ACM SIGCOMM 2010 Session 6: Novel Implementations of Network Components

Viktor Puš

Brno University of Technology, Faculty of Information Technology
Božetechova 2, 612 00 Brno, CZ
ipus@fit.vutbr.cz

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Obsah

- **SwitchBlade**
- **PacketShader**
- **EffiCuts**
- **vlastní poznámky**
- **zajímavé reference**

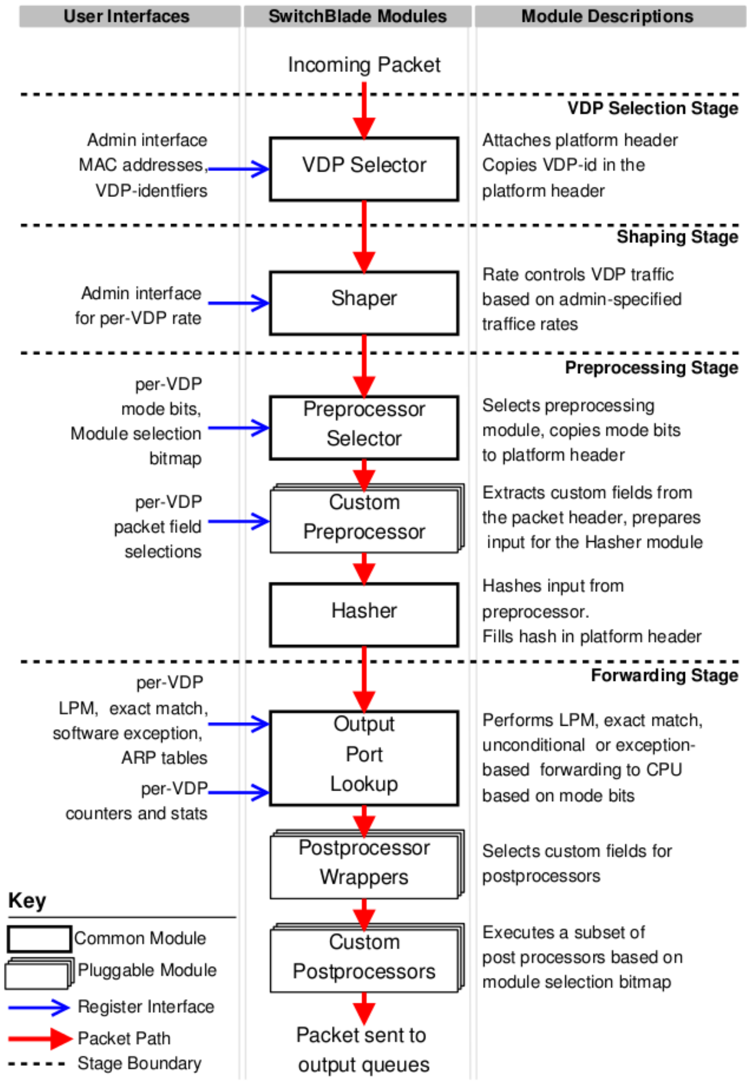
SwitchBlade - úvod

- „A platform for rapidly deploying custom protocols on programmable hardware“
- NetFPGA, wirespeed
- Cíle:
  - Rapid development and deployment on fast hardware
  - Customizability and programmability
  - Parallel custom data planes on a common hardware platform

