14th ANNUAL WORKSHOP 2018

SELINUX SUPPORT IN HFI1 AND PSM2

Dennis Dalessandro, Network SW Engineer

Intel Corp

4/2/2018

NOTICES AND DISCLAIMERS

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO SALE AND/OR USE OF INTEL PRODUCTS, INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT, OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life-saving, life-sustaining, critical control or safety systems, or in nuclear facility applications.

Intel products may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel may make changes to dates, specifications, product descriptions, and plans referenced in this document at any time, without notice.

This document may contain information on products in the design phase of development. The information herein is subject to change without notice. Do not finalize a design with this information.

Intel processors of the same SKU may vary in frequency or power as a result of natural variability in the production process.

Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.

Intel Corporation or its subsidiaries in the United States and other countries may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

Some features may require you to purchase additional software, services or external hardware.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations

Intel, the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Other names and brands may be claimed as the property of others.

Copyright © 2018 Intel Corporation. All rights reserved.

OPTIMIZATION NOTICE

Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel.

Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

LEGAL NOTICES AND DISCLAIMERS

- **Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.**
- **No computer system can be absolutely secure.**
- **Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.**
- **Intel, the Intel logo, Xeon and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.**
- **© 2018 Intel Corporation.**

WHAT IS SELINUX?

- **Not just something you disable on grub boot line!**
  - It was always the first thing I did on pretty much any system
- **Complicated**
  - Have been developing kernel code for years, yet SELinux still confuses me!
  - Have you tried to write a policy?
- **A good idea**
  - Security is an important topic but not one you should have to worry about
  - It needs to just work and have tools to support
  - Tools like audit2allow – makes life much easier

OK SO WHAT IS IT REALLY?!?

- **Detailed explanation well beyond our scope here**
- **See Dan Jurgen's presentation from 2016**
  - Provides a great overview of SELinux and how it applies to RDMA fabrics
- **Mandatory Access Control (MAC)**
  - Goes beyond normal file permissions
  - Multilevel Security (MLS)
    - Just because user is root doesn't mean they have an all access pass
    - Regular users can be granted different access depending on roles
- **RDMA fabrics have PKeys so enforce access controls on them**

Example:

User	Access Public?	Access Restricted?	Access Top Secret?
root	YES	NO	NO
user1234	YES	YES	NO
tux	YES	YES	YES

VERBS SUPPORT ONLY

- **Added to the kernel fairly recently (over the last few versions)**
  - Changes are in the IB core and SELinux core
  - Should work with any driver that supports IB verbs
- **Based on PKeys**
  - Is that the best or only option?
  - Time will tell.
- **User space tooling exists**
- **Does not support other protocols**
  - PSM in particular, which is why we have this talk
  - PSM is our preferred communication library

REQUIREMENTS FOR PSM SELINUX

- **Support large number of labels**
- **Use existing PKey scheme**
  - For now, who knows what the future will bring
- **Support kernel bypass**
- **Require no changes to user space**
  - Enable easy extension if we want to move away from being PKey based
- **Require no changes to IB core or SELinux core**
  - Enable easy extension if we want to move away from being PKey based
- **Be able to work with distro and IFS installations**
  - Only changing the hfi1.ko
- **Require minimal to no modifications to PSM library**
  - Changing user space only if we have to, so far so good

WHAT IS A PKEY?

- **PKey is 16bit non-cryptographic value**
  - Does not need to be a secure value
  - Knowing PKey doesn't mean you have access to it (AH! So SELinux ☺)
- **hfi1 hardware currently supports PKey checking**
  - hardware checks all incoming packets to ensure a valid PKey
  - send/recv contexts and SDMA engines have PKey hardware checks as well
- **hfi1 driver is flexible and can support as many PKeys as needed**

WHAT IS THIS JKEY THING AND WHY USE IT?

