Using the NOVA Microhypervisor for Trusted Computing at Scale

Udo Steinberg
Agenda

❖ NOVA Microhypervisor Overview
  ➢ Architecture, Scalability
  ➢ Innovation Timeline

❖ Trusted Computing
  ➢ Verified vs. Measured Boot / SRTM vs. DRTM
  ➢ Intel Trusted Execution Technology (TXT)
  ➢ Trusted Platform Module (TPM)
  ➢ Integrity Measurement of NOVA and Root-PD

❖ Q & A
Formal Verification of Bare Metal Property™

Ultravisor™ Ultravisor Architecture

VM (Linux)
VM (Windows)
VM (Appliance)
VM (RTOS)
VM (Unikernel)

VMM
UART Multiplexer
VirtIO Socket Multiplexer
Network Multiplexer (Virtual Switch)

UART Driver
Storage Driver
Platform Manager
Service Manager
Network Driver
Host Applications

Master Controller (Root Protection Domain)

NOVA Microhypervisor (ARMv8-A or Intel x86-64)

BedRock Systems Inc

NOVA Microhypervisor
Udo Steinberg - FOSDEM 2024
Scaling NOVA from Embedded to Cloud Servers

**NOVA Microhypervisor**

Udo Steinberg - FOSDEM 2024

**Same NOVA Binary**

make ARCH=aarch64 BOARD=acpi

Same NOVA Binary

make ARCH=x86_64

### Embedded Devices

- **Allwinner** A64
  - NXP i.MX8
  - HiSilicon Hi3660
  - Rockchip RK3399
  - Texas Instruments J721E
  - Amlogic G12B, SM1
  - NXP i.MX8
- **Renesas** R-Car
- **Rockchip** RK3399
- **NVIDIA** Tegra X1/X2, Xavier
- **Qualcomm** Snapdragon 670
- **HiSilicon** Hi3660
- **Texas Instruments** J721E

### Cloud Servers

- **AWS Graviton2**
  - 64C Arm Neoverse N1 (Ares)
  - c6g.metal
  - AWS Graviton3
  - 64C Arm Neoverse V1 (Zeus)
  - c7g.metal

### Intel Xeon-SP

- **1st Gen Intel Xeon-SP**
  - 72T Intel Xeon Platinum 8124M (SKX)
  - c5n.metal
- **2nd Gen Intel Xeon-SP**
  - 96T Intel Xeon Platinum 8259CL (CLK)
  - m5n.metal
- **3rd Gen Intel Xeon-SP**
  - 128T Intel Xeon Platinum 8375C (ICX)
  - c6l.metal

### Intel Xeon-EP

- **4th Gen Intel Xeon-EP**
  - 72T Intel Xeon E5-2686 v4 (BDW-EP)
  - i3.metal

### Intel NUC Pro

- **Intel NUC8PN Pro**
  - 8T Intel Core i7-8665U (WHL)
- **Intel NUC12WS Pro**
  - 16T Intel Core i7-1270P (ADL)
- **Intel NUC13LC Pro**
  - 20T Intel Core i7-1370P (RPL)

### Intel Xeon-D

- **Intel Xeon-D**
  - 32T Intel Xeon D-2775TE (ICX-D)

### Intel Xeon Silver

- **2nd Gen Intel Xeon-Silver**
  - 48T Intel Xeon Silver 4214R (CLK)
- **3rd Gen Intel Xeon-Silver**
  - 80T Intel Xeon Silver 4316 (ICX)
- **4th Gen Intel Xeon-Silver**
  - 24T Intel Xeon Silver 4410Y (SPR)

### ARM

- **Up Squared**
  - 4C Intel Pentium N4200 (APL)
  - UP Squared v2
  - 4C Intel Pentium J6426 (EHL)
  - youyeeetoo X1 SBC
  - 4C Intel Celeron N5105 (JSL)

All trademarks and brand names are the property of their respective owners. All company and product names used in these slides are for identification purposes only. Use of these names, trademarks and brands does not imply endorsement.
Both using UEFI firmware with ACPI support
NOVA / Server: Arm Neoverse V1 (c7g.metal)

