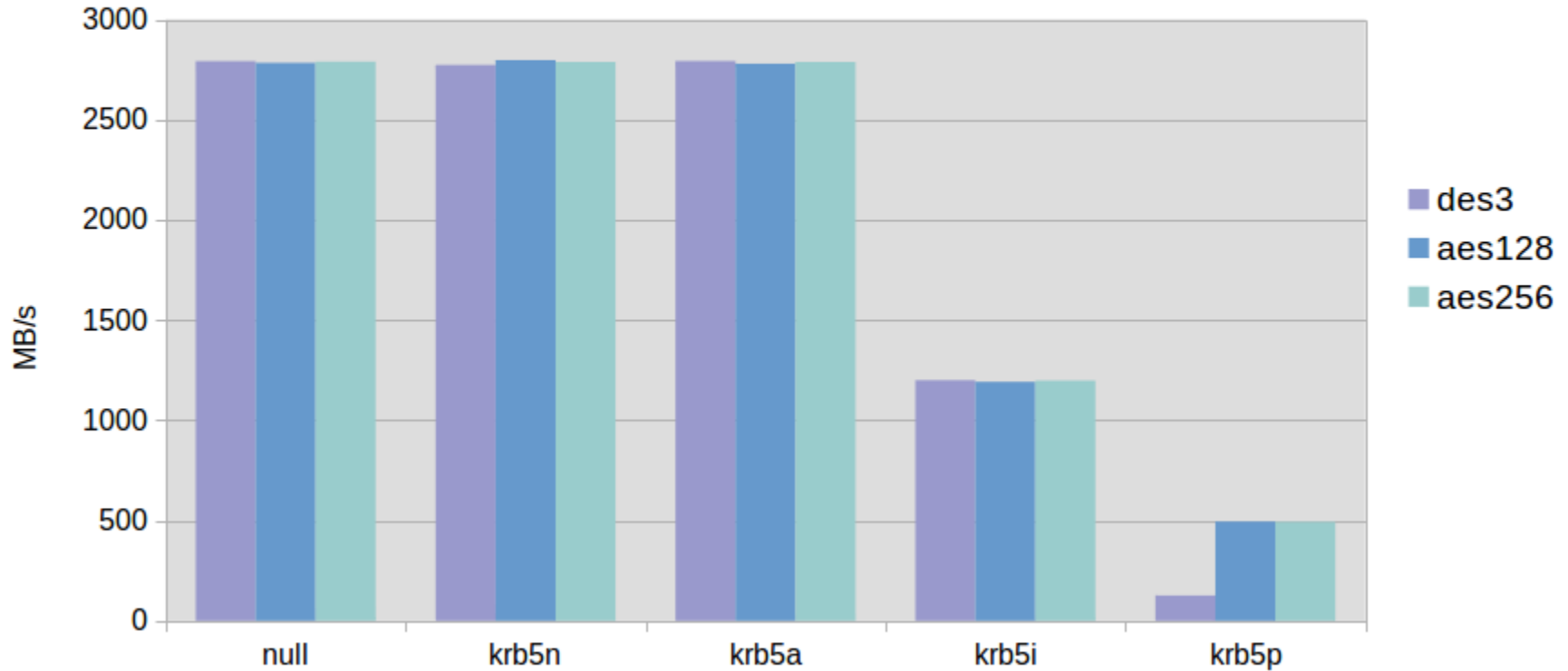
12 tasks



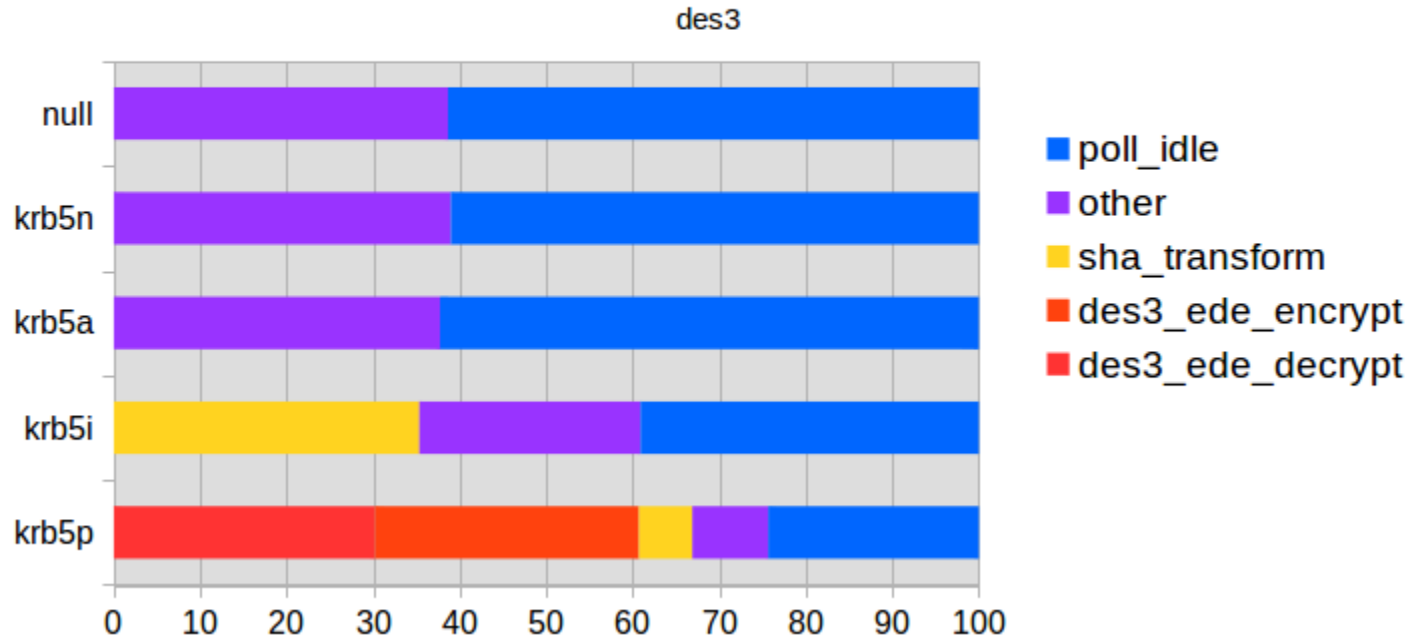
# Impact over data performance

## IOR - read

12 tasks



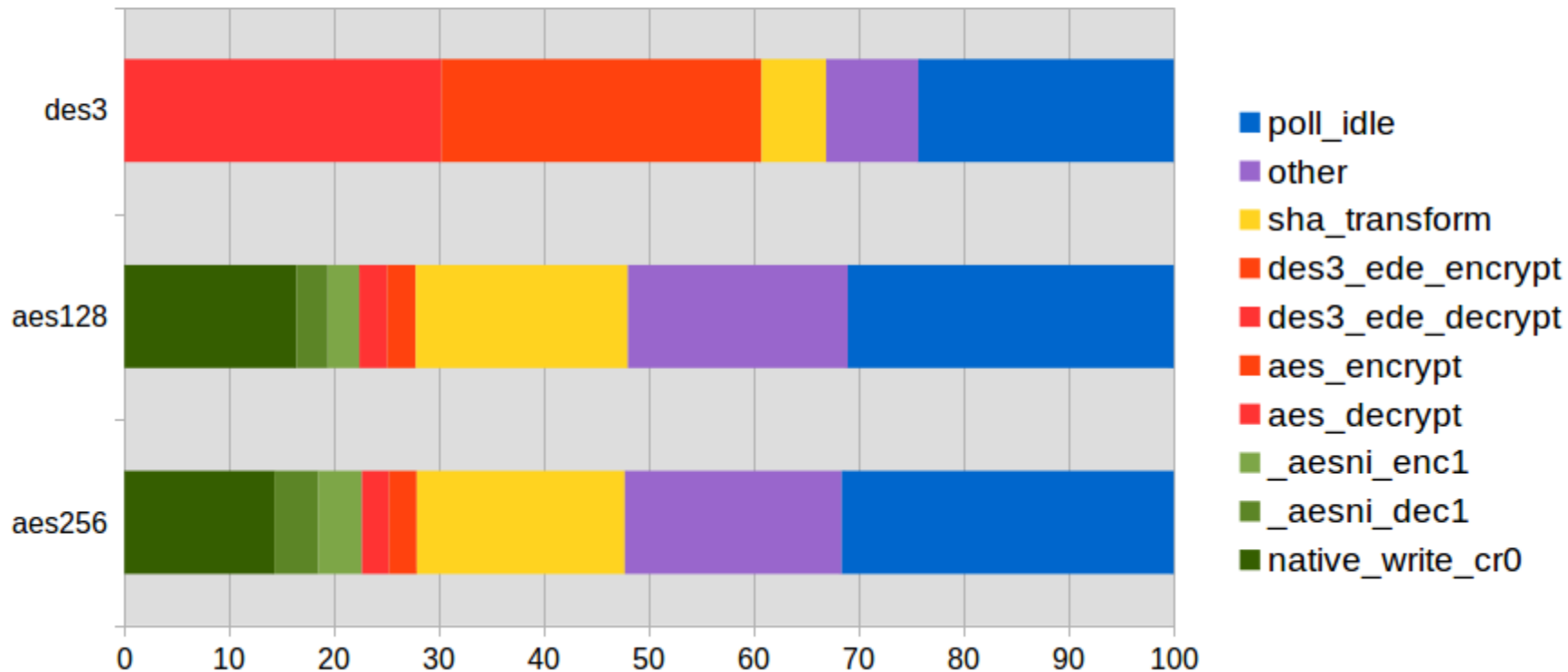
## IOR - client - CPU usage



# Impact over data performance

## IOR - Client - CPU usage

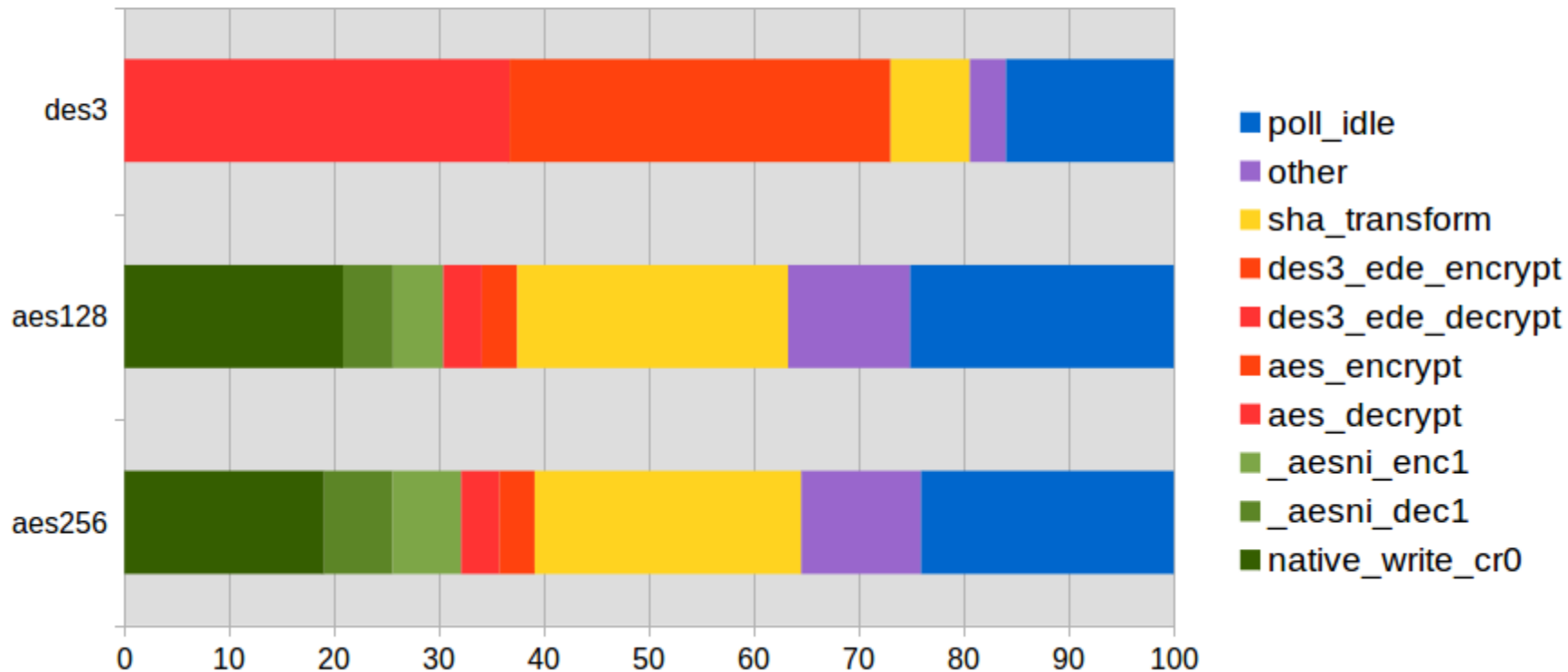