SwitchBlade - návrh

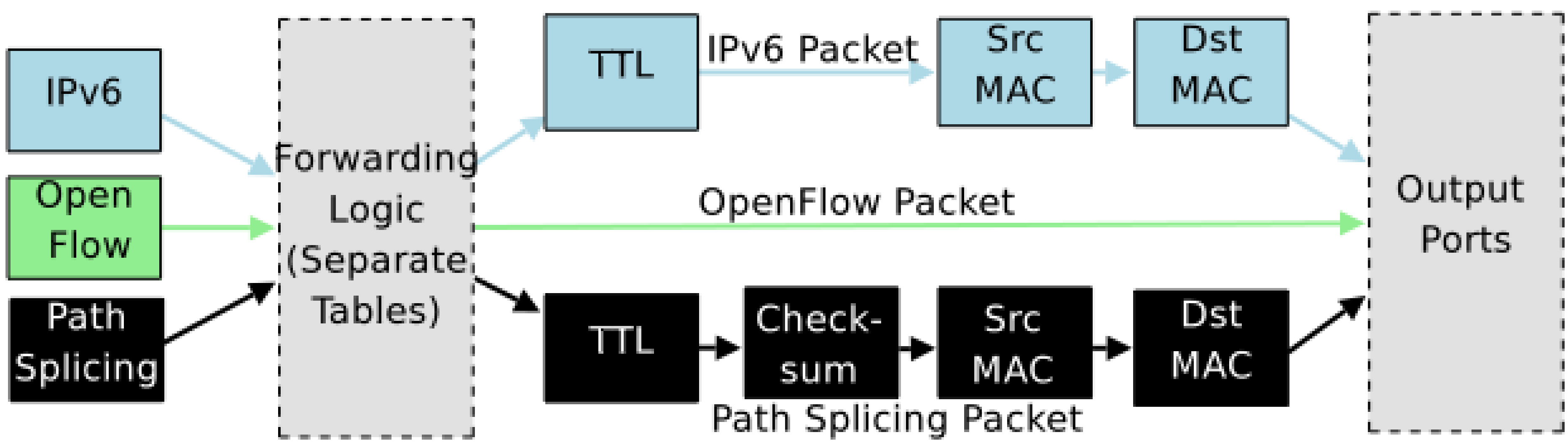
- Programovatelná pipeline
- Možnost vložení vlastních modulů

Field	Value	Description/Action
Mode	0	Default, Perform LPM on IPv4 destination address
	1	Perform exact matching on hash value
	2	Send packet to software for custom processing
	3	Lookup hash in software exceptions table
Module Selector Bitmap	1	Source MAC not updated
	2	Don't decrement TTL
	4	Don't Calculate Checksum
	8	Dest. MAC not updated
	16	Update IPv6 Hop Limit
	32	Use Custom Module 1
	64	Use Custom Module 2
	128	Use Custom Module 3



SwitchBlade - výsledky

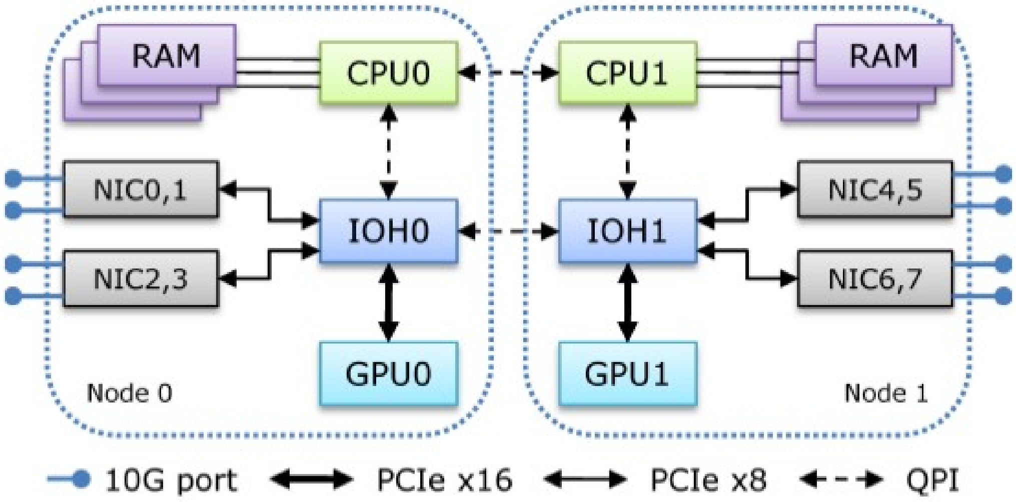
- Implementace ověřena na třech protokolech:
  - OpenFlow
  - IPv6
  - PathSplicing

