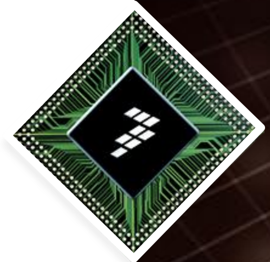
Efficient Guest Agnostic Virtualization With Embedded Power Architecture® KVM

Varun Sethi
Freescale Semiconductor



Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, t
he Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony
are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack,
ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ
Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks
of Freescale Semiconductor, Inc. All other product or service names are the property
of their respective owners. © 2011 Freescale Semiconductor, Inc.

Motivation

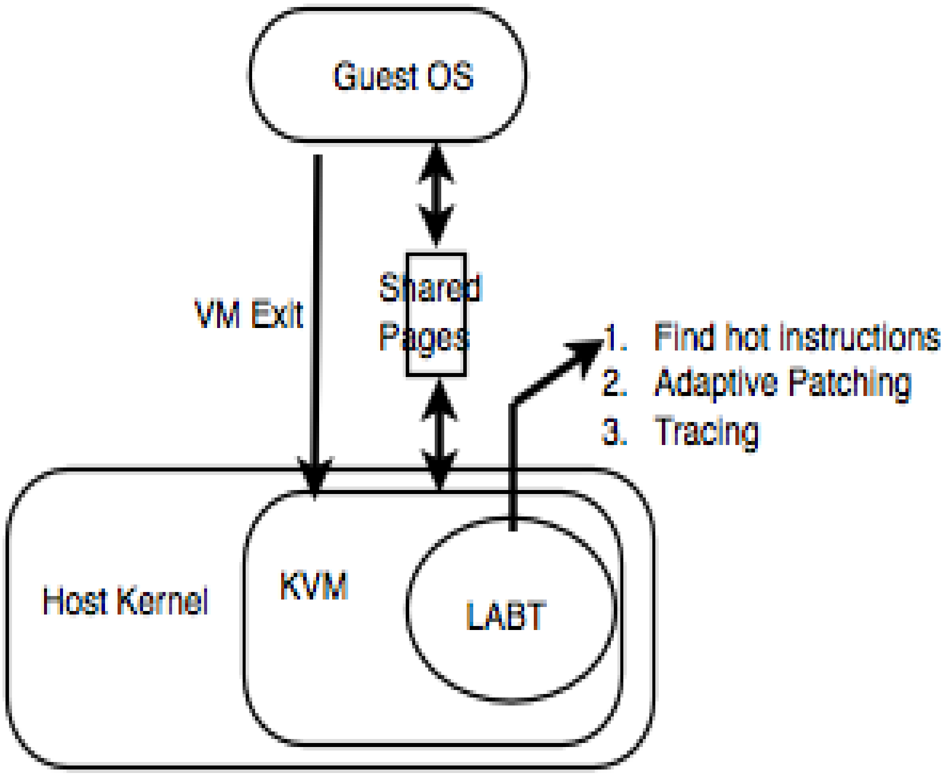
- Growing interest in virtualization using KVM on embedded Power Architecture platforms
- Requirement to run various customer specific operating systems with embedded Power Architecture KVM
- Possible to run unmodified guest on embedded Power Architecture KVM – But this comes at a significant cost associated with VM exits
  - *Problem severe for cores without virtualization assists*
- Possible to “Binary translate” guest privileged instructions from the host side.
  - Continue to run an unmodified guest

freescale™

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Adaptive Binary Translation

- Dynamically infer the instructions that cause a large number of VM exits
- Binary Translate these instructions to a faster emulation code .
- Instruction translation maintained on a shared memory region
- “Memory tracing” (remove R/W permissions to shared page) implemented to control guest access to modified pages
  - Guest access to page generates a DSI exception



Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Design Challenges

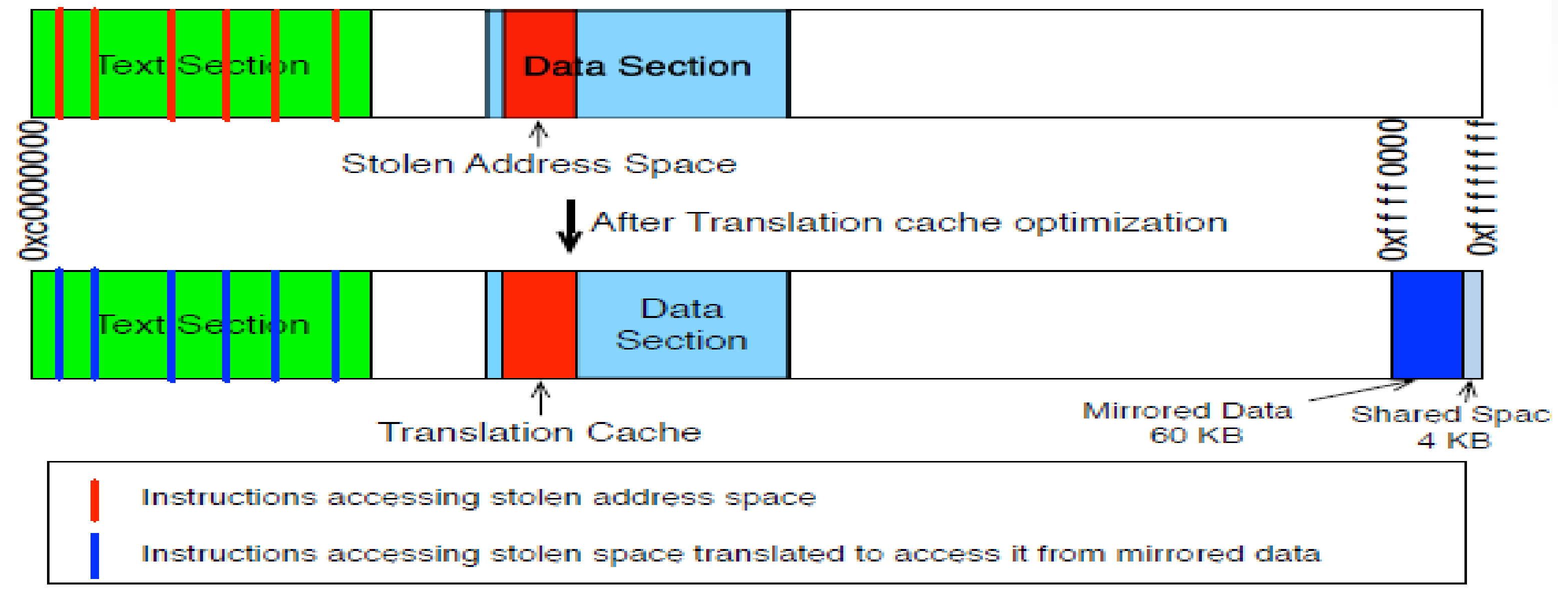
- Handling complex instruction translations
  - Multi line patching could be complex
  - Translation cache placement issue

- Minimizing performance overhead (excessive DSIs) due to Memory tracing
  - Use of huge pages (TLB1)
  - Self referential code
    - Access to sys call table and exception prolog in case of Linux
  - Self modifying code
    - Code from the modified page trying to modify (Write) code on the same page

freescale™

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Solutions to Design Challenges