krb5p



# Impact over data performance

## IOR - OSS - CPU usage

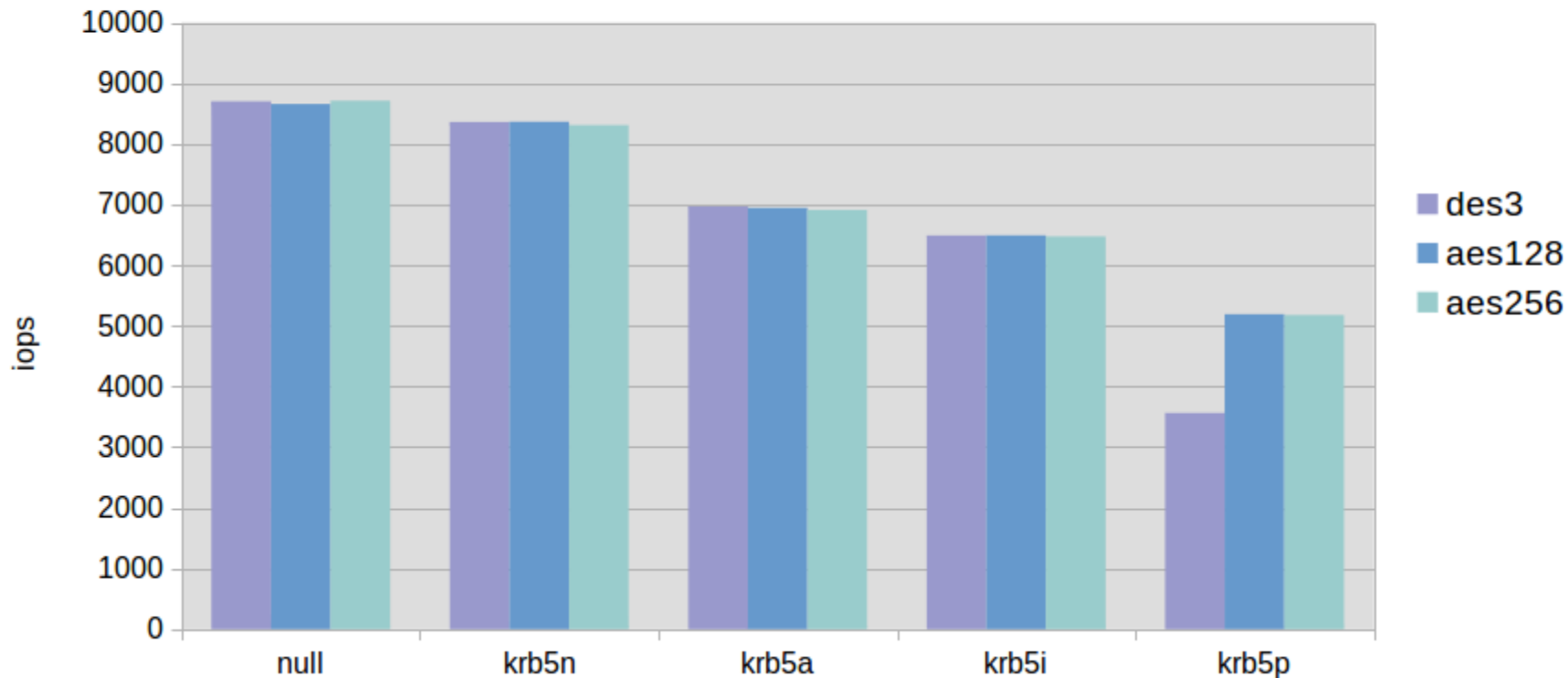
krb5p





## mdtest - file create

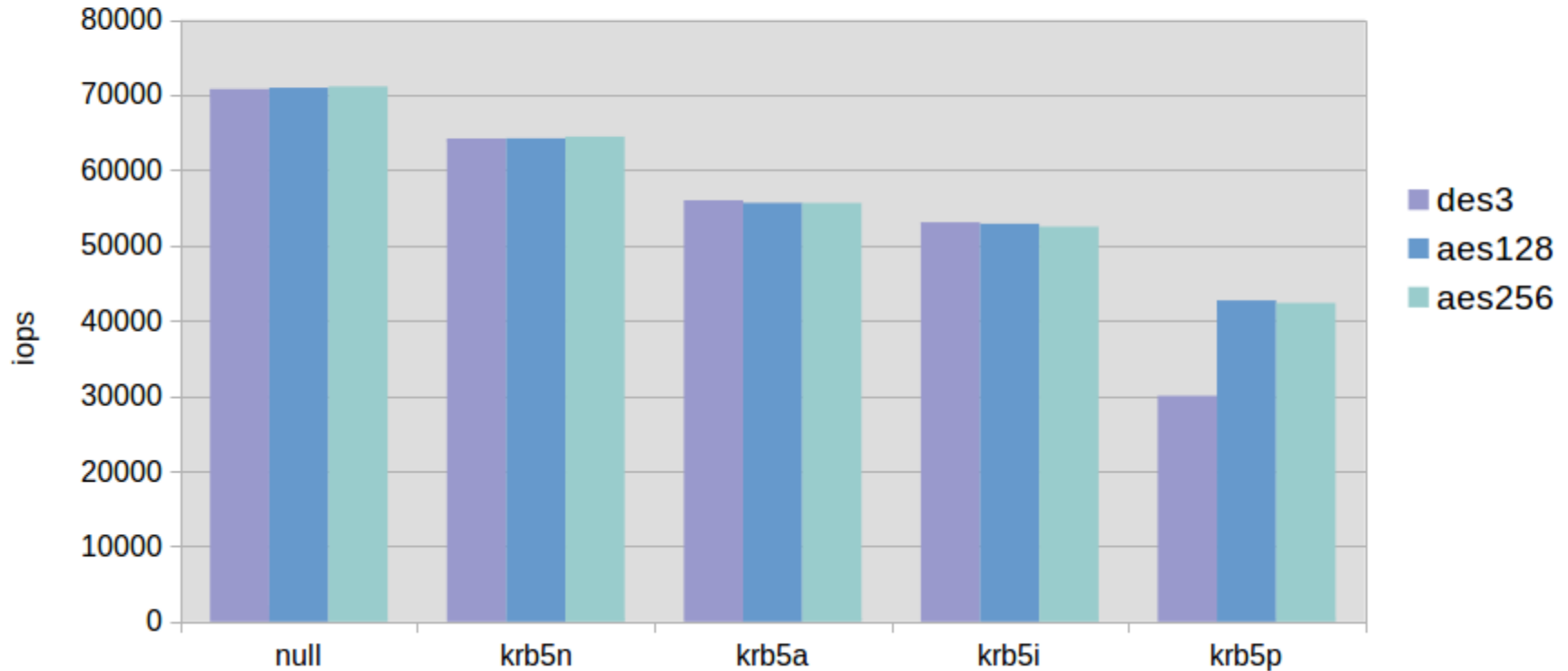
dpp - 12 tasks



# Impact over metadata performance

## mdtest - file stat

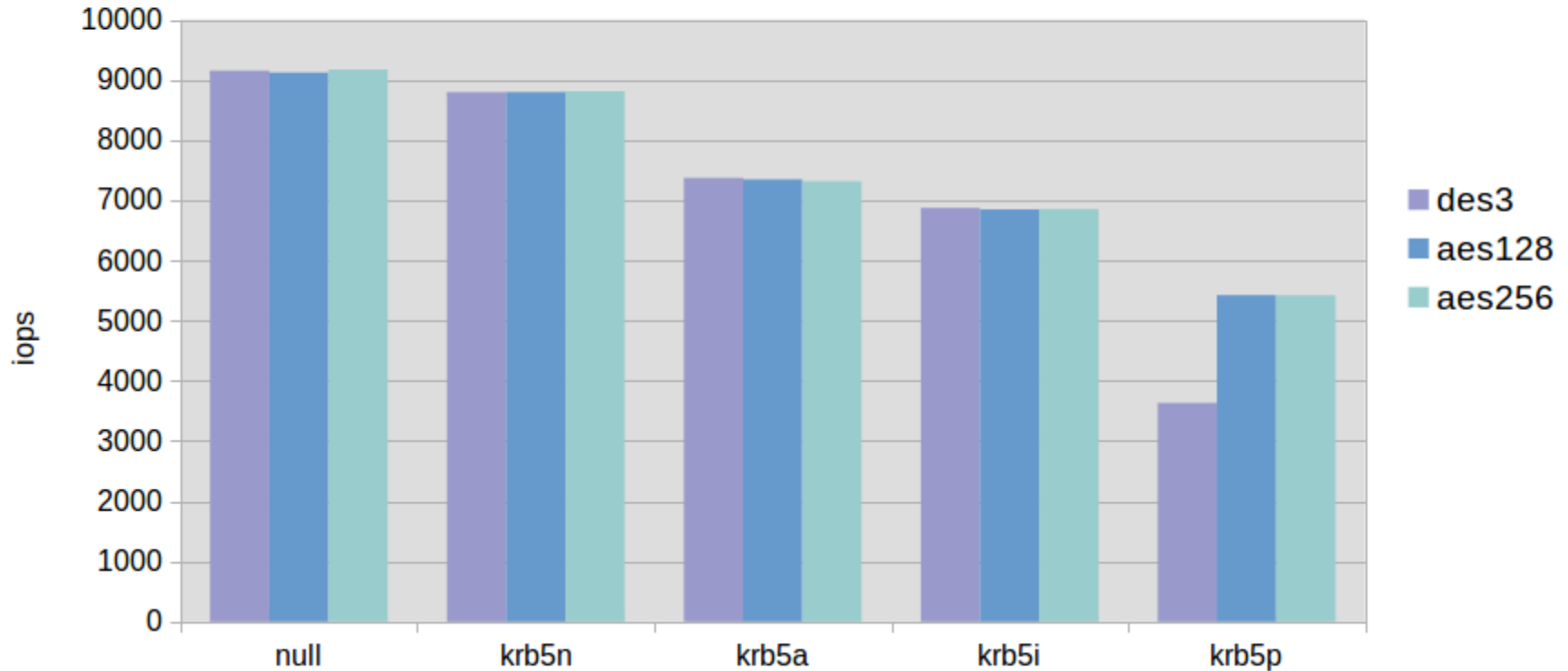
