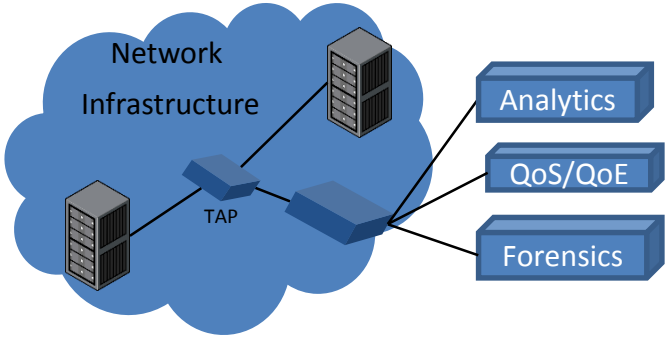

DPDK
DATA PLANE DEVELOPMENT KIT
Userspace 2015 | Dublin
Hyperscan
Software Pattern Matching

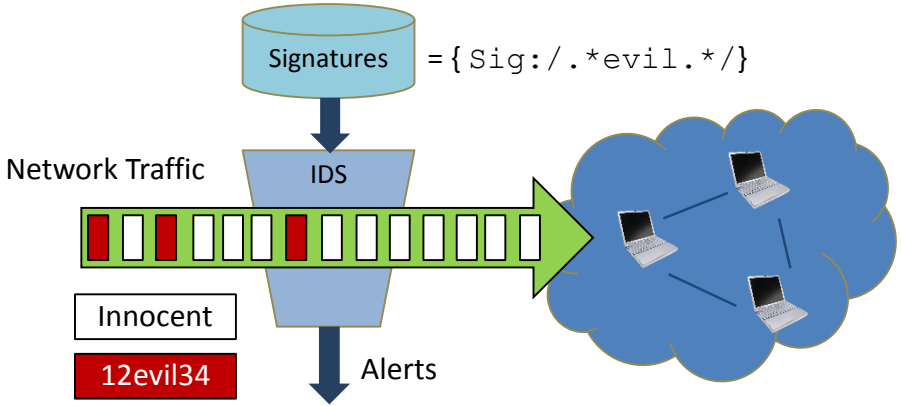
DPI Overview

DPI is a function that classifies Packets with two primary methods

- **Parsing** – Identifies application based on protocol and content

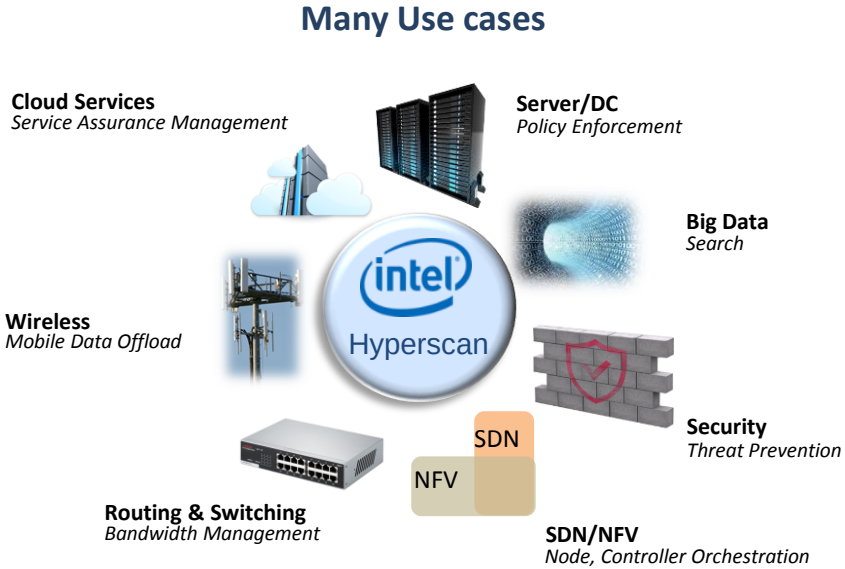


- **Pattern Matching** – Matches signatures in the packet to a database using RegEx or Fixed Strings
  - RegEx = Higher compute, simple to manage
  - Fixed String = Lower compute, complex to manage