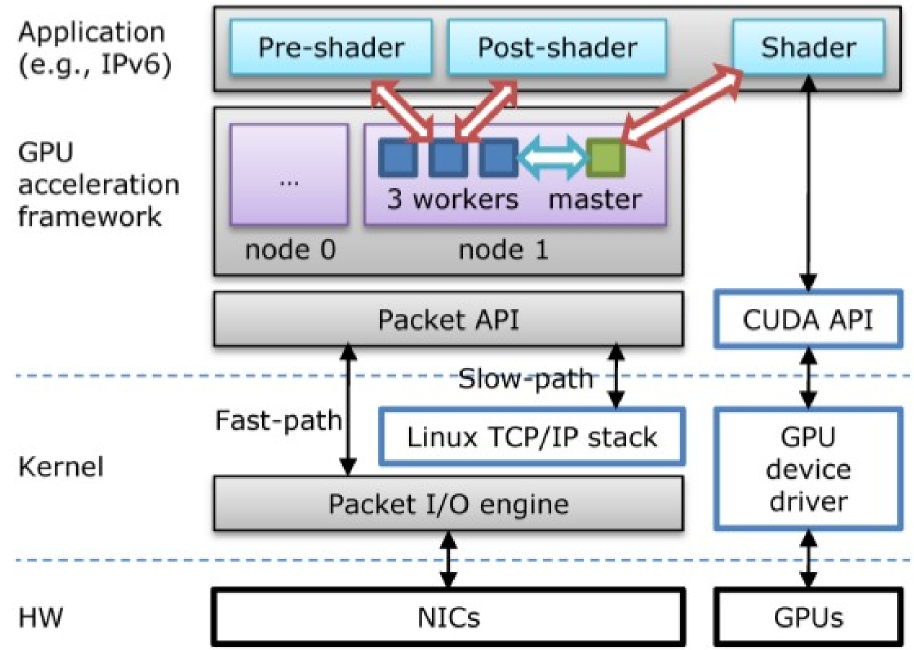


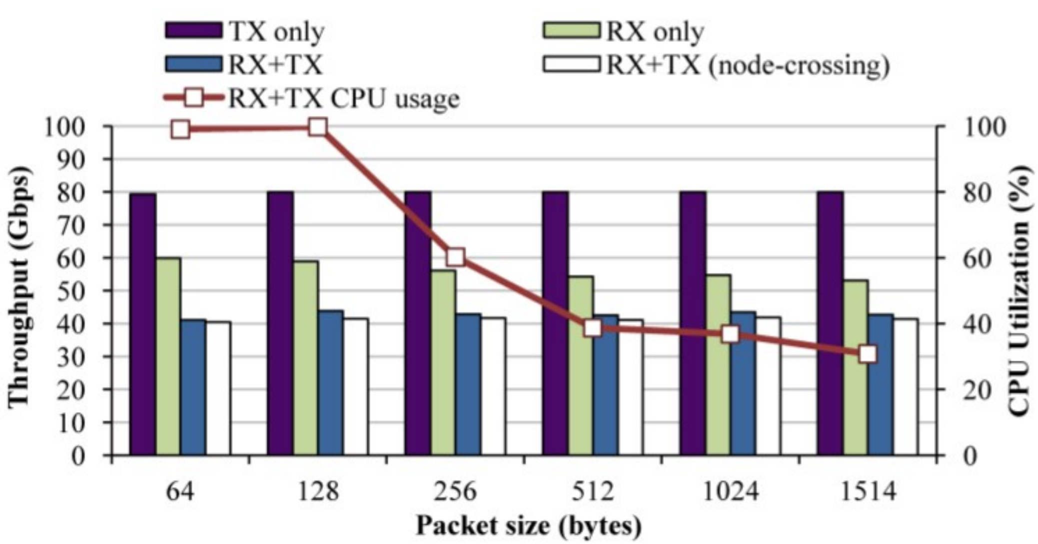
PacketShader - úvod

- Zpracování paketů na PC za $7000
  - 2 x NVIDIA GTX480
  - 2 x Intel Xeon X5550
- Aplikace: IPv4 a IPv6 forwarding, OpenFlow switching, Ipsec tunneling
- Výsledky: 39 Gbps na 64B IPv4 paketech