- **A JKey is a "Job Key"**
  - Unique to PSM
- **Allows splitting up partitions further than Pkey**
  - Many Jkeys in a single partition
- **Determined by the kernel**
  - Jkey = 0 has special meaning
  - The rest of it is up to software
    - Currently: Use the UID to come up with a JKey
  - JKey space is currently divided up into buckets (that we can change!)
    - admin users
    - kernel protocols
    - everyone else
- **Using JKey as security field allows flexibility**
  - To scale to thousands of labels to meet future security needs
  - For alternatives to PKey based security

USING JKEY TO HOLD THE SECURITY FIELD

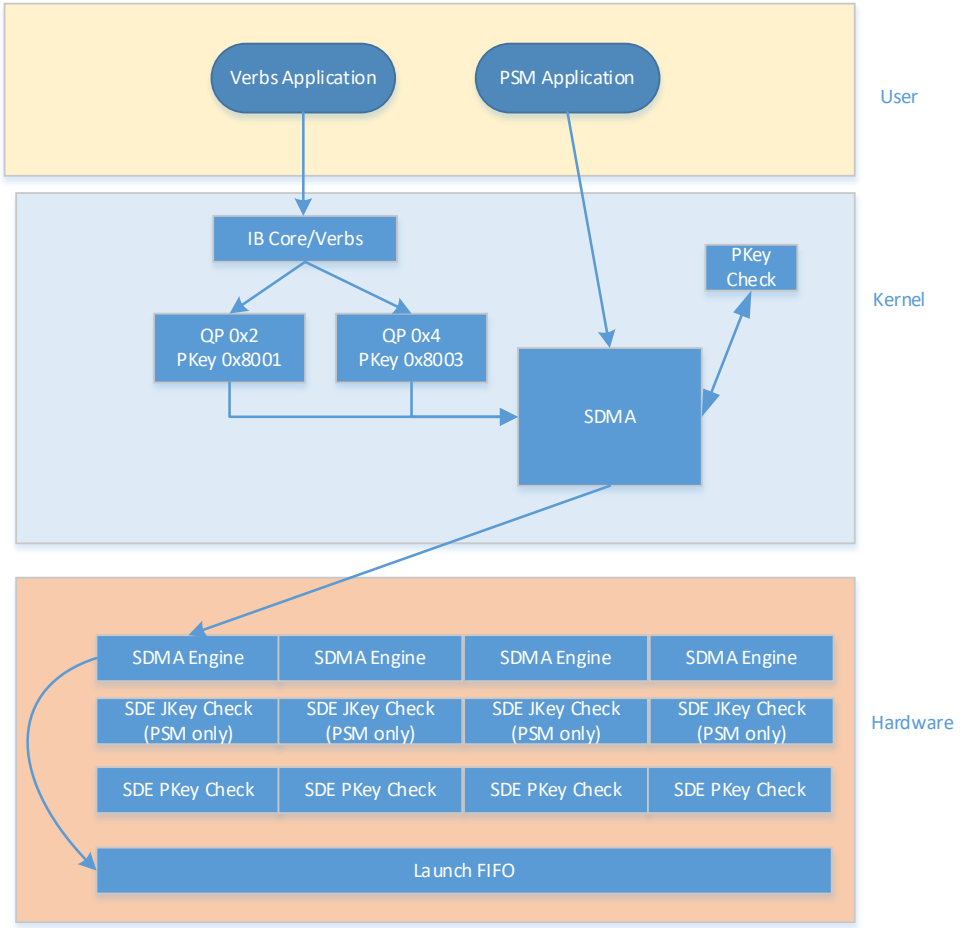
- **JKey can be used to hold a flexible security field**
- **We can use the JKey to hold the index of the PKey**
  - This way we can use the JKey hardware checks to validate the PKey
  - Could be something other than PKey some day
- **Three JKey buckets becomes Four buckets**
  - Exact mapping and format of the JKey is still being finalized
  - Must support sufficient number of JKeys and still maintain job separation

PSM DATA TRANSFERS

- **Two ways to send data**
  - Programmed I/O aka PIO
  - Send Direct Memory Access aka SDMA
  - Differ in how data goes from user space to the hfi1 hardware
- **Hardware supports both**
  - Tradeoffs to using one vs the other beyond the scope of this talk
- **PIO**
  - Kernel bypass
- **SDMA**
  - Goes through the kernel for SDMA engine programming