NOVA Microhypervisor #8b6c28a (aarch64): Feb 1 2024 10:57:11 [gcc 13.2.0] 6

XSDT: 0x15e1fe98 OEM:AMAZON TBL:GRVTN003 REV: 1 LEN:    140 (0x9f) ok
FACP: 0x15e1fb98 OEM:AMAZON TBL:GRVTN003 REV: 6 LEN:    276 (0x30) ok
HEST: 0x15e1fa98 OEM:AMAZON TBL:GRVTN002 REV: 1 LEN:    168 (0x8a) ok
GTDT: 0x15e1fd18 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:     96 (0xbb) ok
APIC: 0x15e17518 OEM:AMAZON TBL:GRVTN003 REV: 5 LEN:   5268 (0x45) ok
SRAT: 0x15e1d898 OEM:AMAZON TBL:GRVTN003 REV: 3 LEN:   1496 (0x2f) ok
SLIT: 0x15e1fe18 OEM:AMAZON TBL:GRVTN003 REV: 1 LEN:     45 (0x24) ok
MCFG: 0x15e1df18 OEM:AMAZON TBL:GRVTN003 REV: 1 LEN:    108 (0xad) ok
PPTT: 0x15e19098 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:   3262 (0xe4) ok
SDEI: 0x15e1e918 OEM:AMAZON TBL:GRVTN003 REV: 1 LEN:     36 (0x4f) ok
IORT: 0x15e1e018 OEM:AMAZON TBL:GRVTN003 REV: 0 LEN:    800 (0xb2) ok
SSDT: 0x15e1ce18 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:   1500 (0x8c) ok
SSDT: 0x15e1e518 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:     74 (0xed) ok
SSDT: 0x15e1ce18 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:   1500 (0x8c) ok
SPCR: 0x15e1ff98 OEM:AMAZON TBL:GRVTN003 REV: 2 LEN:     80 (0x86) ok
ACID: Version 6.1 Profile 4 Features 0x30100 Boot Oxl
PCi: VERSION 1.1 Status 0000
PCie: Console 0000:00.0

[56] CORE: 01:24:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[56] GICC: REGS
[56] GICH: REGS APR:1 LR:4
[56] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[57] CORE: 01:25:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[57] GICC: REGS
[57] GICH: REGS APR:1 LR:4
[57] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[58] CORE: 01:26:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[58] GICC: REGS
[58] GICH: REGS APR:1 LR:4
[58] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[59] CORE: 01:27:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[59] GICC: REGS
[59] GICH: REGS APR:1 LR:4
[59] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[60] CORE: 01:28:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[60] GICC: REGS
[60] GICH: REGS APR:1 LR:4
[60] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[61] CORE: 01:29:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[61] GICC: REGS
[61] GICH: REGS APR:1 LR:4
[61] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[62] CORE: 01:30:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[62] GICC: REGS
[62] GICH: REGS APR:1 LR:4
[62] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[63] CORE: 01:31:00.00 Arm Neoverse V1 (Zeus) r1p1 PA:5 XNX:1 GIC:3 (EL2)
[63] GICC: REGS
[63] GICH: REGS APR:1 LR:4
[63] TIMR: EL2p:10L EL1v:11L 1050000000 Hz
[ 0] TIME: 8842ms 16ms/786ms
[ 0] ROOT: No image

NOVA Microhypervisor
Udo Steinberg - FOSDEM 2024
NOVA Microhypervisor #8b6c28a (x86_64): Feb  1 2024 10:56:51 [gcc 13.2.0]

XSDT: 0x74bfc0e8 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 560 (0xd8) ok
FACP: 0x74bf7000 OEM:INTEL TBL:INTEL ID REV: 6 LEN: 276 (0x174) ok
ERST: 0x74bf8000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 2329 (0xc8) ok
SSDT: 0x74bf9000 OEM:INTEL TBL:RAS_ACPI REV: 2 LEN: 1861 (0x4f) ok
SSDT: 0x74bf9000 OEM:INTEL TBL:ADDRXLAT REV: 2 LEN: 1864 (0x19) ok
SSDT: 0x74bfa000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 560 (0xd8) ok
APIC: 0x74bf4000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 384 (0xc4) ok
MCFG: 0x74bf3000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 100 (0xff) ok
HPET: 0x74bf5000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 56 (0x51) ok
MSCT: 0x74bf3000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 60 (0x44) ok
APIC: 0x74bbb000 OEM:INTEL TBL:INTEL ID REV: 4 LEN: 1118 (0x23) ok
SLIT: 0x74bba000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 48 (0x44) ok
SRAT: 0x74bb6000 OEM:INTEL TBL:INTEL ID REV: 3 LEN: 12848 (0x2a) ok
OEM4: 0x74b2d000 OEM:INTEL TBL:CPU CST REV: 2 LEN: 802145 (0x1b) ok
SRAT: 0x74bb6000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 18 (0x12) ok
OEM1: 0x74b09000 OEM:INTEL TBL:CPU EIST REV: 2 LEN: 558729 (0x46) ok
OEM2: 0x6e36d000 OEM:INTEL TBL:CPU  CST REV: 2 LEN: 802145 (0x1b) ok
HPET: 0x74bf5000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 380 (0x4a) ok
ACPI: 0x6e2b9000 OEM:INTEL TBL:SSDT  PM REV: 2 LEN: 242286 (0x6c) ok
SSDT: 0x6e2ba000 OEM:INTEL TBL:SSDT  PM REV: 2 LEN: 242286 (0x6c) ok
DMAR: 0x74b07000 OEM:INTEL TBL:INTEL ID REV: 1 LEN: 368 (0x9c) ok
ACPI: Version 6.2 Profile 4 Features 0x4a5 Boot 0x10
FACS: Hardware 0x0 Flags 0x0 Wake 0x0/0x0