PacketShader - architektura







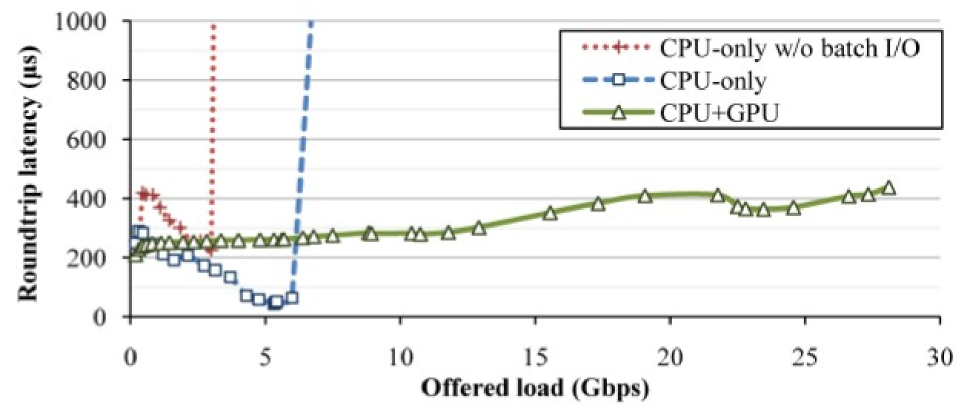
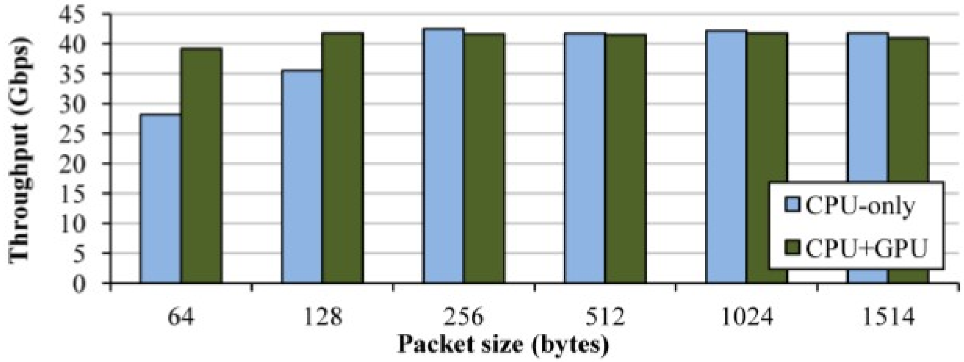
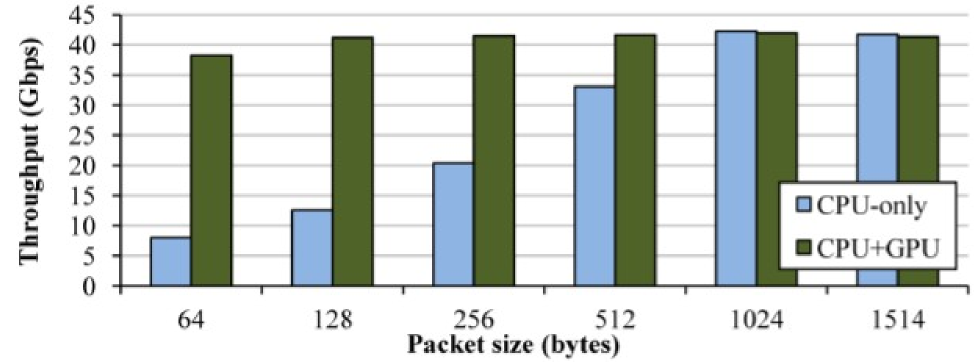