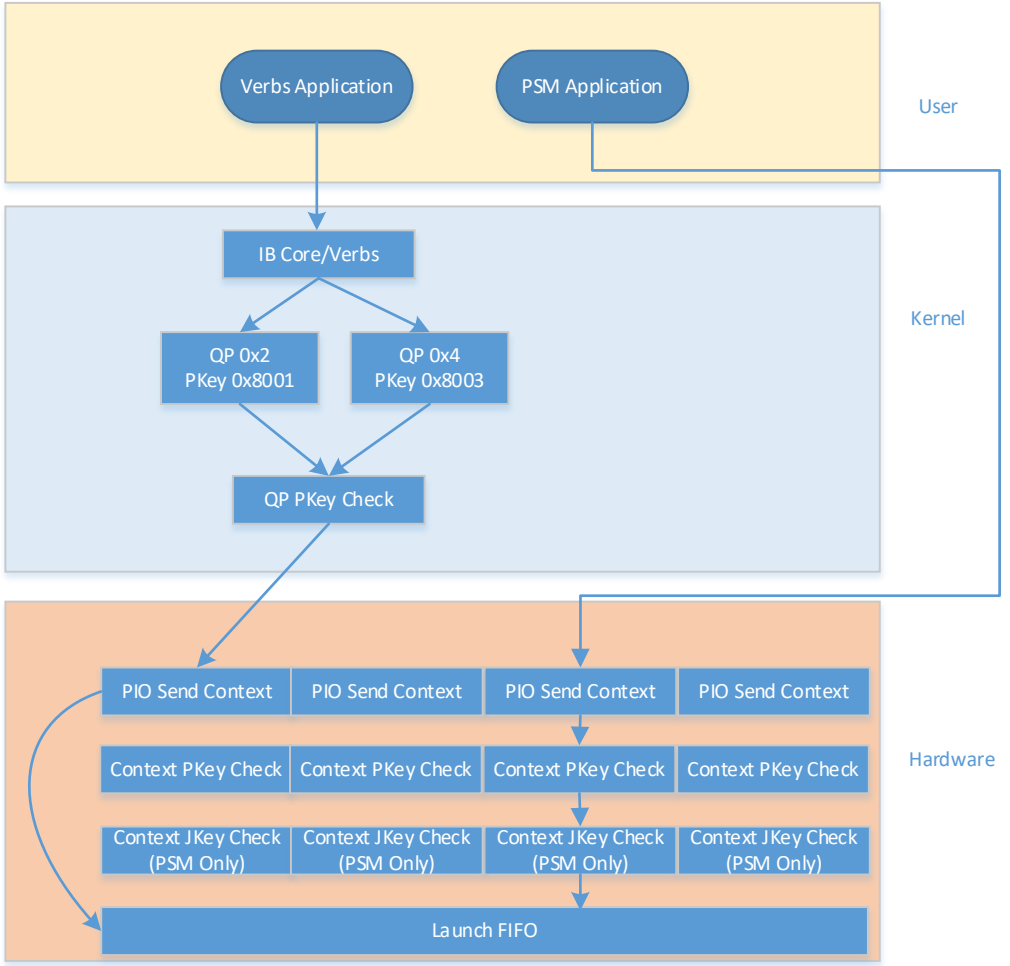
SDMA SEND BEFORE SELINUX

- **PSM calls into kernel**
- **PKey is checked in software**
  - Verbs and PSM share SDMA engines
  - Also means JKey can not be checked in hardware
  - Even verbs can have multiple PKeys in packets for a particular SDMA Engine