PCIE: 0x80000000 Segment 0xe000  Bus 0x00  D0 00-0ff
- [PCIe] 0x80169a0 08-80-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806f9a4 08-80-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806f998 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806e000 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d99c 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d99b 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d99a 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d998 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806d996 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d994 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d992 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d990 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d988 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806d986 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d984 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d982 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d980 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806d978 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d976 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d974 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d972 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806d970 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d968 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d966 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d964 06-00-00 DO PCIe 0000:00.00.0
- [PCIe] 0x806d962 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d960 08-80-00 DO PCIe PMI FLR 0000:00.01.0
- [PCIe] 0x806d958 13-00-00 DO PCIe 0000:00.02.0
- [PCIe] 0x806d94e ff-00-00 DO PMI 0000:00.11.0

[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
[ 48] CORE: 35-29-8 using 35 (HYP)
[ 48] RDTA: L3:15 L2:0 Mib15
[ 48] CPU: State:0x13 (0xfffff0fa:0x951dafa:0x0a)
NOVA Microhypervisor: Innovation Timeline

APIC Register Virtualization
- X2APIC Support
- Atom SoC Support

Trusted Execution Technology Support
- NOVA Integrity Measurement
  - Minimal TPM Support
  - Extensible Kernel Heap

Improved Capability Management and Reference Counting

Portable PCI Infrastructure
- PCI Segment Groups
- NIST FIPS 180-4 SHS
- Root-PD Integrity Measurement
- TPM 1.2 / 2.0 PCR Commands

PORTABLE RCU INFRASTRUCTURE
- Generic Completion Wait
- PCI Capabilities

NOVA release cadence is ~2 months

Deep Core C-States
- Multiple TPM Interfaces
- Crypto Agile Event Log
What Problem Does Trusted Computing Solve?

❖ Once you have a formally verified software stack
  ➢ and a compiler that produced a qualified set of binaries for the target architecture

❖ How do you ensure that some computer is running those binaries
  ➢ and not some other (malicious) software instead
  ➢ before you entrust that computer with your data or secrets

❖ In other words, how can you
  ➢ either restrict the software that a computer will launch
  ➢ or determine what software has been launched on a computer
**Verified Boot: Static Root of Trust**

- **Boot policies are enforced during the boot process**
- **Starting with the Core Root of Trust for Verification, the currently executing module verifies the integrity of the next module against a boot policy (e.g. UEFI db/dbx) ⇒ Chain of Trust**
- **Integrity measurement is a cryptographic hash ⇒ unique + indicative to changes in the module**
Integrity measurements are extended into TPM PCRs during the boot process.

Starting with the Core Root of Trust for Measurement, the currently executing module extends the launch integrity measurement for the next module into the TPM.
DRTM Flow lets system boot into an untrustworthy state (initially)

- Measured Launch later “resets” system into a trustworthy safe state
- Takes control of all CPUs and forces them down a protected and measured code path
Intel Trusted Execution Technology (TXT / CBnT)

❖ Provides a Dynamic Root of Trust (DRTM)

❖ Prerequisites
  ➢ CPU support (SMX features)
  ➢ TXT-capable chipset (DMA protection)
  ➢ TPM 2.0 (preferably) or 1.2
  ➢ SINIT Authenticated Code Module (ACM)

❖ Use Cases
  ➢ Remote Attestation (via TPM Quote)
  ➢ Local Attestation (via Launch Control Policy)
GETSEC[SENTER] Late Launch Sequence

- **Design Decision**
  - NOVA late-launches itself (~650 LOC)