Figure 12: Average roundtrip latency for IPv6 forwarding

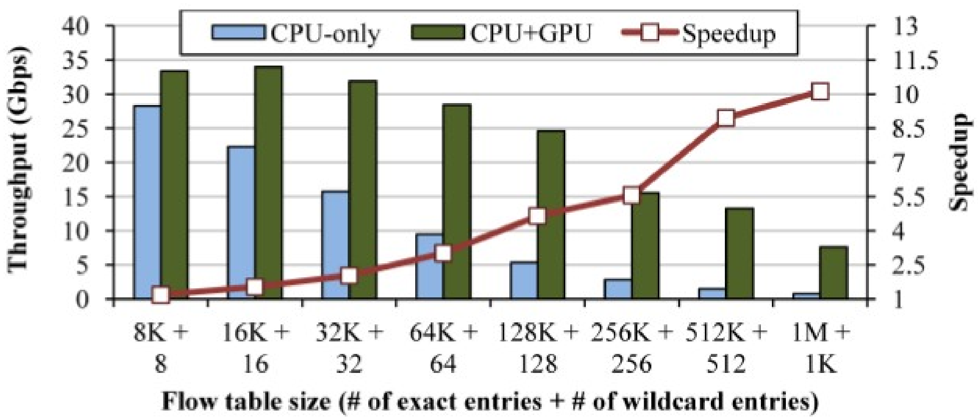
PacketShader - výsledky



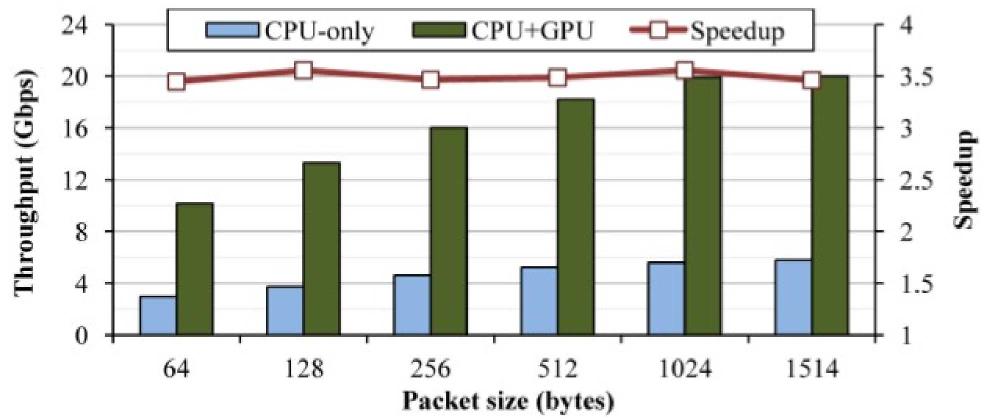
(a) IPv4 forwarding



(b) IPv6 forwarding



(c) OpenFlow switch (with 64B packets)



(d) IPsec gateway

EffiCuts - úvod

- Klasifikační algoritmus, zlepšení HyperCuts
- 4 zlepšení:
  - Separable trees
  - Selective tree merging
  - Equi-dense cuts
  - Node Co-location

EffiCuts – Separable trees

- HiCuts i HyperCuts dělí prostor na stejné části
- Je-li pravidlo ve více částech, je replikováno
    - (i 10 000x)
- Rozdělení pravidel na „velká“ a „malá“ v jednotlivých dimenzích