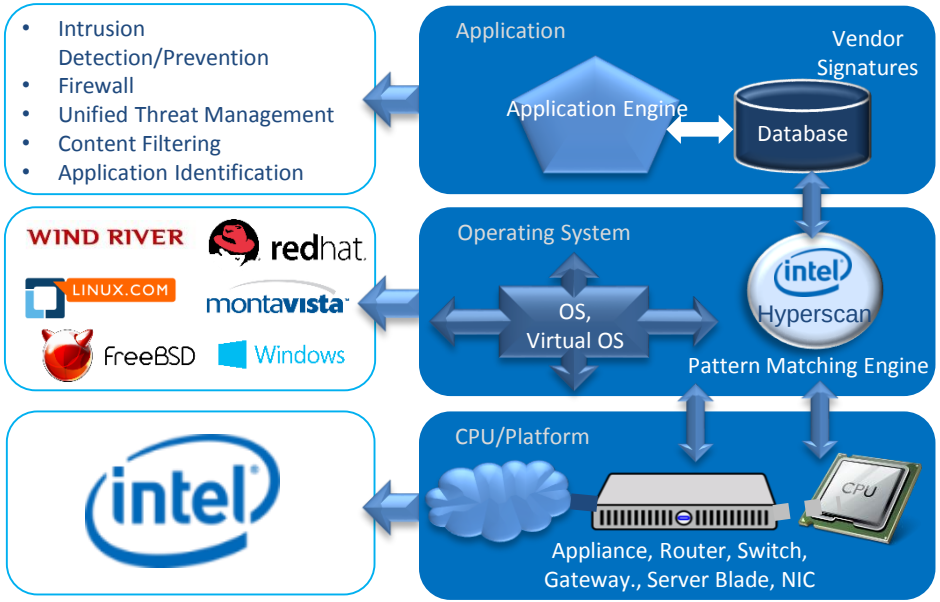
Pattern Matching

- Pattern matchers are at the heart of most security applications.
- As threats become complex, more intensive inspection is needed, but without application slow-down.
- Purpose-build hardware may cope with line-rate performance but time-to-market and maintenance cost is high.
- Software pattern matching provides the performance and scalability needed for the rapidly changing landscape.



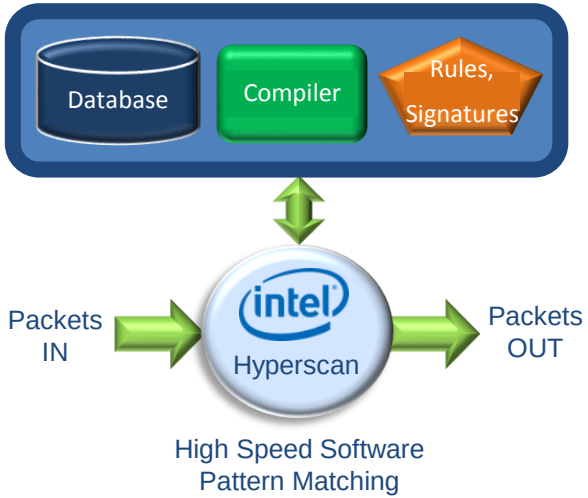
Hyperscan

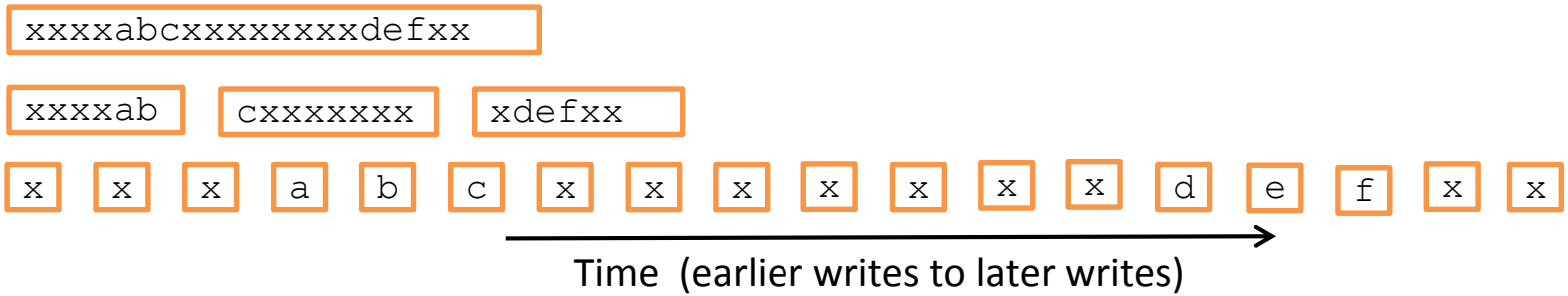
- Software Pattern Matching engine
  - Regex and Fixed-string matching
  - High performance
  - Low latency, compile time, memory
- Scales IA (Atom to Xeon)
  - Utilizes SIMD (SSE.x) for highest performance
- Portable, Easy to Integrate
  - Simple API; 32/64-bit systems
  - OS independent
- Recent Release
  - Hyperscan 3.4



