

# How to fight with “bloated repository”

Hideki Yamane

(Debian Project:Debian Developer)

<henrich @ debian.org/or.jp>

<http://wiki.debian.org/HidekiYamane>



# Today's Agenda

- How large Debian Repository is
- One day, I found a solution...
- Is it really effective?
- Problem on slower Arch
- How shrink



## Debian supports...

- **Many many packages**
- **Many CPU architectures**
- **Some kernels**



# How large Debian Repository is?

- Arch: source, all, amd64, armel, armhf, hurd-i386, i386, ia64, kfreebsd-amd64, kfreebsd-i386, mips, mipsel, powerpc, s390, s390x, sparc



# How large Debian Repository is?

- Arch: source 52GB, all 57GB, amd64 53GB, armel 38GB, armhf 26GB, hurd-i386 14GB, i386 50GB, ia64 42GB, kfreebsd-amd64 37GB, kfreebsd-i386 36GB, mips 35GB, mipsel 34GB, powerpc 42GB, s390 36GB, s390x 24GB, sparc 39GB...

- **Total?**

(<http://www.debian.org/mirror/size>)



# How large Debian Repository is?

- Arch: source 52GB, all 57GB, amd64 53GB, armel 38GB, armhf 26GB, hurd-i386 14GB, i386 50GB, ia64 42GB, kfreebsd-amd64 37GB, kfreebsd-i386 36GB, mips 35GB, mipsel 34GB, powerpc 42GB, s390 36GB, s390x 24GB, sparc 39GB...
- **Total: 615GB!!**

(<http://www.debian.org/mirror/size>)

How can we  
improve this?





Can we shrink this?

**Yes,** in some ways...

**Drop** support architectures

**Delete** packages from archive





Can we shrink this?

However, we don't want such  
solutions

~~Drop support architectures~~

~~Delete packages from archive~~



Use XZ!

Default compression is  
gzip  
**xz** can **reduce** file size



# Use XZ!

ex)

**fonts-horai-umefont** (I'm maintainer :-)

**By gzip -9 : 43,664kb**

**By xz : 25,476kb**



# Use XZ!

ex)

**fonts-horai-umefont** (I'm maintainer :-)

**By gzip -9 : 43,664kb**

**By xz -9 . 25,476kb**

→ **5,916kb**



# WARNING!

The archive software now accepts packages using xz for compression in addition to gzip and bzip2 for both source and binary packages.

(snip)

Additionally please only use xz (or bzip2 for that matter) if your package really profits from its usage (for example, it provides a significant space saving). While those methods may compress better they often use more CPU time to do so and a very small decrease in package size is hardly worth the extra effort placed on slower systems. Think of both user systems and the Debian build systems which will waste more time – an especially bad problem on slower architectures.

(“The archive now supports xz compression” by Ansgar Burchardt <[ansgar@debian.org](mailto:ansgar@debian.org)>  
<http://lists.debian.org/debian-devel-announce/2011/08/msg00001.html>)

debian



# WARNING!

The archive software now accepts packages using xz for compression in addition to gzip and bzip2 for both source and binary packages.

(snip)

Additionally please only use xz (or bzip2 for that matter) if your package really profits from its usage (for example, it provides a significant space saving). While those methods may compress better they often use more CPU time to do so and a very small decrease in package size is hardly worth the extra effort placed on slower systems. Think of both user systems and the Debian build systems which will waste more time – an especially bad problem on slower architectures.

(“The archive now supports xz compression” by Ansgar Burchardt <[ansgar@debian.org](mailto:ansgar@debian.org)>  
<http://lists.debian.org/debian-devel-announce/2011/08/msg00001.html>)

debian



XZ on Slower arch is problem...

**It'll eat  
most CPU time**



XZ on Slower arch is problem...

Then...

if only on Powerful arch?