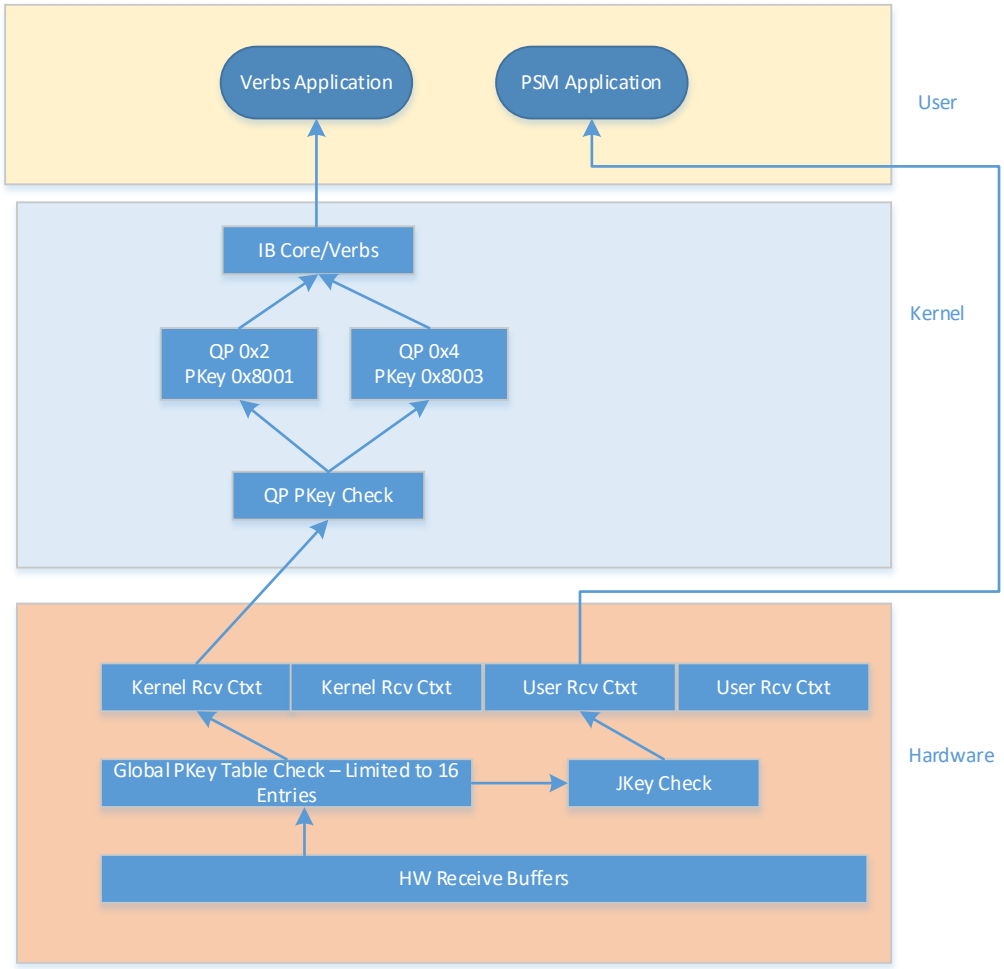
PIO SEND BEFORE SELINUX

- **verbs**
  - Goes through kernel
  - Multiple QPs and PKeys could be mapped to the same context
  - Thus no HW PKey checks
- **PSM**
  - Kernel bypass
  - Must use HW checks
    - Includes JKey check