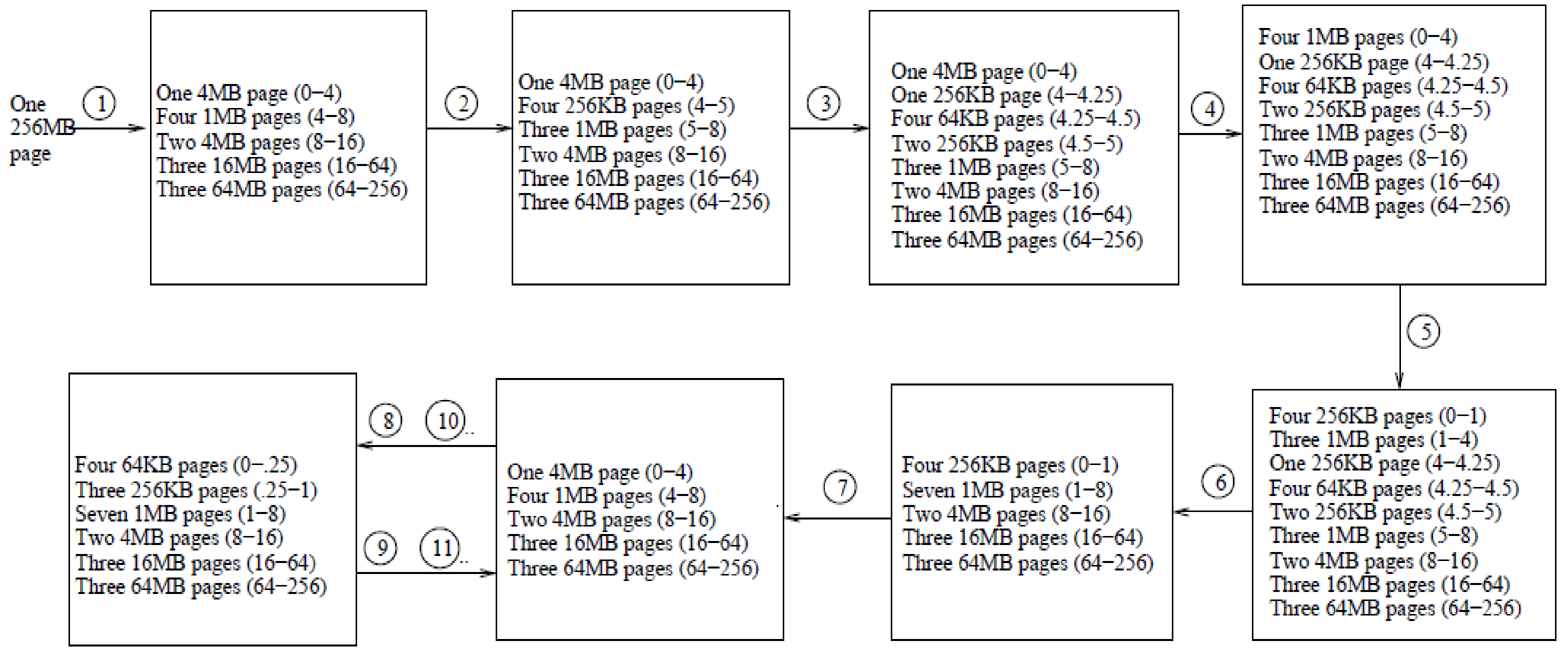
Text Section
Data Section
0xc0000000
Stolen Address Space
After Translation cache optimization
0xffff0000
0xffffffff
Text Section
Data Section
Translation Cache
Mirrored Data 60 KB
Shared Spac 4 KB
Instructions accessing stolen address space
Instructions accessing stolen space translated to access it from mirrored data

- Adaptive data mirroring algorithm to address self referential code
  - Data causing excessive DSIs mirrored to a guest memory region having R/W permission
- Addressing Translation Cache placement issue
  - Mechanism developed for stealing space from read only guest section
  - The data from the stolen section mirrored to new location