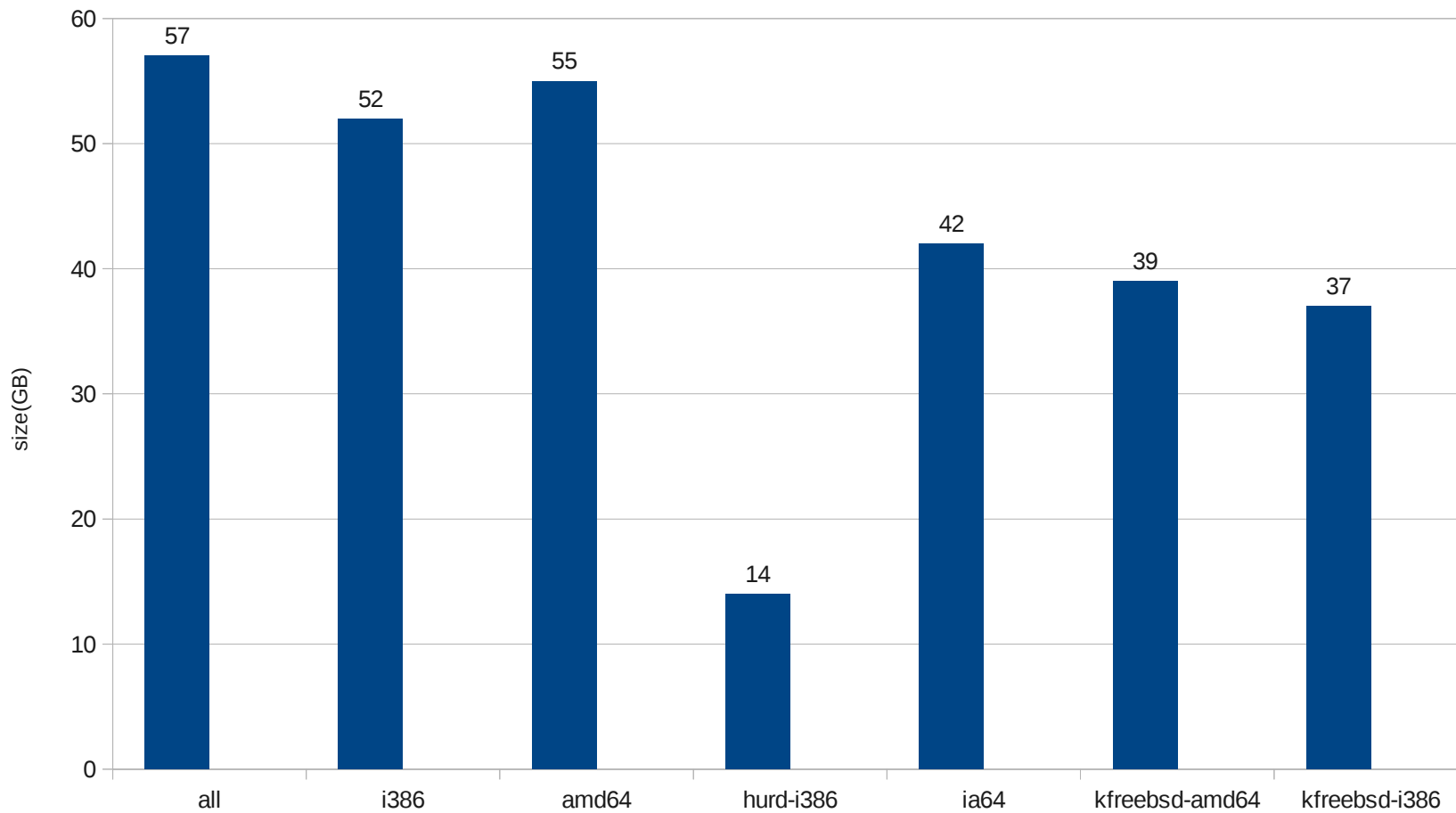
XZ on Powerful arch is NOT problem

assumption:

use XZ on Intel/AMD arch by default

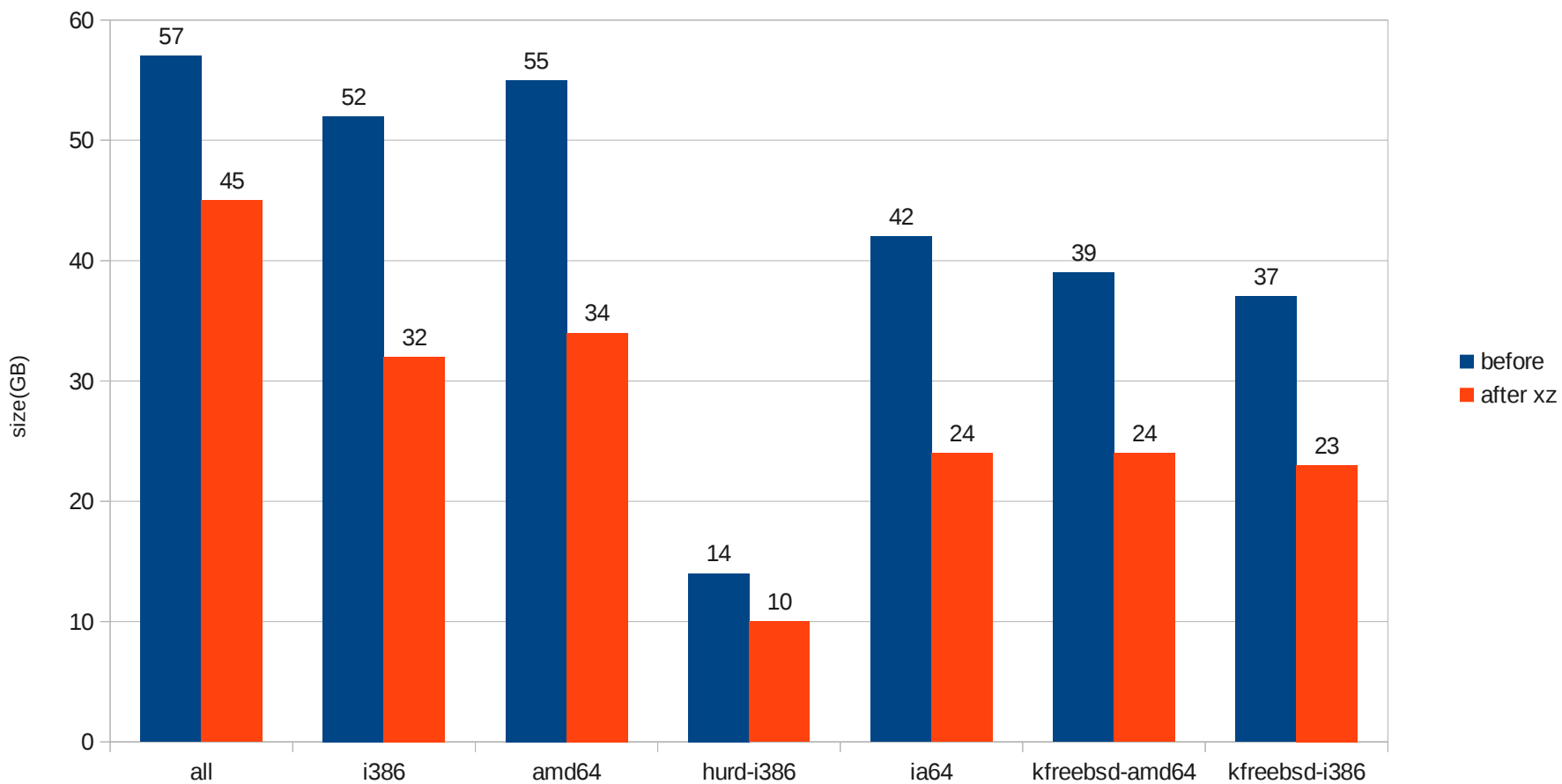


# Before XZ...



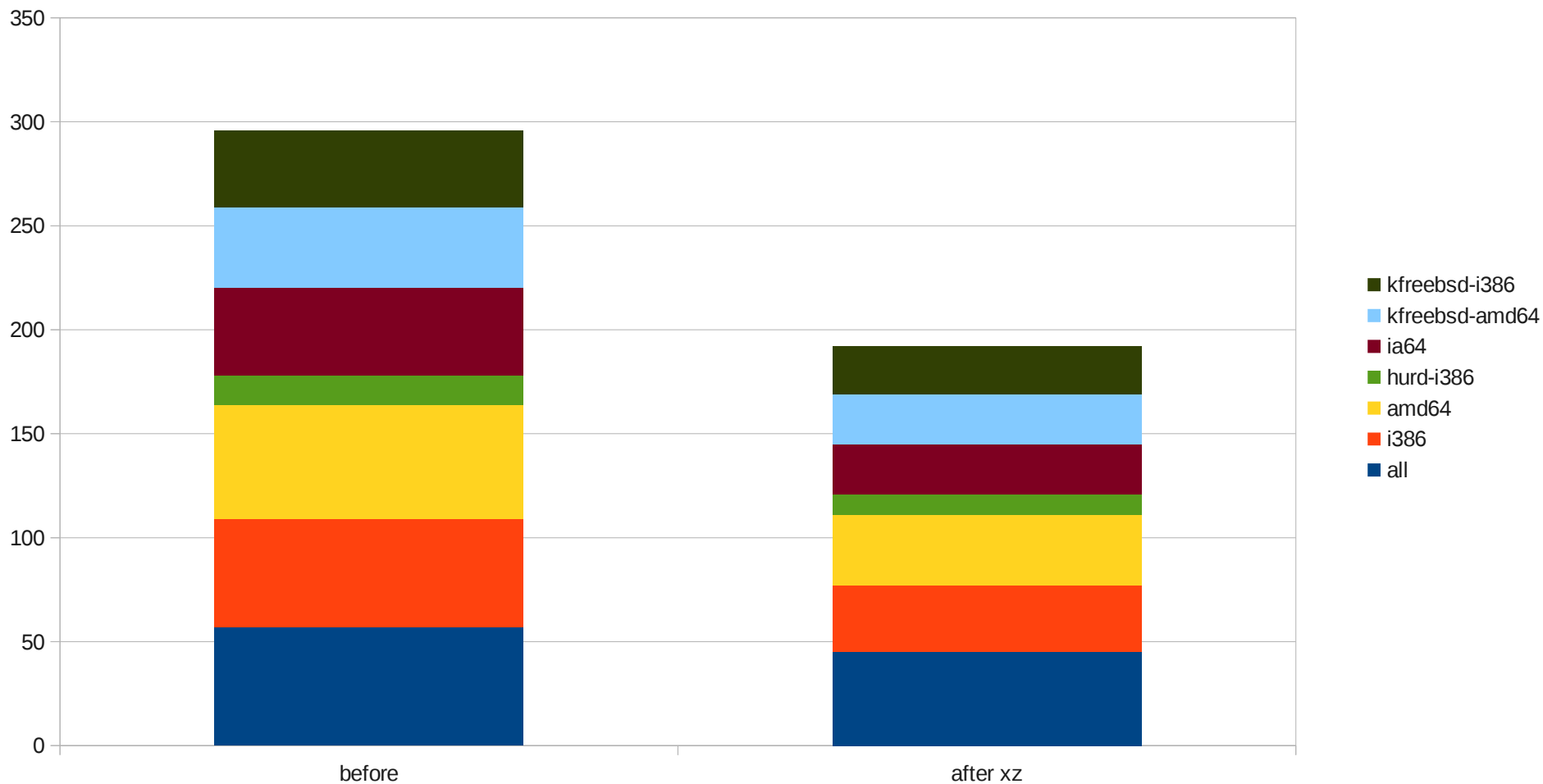


# After XZ!





# How shrink it is





# How shrink it is

architecture	before	after xz	difference	Reduction Rate
all	57	???	---	---
i386	52	???	---	---
amd64	55	???	---	---
hurd-i386	14	???	---	---
ia64	42	???	---	---
kfreebsd-amd64	39	???	---	---
kfreebsd-i386	37	???	---	---
<b>total</b>	<b>296</b>	<b>???</b>	<b>---</b>	<b>---</b>



# How shrink it is

architecture	before	after xz	difference	Reduction Rate
all	57	45	-12	21%
i386	52	32	-20	38%
amd64	55	34	-21	38%
hurd-i386	14	10	-4	29%
ia64	42	24	-18	43%
kfreebsd-amd64	39	24	-15	38%
kfreebsd-i386	37	23	-14	38%
<b>total</b>	<b>296</b>	<b>192</b>	<b>-104</b>	<b>35%</b>

# Conclusion

- How big Debian Repository is : 615GB
- One day, I found a solution... : use xz
- Is it really effective? : YES!
- Problem on slower Arch : x86 + all
- How shrink : **100GB!**

**So, recommend to apply XZ to all, \*i386 and \*amd64 by default (exclude Priority:require)**

# Also, Thanks to nice pictures

- SpaceFun

<http://wiki.debian.org/DebianArt/Themes/SpaceFun>

By Valessio Brito

licensed under GPL-2

- Debian Theme (etch?)

- Debian Theme (by @nogajun)

- Thinking

<http://www.flickr.com/photos/nachoissd/3499105933/>

By Victor Pérez :: victorperezp.com

licensed under Creative Commons Attribution 2.0 Generic (CC BY 2.0)

- A successful tool is one that was used to do something undreamed of by its author.

<http://www.flickr.com/photos/katerha/5746905652/>

By katerha

licensed under Creative Commons Attribution 2.0 Generic (CC BY 2.0)