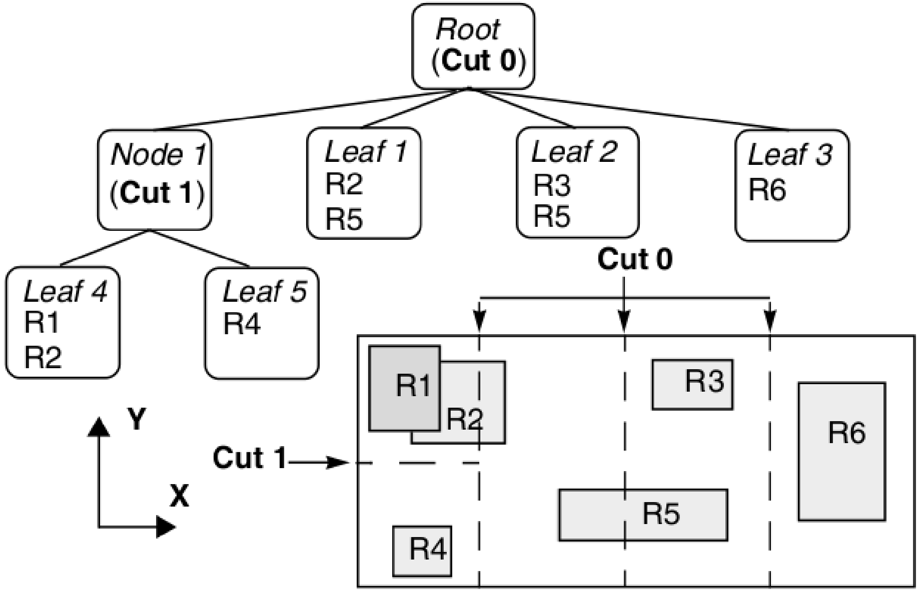


FIGURE 1. HiCuts Example in a 2D Rule Space

EffiCuts – Selective tree merging

- Až 26 kategorií pravidel → stromů
  - V praxi do 15
- Hloubka stromu vzniklého sloučením je většinou menší než součet hloubek
- Ale je větší šance replikace pravidel
- Heuristika

EffiCuts – Equi-dense cuts

- Prostor se dělí tak, aby části obsahovaly stejný počet pravidel
  - Ne aby byly stejně velké

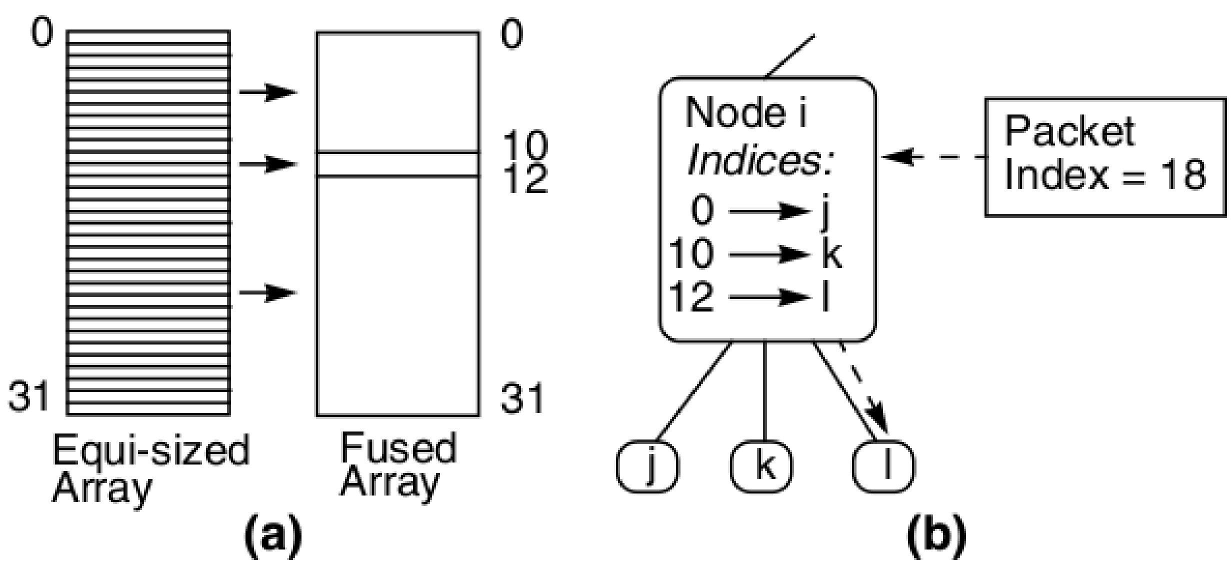


FIGURE 4. Equi-dense Cuts: (a) Fusion (b) Lookup

EffiCuts – Node co-location

- Dříve: Na základě hlavičky se vyčetl ukazatel, a tam se skočilo
- Nyní: S ukazatelem se rovnou vyčte hlavička, ta určí offset od ukazatele
  - Uzly jsou uloženy za sebou jako u TreeBitmap