dpp - 12 tasks



# Impact over metadata performance

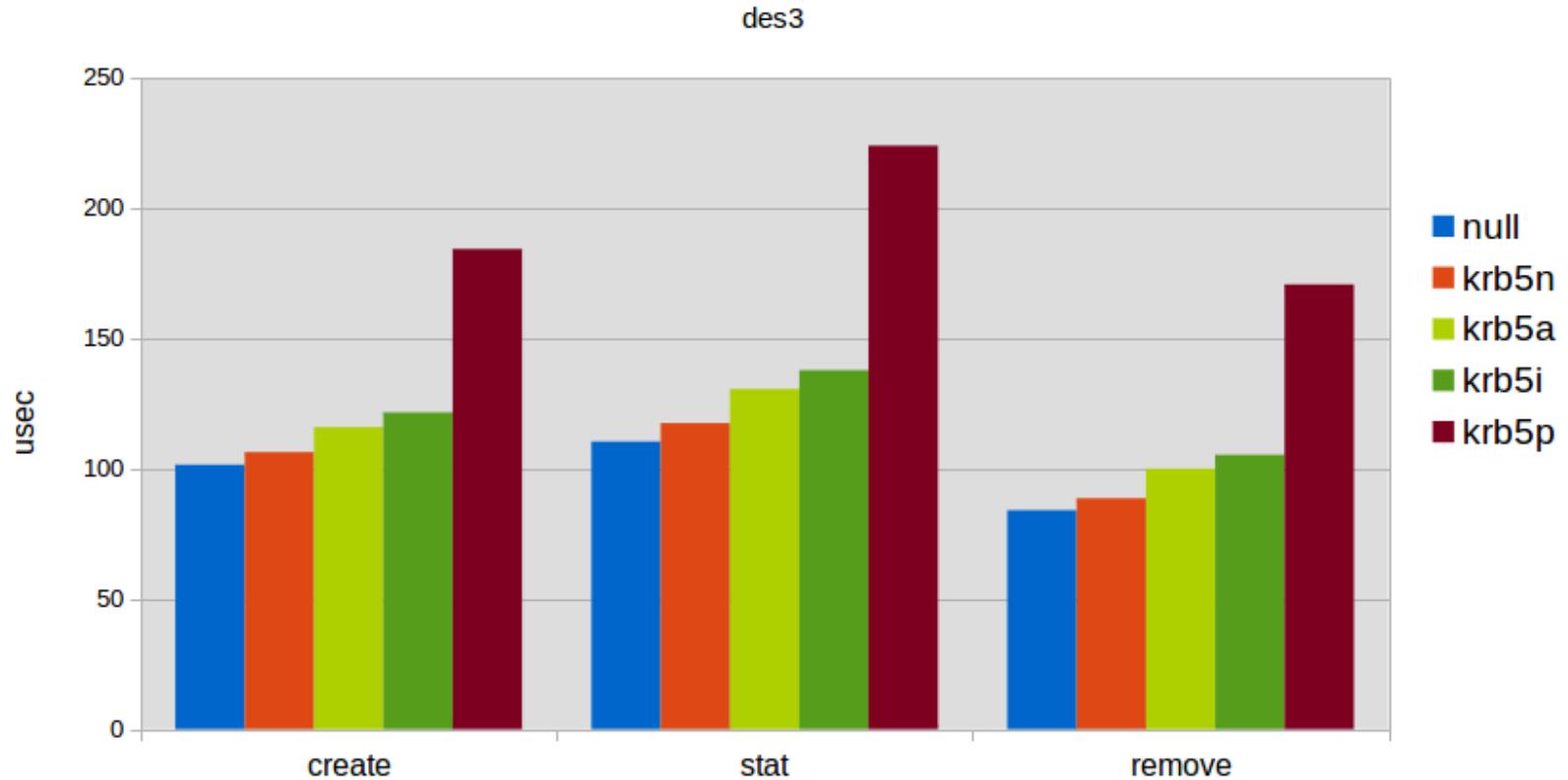
## mdtest - file remove

dpp - 12 tasks



# Impact over metadata performance

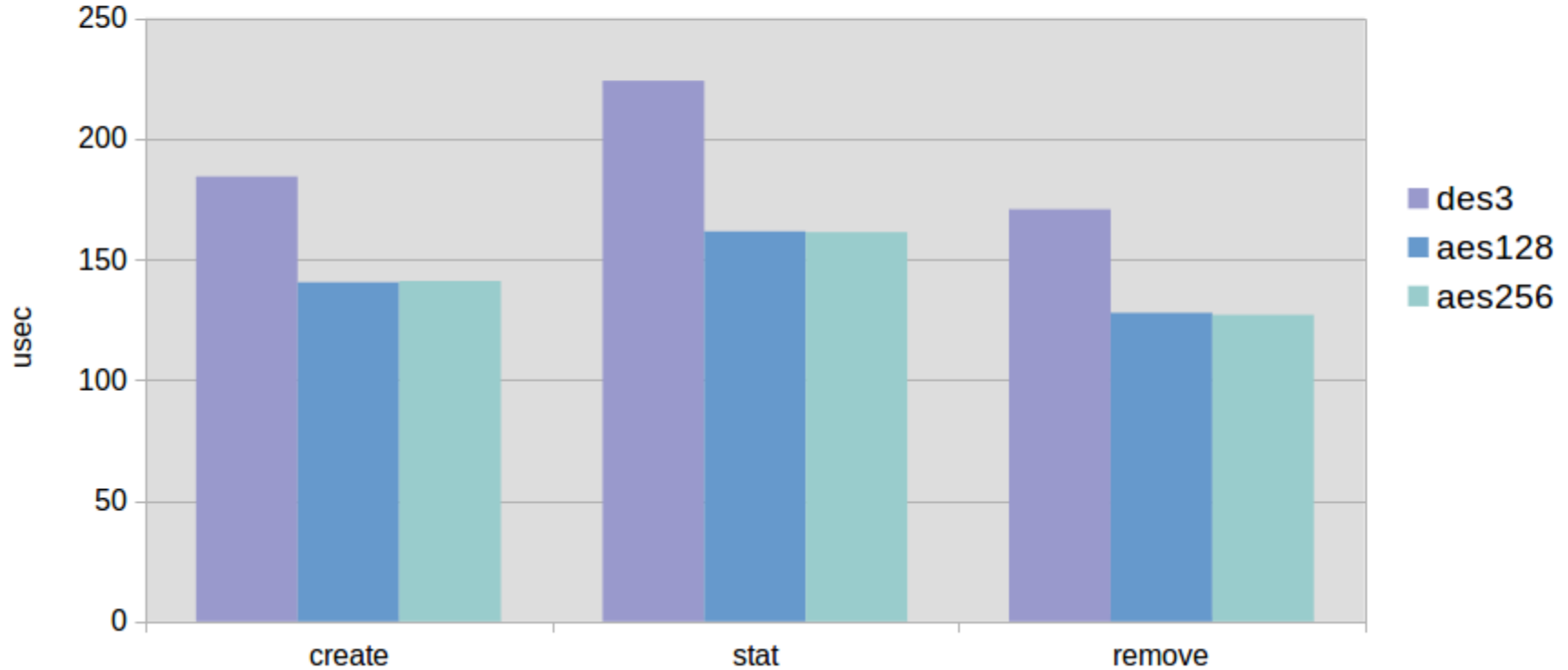
mdtest - client - req\_waittime