freescale™

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

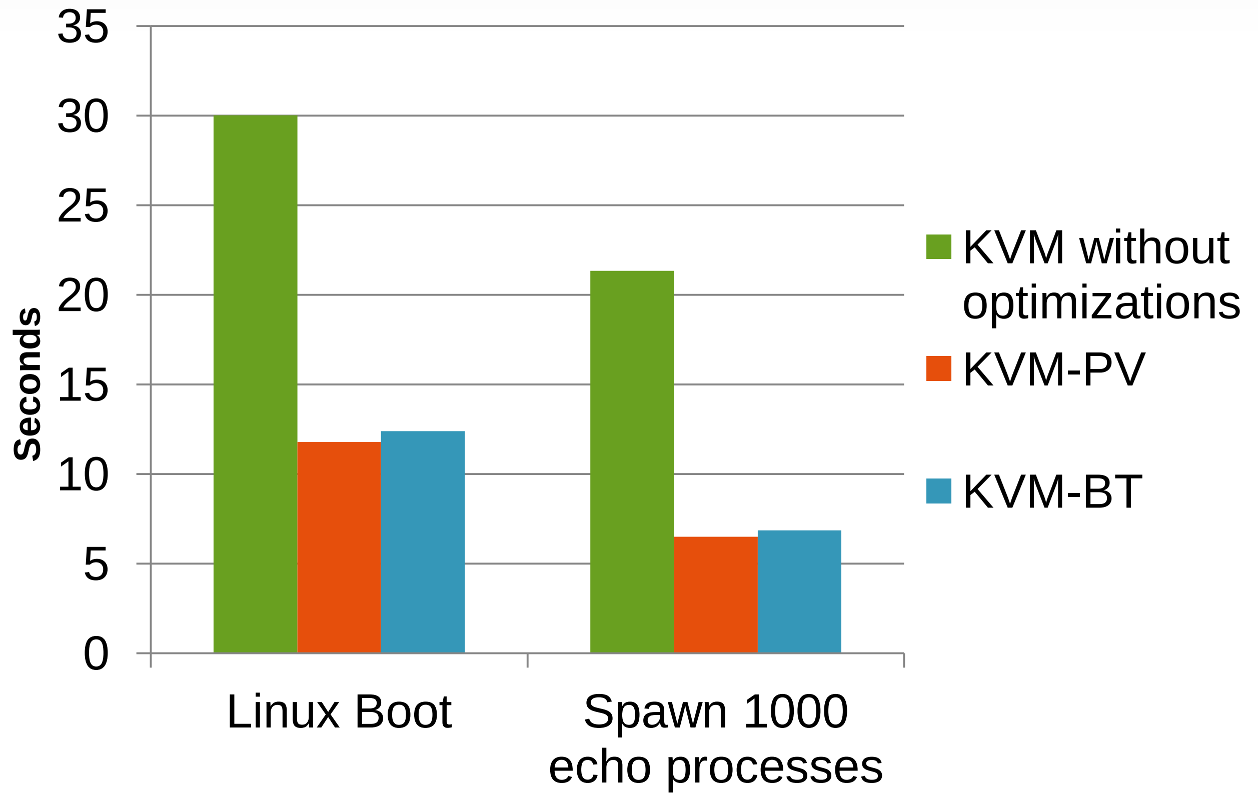
Solutions to Design Challenges



- Addressing Memory tracing issues:
  - Adaptive page resizing algorithm for addressing issue arising out of the use of Huge TLB and self modifying code
    - Dynamically splitting/merging traced pages

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Performance



Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Summary

- With “Adaptive Binary Translation” it’s possible to run an unmodified guest efficiently with embedded Power Architecture KVM
- Possible to mitigate memory tracing overhead using “Adaptive Page Resizing” and “Adaptive Data Mirroring”
- Performance is comparable to the existing PV solution

freescale™

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

Questions?

freescale™

Freescale, the Freescale logo, AltiVec, C-5, CodeTEST, CodeWarrior, ColdFire, C-Ware, the Energy Efficient Solutions logo, mobileGT, PowerQUICC, QorIQ, StarCore and Symphony are trademarks of Freescale Semiconductor, Inc., Reg. U.S. Pat. & Tm. Off. BeeKit, BeeStack, ColdFire+, CoreNet, Flexis, Kinetis, MXC, Platform in a Package, Processor Expert, QorIQ Qonverge, Qorivva, QUICC Engine, SMARTMOS, TurboLink, VortiQa and Xtrinsic are trademarks of Freescale Semiconductor, Inc. All other product or service names are the property of their respective owners. © 2011 Freescale Semiconductor, Inc.

freescale™