- **ILP**
  - Load ACM
  - Load MLE
  - Launch (DRTM)
  - SENTER Event

- **RLPs**
  - GETSEC[SENTER]
  - ILP broadcasts SENTER message
  - Each CPU responds to SENTER event
  - Each CPU issues ACK
  - ILP continues once all ACKs received

- **SENTER Event**
  - ILP continues once all ACKs received

- **SENTER**
  - Unrecoverable failure causes TXT Shutdown

- **SINIT**
  - Launches ACM
  - Launches MLE

- **MLE (NOVA)**
  - MLE (NOVA)

- **Measured Launch Environment (NOVA)**
  - Load ACM
  - Load MLE
  - Launch (DRTM)

- **GETSEC[WAKEUP]**
  - GETSEC[WAKEUP] for RLPs
  - SINIT launches MLE
  - ILP launches SINIT ACM

- All CPUs in secure environment

---

NOVA Microhypervisor
Udo Steinberg - FOSDEM 2024
A verifier can use the crypto agile event log to recompute/validate the composite value in each PCR.
Integrity Measurement (NOVA)

What (not) to include in integrity measurement requires careful consideration

- Sensitive to: modifications in NOVA measured region, command-line parameters
- Insensitive to: hardware platform, memory map, multiboot modules, code self-patching

Build system prints reference integrity measurements for NOVA
Extending the Chain of Trust into User-Mode

- Compute the Root-PD launch integrity measurement
  - Define the attestable region to measure
  - Compute the integrity measurement of that region
    - Either using TPM Hash/Event Sequence (maximum crypto agility)
    - Or using a SW implementation in NOVA (maximum performance)
- Drive the Trusted Platform Module
  - Different MMIO Interfaces
  - Different TPM Families
- Append integrity measurement to the TPM event log

Time for measuring a 2 MiB image into a PCR

<table>
<thead>
<tr>
<th></th>
<th>NOVA (perf)</th>
<th>TPM2 (agility)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 + 2 ms</td>
<td>13700 ms</td>
</tr>
</tbody>
</table>
Integrity Measurement (Root Protection Domain)

- NOVA must measure Root-PD **before** launching it
  - Attestable region must be conveyed to NOVA without any invocations from Root-PD
  - Using ELF PHDRs (first non-writable PT_LOAD segment)

- **Hash Computation**
  - Digests computed by NOVA in C++ (for all supported hash algorithms)
  - Subsequently extended into PCR 19 (for all PCR banks)

- **NOVA implements the FIPS 180-4 Secure Hash Standard** (~130 LOC)
  - SHA1: -160
  - SHA2: -224, -256, -384, -512
Trusted Platform Module Infrastructure in NOVA

- Supports all TPM interface types (FIFO and CRB) (~250 LOC)
- Supports relevant command `subset` (Family 1.2 and 2.0) (~500 LOC)
  - Determine TPM capabilities (PCRs, Algorithms, …)
  - Perform PCR operations

- TPM localities control PCR access
  - Loc 2 belongs to NOVA Microhypervisor
  - Loc 1 belongs to User-Mode Environment
  - Loc 0 is for Legacy Use

<table>
<thead>
<tr>
<th>Locality</th>
<th>Usage</th>
<th>Can Extend</th>
<th>Can Reset</th>
<th>Next Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>CRTM</td>
<td>0-18, 23</td>
<td>17-22</td>
<td>PCR 17</td>
</tr>
<tr>
<td>3</td>
<td>SINIT ACM</td>
<td>0-20, 23</td>
<td>16, 20-23</td>
<td>PCR 17</td>
</tr>
<tr>
<td>2</td>
<td>NOVA</td>
<td>0-23</td>
<td>16, 20-23</td>
<td>PCR 19</td>
</tr>
<tr>
<td>1</td>
<td>Root PD</td>
<td>0-16, 20, 23</td>
<td>16, 23</td>
<td>PCR 20</td>
</tr>
</tbody>
</table>
Confidential & Trusted Computing Building Blocks

❖ **Availability**
➢ Cache & Memory Bandwidth Allocation Technology (CAT/CDP/MBA) - since 22.26

❖ **Integrity**
➢ Control-Flow Enforcement Technology (CET IBT+SSS) - since 22.17

❖ **Confidentiality**
➢ Total Memory Encryption with Multiple Keys (TME-MK) - since 22.52

❖ **Measured Launch & Attestation**
➢ Trusted Execution Technology (TXT/CBnT) - since 23.26
Questions and Discussion

The NOVA microhypervisor is licensed under GPLv2

Releases: https://github.com/udosteinberg/NOVA/tags

More Information: bedrocksystems.com and hypervisor.org