# Impact over metadata performance

mdtest - Client - req\_waittime

krb5p



- ▶ Kerberos support in Lustre is back!
- ▶ Performance impact
  - with authentication: very modest
  - with integrity/privacy: no pain, no gain...
- ▶ Remaining work:
  - land patches
  - document:
    - update OpenSFS wiki

---

# Thanks

For more information please contact:  
sebastien.buisson@atos.net

Atos, the Atos logo, Atos Consulting, Atos Worldgrid, Worldline, BlueKiwi, Canopy the Open Cloud Company, Yunano, Zero Email, Zero Email Certified and The Zero Email Company are registered trademarks of Atos. © 2015 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

---

04-13-2015

# Data performance tests summary

<b>write</b>	krb5n	krb5a	krb5i	krb5p
des3	<i>similar</i>	<i>similar</i>	- 50 %	- 95 %
aes128	<i>similar</i>	<i>similar</i>	- 50 %	- 75 %
aes256	<i>similar</i>	<i>similar</i>	- 50 %	- 75 %
<b>read</b>	krb5n	krb5a	krb5i	krb5p
des3	<i>similar</i>	<i>similar</i>	- 60 %	- 95 %
aes128	<i>similar</i>	<i>similar</i>	- 60 %	- 80 %
aes256	<i>similar</i>	<i>similar</i>	- 60 %	- 80 %



<b>create</b>	krb5n	krb5a	krb5i	krb5p
des3	- 5 %	- 20 %	- 25 %	- 60 %
aes128	- 5 %	- 20 %	- 25 %	- 40 %
aes256	- 5 %	- 20 %	- 25 %	- 40 %
<b>stat</b>	krb5n	krb5a	krb5i	krb5p
des3	- 10 %	- 20 %	- 25 %	- 60 %
aes128	- 10 %	- 20 %	- 25 %	- 40 %
aes256	- 10 %	- 20 %	- 25 %	- 40 %
<b>remove</b>	krb5n	krb5a	krb5i	krb5p
des3	- 5 %	- 20 %	- 25 %	- 60 %
aes128	- 5 %	- 20 %	- 25 %	- 40 %
aes256	- 5 %	- 20 %	- 25 %	- 40 %