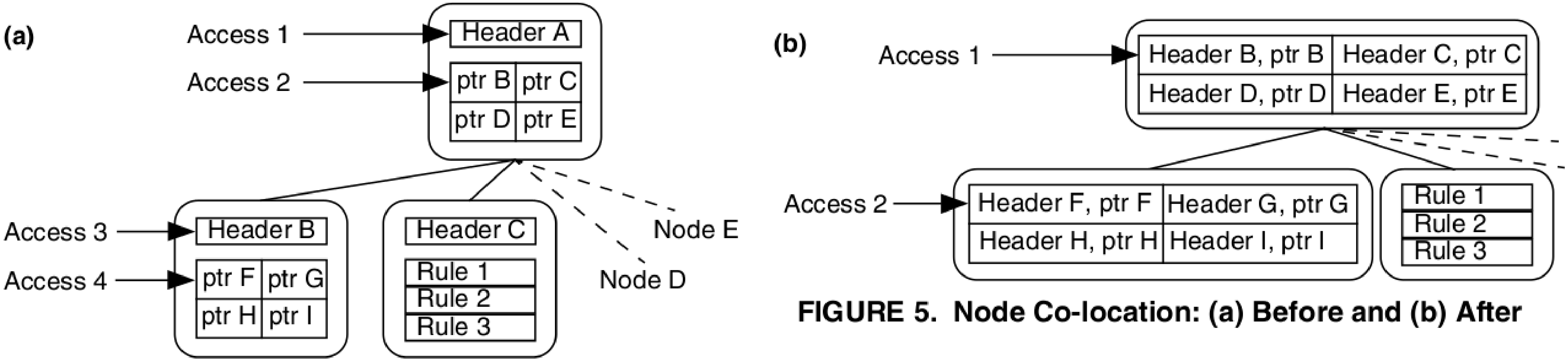
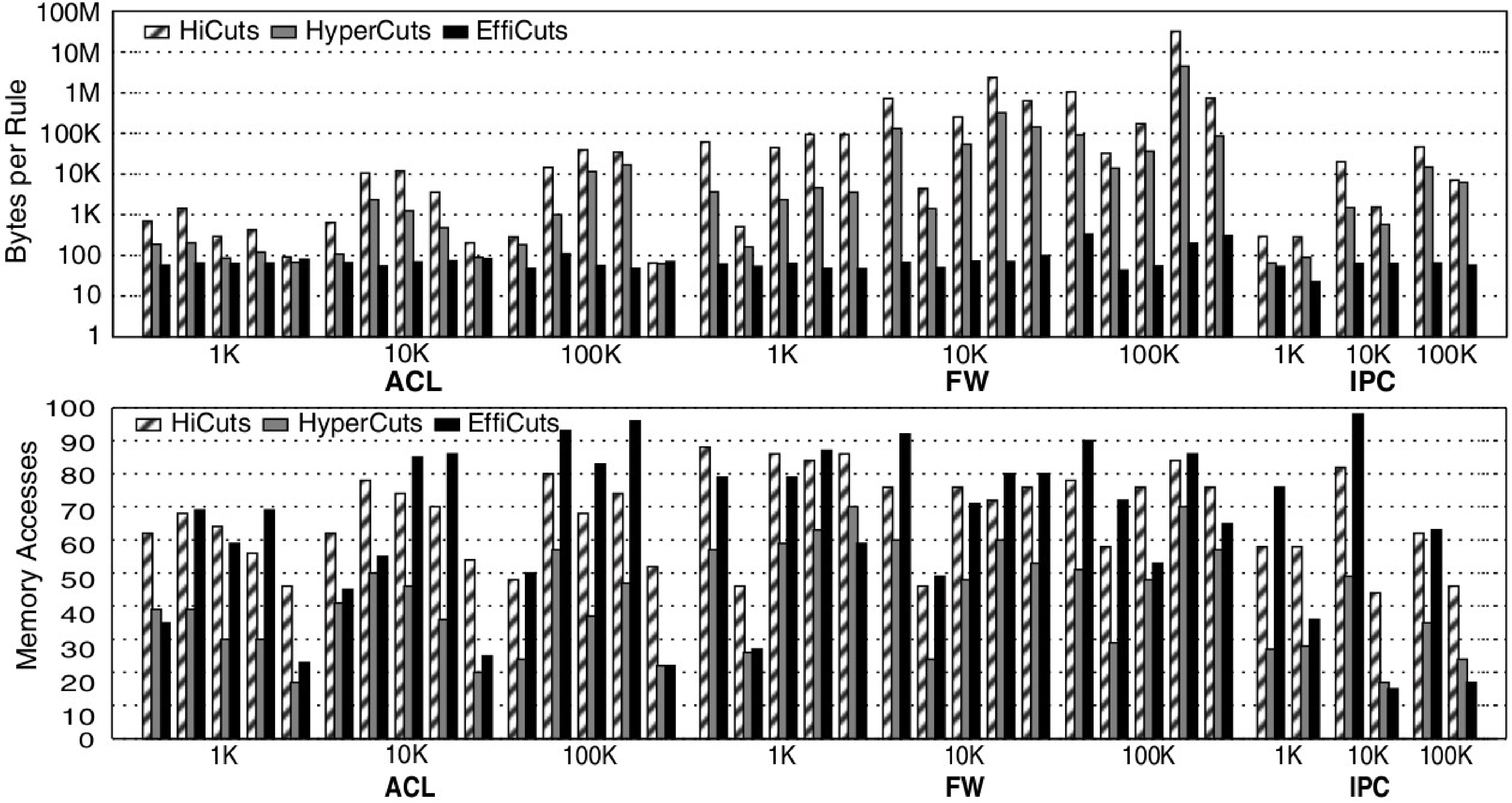