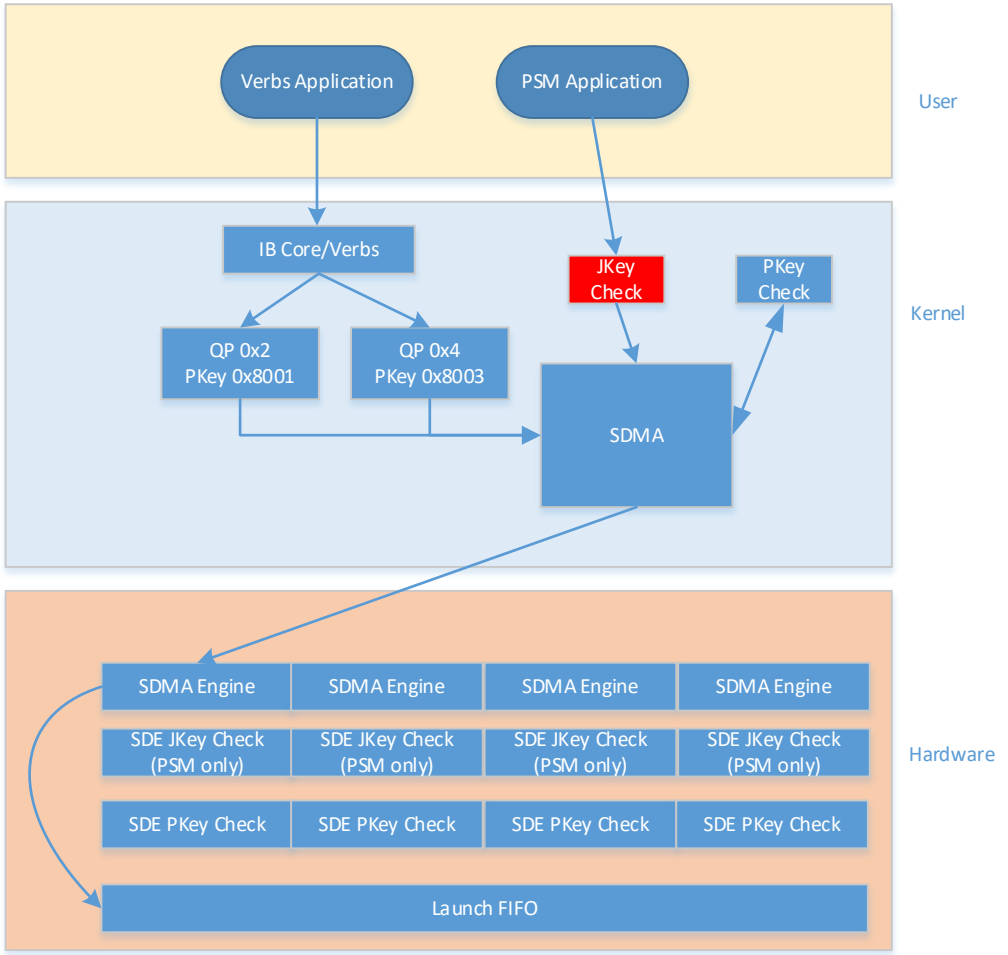
RECEIVING DATA BEFORE SELINUX

- **Verbs**
  - Global PKey table check
  - Limited to 16 Pkeys currently. We can support more.
  - Still has to be a check per QP
- **PSM**
  - Kernel bypass
  - Must use HW
  - Also includes JKey



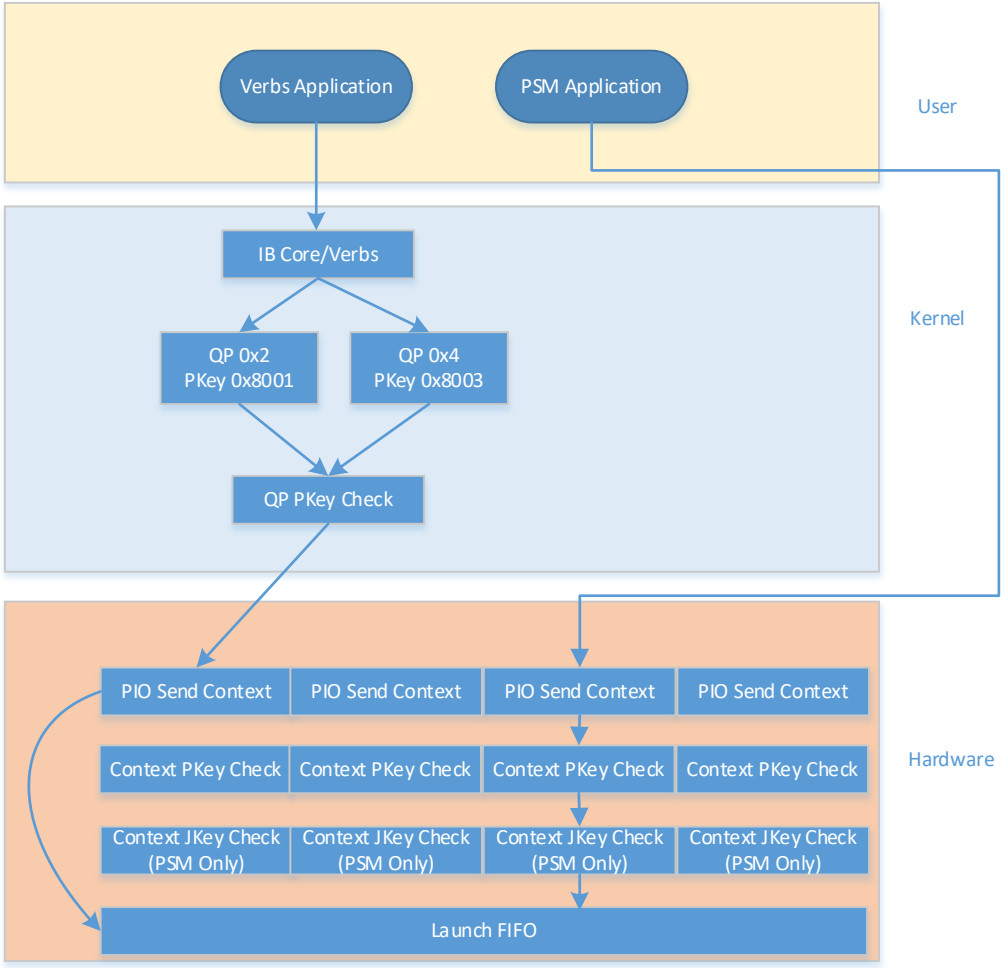
SELINUX SDMA SUPPORT

- **Verbs**
  - Still does PKey check in SW
- **PSM**
  - Still does PKey check in SW
  - Adds additional check of JKey in software
  - Can not use HW JKey checking since SDMA Engines are shared with verbs
  - Same reason we can not use HW PKey checking