Hyperscan Structure Summary

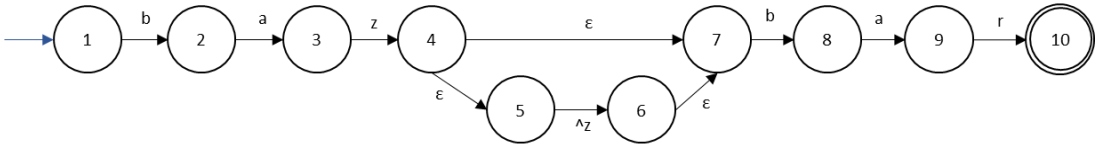
- Regular expressions are parsed into state machines.
  - Non-deterministic finite automata (NFA)
  - Deterministic finite automata (DFA)
- Engines are compiled into databases in terms of bytecode.
- During runtime, bytecode are used to search for patterns in data streams.
  - Block/streaming mode



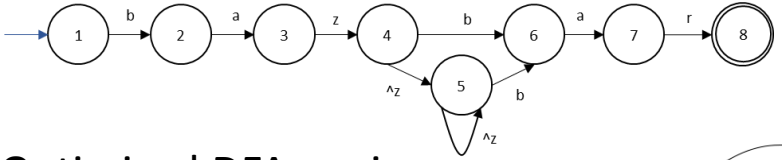


Example Automata Engines

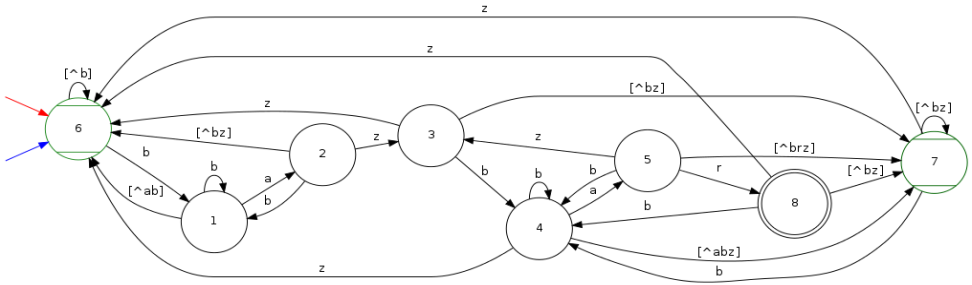
- Sample regular expression
  `/baz[^z]*bar/`
  Search string
  `"babazcbar"`
- NFA engine