FIGURE 5. Node Co-location: (a) Before and (b) After

EffiCuts – výsledky 1



EffiCuts – výsledky 2

Table 3: Typical Power and Throughput Comparison

Classifier Type	TCAM			HyperCuts						EffiCuts					
	Cycle time (ns)	Energy (nJ)	Power (W)	# Memory accesses	Size (MB)	Cycle time (ns)	Energy (nJ)	# Copies	Power (W)	# Memory accesses	Size (MB)	Cycle time (ns)	Energy (nJ)	# Copies	Power (W)
ACL 100K	7.46	169.5	23	37	1084	0.18	5.6	1	31	83	5.3	0.18	0.51	2	6
FW 100K	7.46	169.5	23	48	2433	0.21	8.3	2	81	53	3.7	0.20	0.43	2	4
IPC 100K	7.46	169.5	23	24	575	0.17	4.3	1	26	17	5.5	0.18	0.51	1	3

Zajímavé reference

- L. D. Carli, Y. Pan, A. Kumar, C. Estan, and K. Sankaralingam. Flexible lookup modules for rapid deployment of new protocols in high-speed routers. In Proc. ACM SIGCOMM, Barcelona, Aug. 2009.
- OpenVZ: Server Virtualization Open Source Project. http://www.openvz.org.
- J. Turner, P. Crowley, J. DeHart, A. Freestone, B. Heller, F. Kuhns, S. Kumar, J. Lockwood, J. Lu, M. Wilson, et al. Supercharging PlanetLab: A High Performance, Multi-application, Overlay Network Platform. In Proc. ACM SIGCOMM, Kyoto, Japan, Aug. 2007.
- Receive-Side Scaling Enhancements in Windows Server 2008. http://www.microsoft.com/whdc/device/network/ndis_rss.mspx.
- M. Dobrescu, N. Egi, K. Argyraki, B.-G. Chun, K. Fall, G. Iannaccone, A. Knies, M. Manesh, and S. Ratnasamy. RouteBricks: exploiting parallelism to scale software routers. In SOSP, 2009.
- University of Oregon RouteViews project. http://www.routeviews.org/.
- P. Gupta, S. Lin, and N. McKeown. Routing lookups in hardware at memory access speeds. In IEEE INFOCOM, 1998.
- M. Waldvogel, G. Varghese, J. Turner, and B. Plattner. Scalable high speed IP routing lookups. In SIGCOMM, 1997.
- N. Muralimanohar, R. Balasubramonian, and N. P. Jouppi. CACTI 6.0: A Tool to Model Large Caches. Technical Report HPL-2009-85, HP Labs.
- E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The Click modular router. ACM TOCS, 2000.
- J. Hasan, S. Cadambi, V. Jakkula, and S. Chakradhar. Chisel: A Storage-efficient, Collision-free Hash-based Network Processing Architecture. In Proceedings of the 33rd Annual International Symposium on Computer Architecture, 2006.

Diskuze