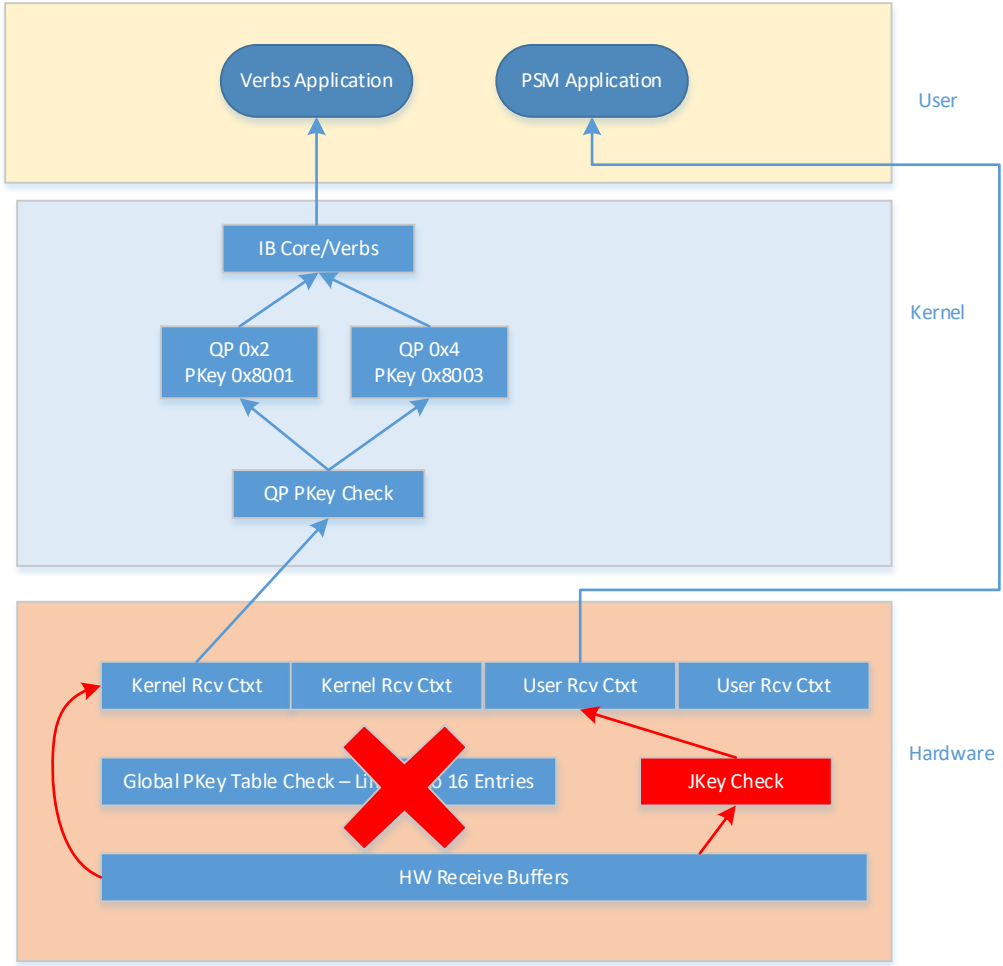
SELINUX PIO SUPPORT

- **Verbs**
  - Nothing changes
  - Pure SW based
- **PSM**
  - Nothing changes
  - Pure HW based



SELINUX RECEIVE SIDE SUPPORT

- **Verbs**
  - Disable PKey checks since
    - Required to support > 16 PKeys
  - Still have to do a PKey check for the QP in software so HW check really not needed anyway
- **PSM**
  - Use JKey check in HW
  - Preserves kernel bypass



CURRENT STATUS

- **Under development**
- **Targeting kernel 4.19 – rough estimate and subject to change**
- **So far no changes required to**
  - PSM
  - IB Core
  - SELinux Core

14th ANNUAL WORKSHOP 2018

THANK YOU

Dennis Dalessandro, Network SW Engineer

Intel Corp

4/2/2018