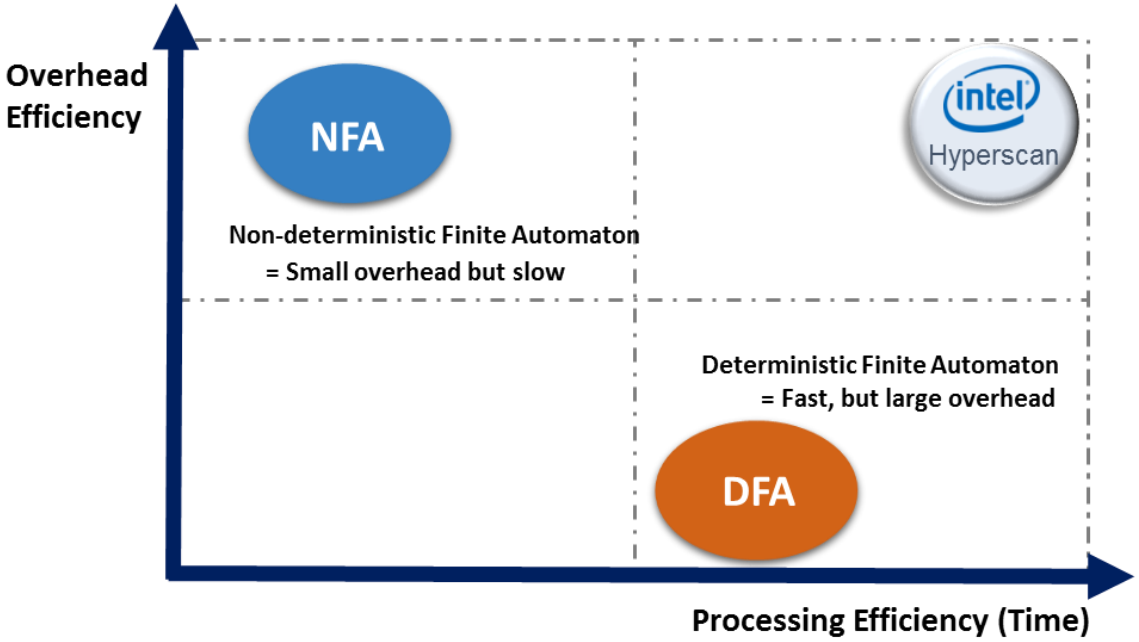
- DFA engine



- Optimized DFA engine



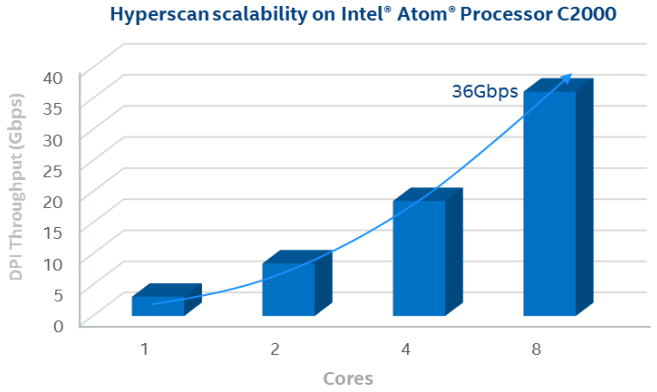
Performance Tradeoffs

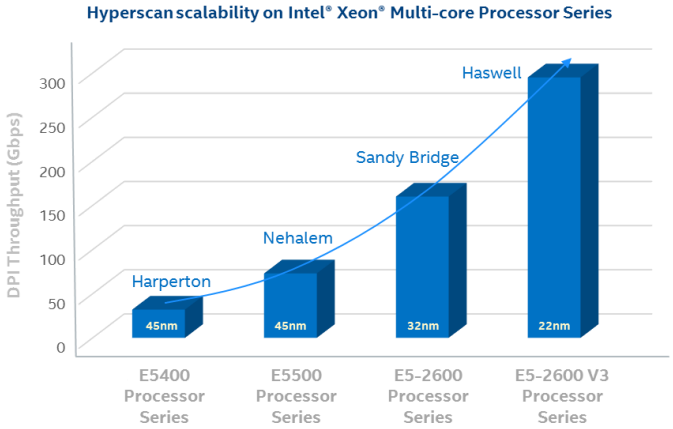


Hyperscan Performance

- Using Tier-1 OEM commercial IPS signature database
- HTTP test traffic; real world
- Rangeley (8-core, 2.4Ghz): ~3Gbps (1 core) scaling to 36Gbps (8-core)
- Haswell-EP: 293Gbps
  - Intel® Xeon ® CPU E5-2658 v3 @ 2.20GHz
  - With hyperthreading

Note: Numbers are subject to change using different benchmarking





IA Drives Performance

- Cache rich architecture
  - High bandwidth to Level 1 and Level 2 cache
  - Large L2 and L3 allows matching tables for literal matching to stay cache resident
  - Large L2 is unshared which means, unlike much of IA competition, scaling keeps going – unshared L2 bandwidth is per-core not per-chip
- Hyperthreading enables additional performance (15-20% is typical)
- Instruction sets
  - Process large numbers of characters using SIMD: SSE2, SSSE3
  - SIMD operations are resource friendly and fast on IA; enables large matching engines e.g. NFAs with big state counts
  - AVX2.0 enables processing of large amounts of input data in one step
  - BMI1/BMI2 also a 1:1 match for many pattern matching primitives: PEXT/PDEP replace a 10-30 instruction loop with 1 instruction

Hyperscan
(Software DPI)

DPDK

Thanks