

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, DESIGN AND MANUFACTURING KANCHEEPURAM CHENNAI - 600127

Synopsis Of

High Speed Packet Classification Architectures

A Thesis

To be submitted by

DHAYALAKUMAR M

For the award of the degree

Of

DOCTOR OF PHILOSOPHY

1 Abstract

This research enhances the performance of high-speed network devices in FPGA (Field Programmable Gate Array) and ASIC (Application Specific Integrated Circuit) domains by optimizing priority selection, prefix processing, and range processing hardware efficiency. The approach uses analysis to achieve scalability and efficiency in priority selection, enabling best-match and multi-match functions through strategic rule grouping. This method integrates into existing hardware pipelines due to its gate-level implementation.

A deterministic approach called Range Enhanced Reconfigurable Packet Classification Engine (RER PCE) also tailors FPGA configurations for rule sets. The proposed FPGA solution, combined with the modified priority selection, achieves a 4Mb TCAM (Ternary Content Addressable Memory) size and 400MPPS throughput on Virtex-7 FPGA.

The proposed ASIC solution, Ternary and Range (TeRa) based Packet Classification Engine, adds scalability and range processing capacity. It includes a unique subtraction based range processing hardware that minimizes rule expansion. With the modified priority selection, this solution reaches 6.6BPPS throughput for 104-bit rules.

The research introduces a sixteen state (Sedinary) encoding for prefix and range processing, achieving 11.9BPPS throughput using the proposed configurable logic block. These ASIC, FPGA, and configurable hardware solutions maintain scalability, offering a versatile toolkit for high-speed networking challenges.

2 Objectives

The research focuses on developing high-performance packet classification hardware that efficiently handles large rule databases and high packet flow rates in real-time networks. The goal is to remove the significant hurdle to scalability by creating hardware elements independent of the number of rules to be processed, the number of headers in a rule, the matching scheme of each header, priority selection between rules, and link speed (wired or wireless). These properties will explore hardware architectures and algorithms to provide efficient and scalable packet classification capabilities.

2.1 Research Problem

The research aims to overcome hardware limitations in packet classification, focusing on scalability, reconfigurability, high-speed processing, range handling, and priority selection. These limitations necessitates high-throughput hardware components that are not affected by the following:

• The number of rules that need to be processed and it's priority effect

- The number of headers contained in a rule
- The matching scheme for each header
- The number of packets processed per second or the link speed (wired or wireless)

The priority selection overheand is an critical one compare to all others since it needs a recursive function to compute the priority effect which is completely serial in nature.

2.2 Research Findings

Based on the literature and analysis results, the research will explore FPGA, ASIC, and memory based hardware solutions to address the major challenges of scalability and high throughput in packet classification as follows

- (a) **Best-Match and Multi-Match Selection:** The proposed method reorder the rules without changing the priority effect and grouping the rules into subgroup to reduce the priority selection overhead to obtain the best-match and multi-match results in the same hardware.
- (b) Modified Priority Order: Proposes a recursive solution for N to N, and N to log₂N priority encoder, with modified priority order to reduce the computational complexity of priority selection.
- (c) Resource Deterministic Approach for BRAM / LUT-RAM based Ternary Match: This approach enables determining the required number of LUT (Look Up Table), LUT-RAM, and BRAM(Block RAM) in FPGA based on the number of rules in specified ACL. The mechanism also finds the maximum possible TCAM availability in any given FPGA. Using the proposed approach, the Virtex-7 FPGA gives a maximum of 4Mb TCAM space.
- (d) Enhanced Range Match and Negation Prefix Logic: Processing converted prefixes in a single rule minimizes the need for rule expansion for range fields. Both prefixes and negated prefixes are represented by inversely encoding the match bit in LUT-RAM/BRAM of FPGA, allowing the exclusion headers representation without additional hardware resources.
- (e) **Resource Accountable Modified Scalable Priority Selection for FPGA:** A resource measurable approach is employed for scalable priority selection logic in FPGA, ensuring efficient resource utilization while maintaining scalability.
- (f) **Compressed Header Representation:** Compressed header field representation with respect to classbench rule header format.
- (g) **Optimal Prefix and Specialized Range Match Logic:** Optimal ternary processing logic for processing the prefix fields, and a specialised subtraction based hardware logic to process the range field without increasing the memory and hardware complexity.
- (h) **Sedinary State Encoding:** Sixteen-state encoding with a two-bit stride to represent the prefix leads to a 50% reduction in hardware without necessitating an increase in memory compared to the conventional approach. The hardware responsible for processing the sixteen-state encoding is referred to as Sed-CAM.
- (i) Configurable Range and Prefix Processing Hardware (Transfigure Architecture): A configurable block to process the 2*n*-bit prefix or *n*-bit range along with match inversion logic.

3 Fast Priority Selection Algorithm using Priority Reordering and Grouping

Priority selection serves as a critical mechanism that dictates the sequence in which classification rules are applied to incoming packets. This ordering ensures that higher-priority rules are evaluated before lower-priority ones, enabling efficient decision-making and appropriate handling of network traffic. Priority selection mitigates bottlenecks, tailors treatment for different applications, and boosts overall network throughput. It is a vital mechanism for efficient network operation and timely handling of critical data.

3.1 Priority Grouping

The research solution focuses on rearranging high-priority non-overlapping rules into a High-priority Best-match Rule (HBR) group. This recursive grouping process divides the rules into subgroups, and each subgroup contains nonoverlap rules with a single rule match condition for any incoming packet header. This approach ensures that rule significance is maintained and accurate classification is achieved.



Figure 1: Priority reordering

Though the rules R2 and R7 in Fig. 1 have lower priority in the given list, they are not losing their significance compared to other high-priority rules. As a result, these two rules moved to the HBR group along with rule R1. On the other hand, rules R3, R4, and R5 share overlap significance with only one high-priority rule and are categorized into the Mid-priority Multi-match Rule (MMR) group. Hence, the Low-priority Multimatch Rule (LMR) group gets the remaining rules R6, R8, and R9.

3.2 Possibility of Best-Match and Multi-Match

The multi-match requires representing all possible match addresses that need large memory for the number of rules. Only one rule gets a match in each level of the proposed grouping method. This compact representation reduces the number of bits required to represent each subgroup result and optimizes the memory for multi-match results.

The sublist (SL1) mentions the optimal level for the rules in the MMR group, where the HBR and LMR groups are represented in a single level. The sublist (SL2) divides

					with LMR		without LMR	
Data Set	# Rules (N)	HBR (H)	MMR (M)	LMR (L)	# bits for Multi- match	# of Sublist SL1	# bits for Multi- match	# of Sublist SL2
ACL1_10K	9603	7957	1615	31	89	8	64	16
ACL1_5K	4415	3560	830	25	92	12	76	24
IPC1_10K	9037	5042	3906	89	386	61	325	70
IPC1_5K	4460	2378	1951	131	323	34	234	54
FW1_10K	9311	2481	6792	38	179	20	158	30
FW1_5K	4653	1160	3459	34	71	24	153	34

Table 1: Classbench dataset and its priority effect

the MMR group into multiple sublists, each containing only NOL rules. This division allows each sublist to produce a single match result, but it comes at the cost of additional levels. The Table 1 explains multi-match bit needs with respect to SL1 and SL2. The cumulative value of each level's match address and the LMR match result provides the minimum number of match bits required for multi-match. The proposed method can efficiently represent the multi-match result with fewer bits without compromising the actual priority effect in multiple sublists. As a result, the best-match result is also available in the sublist with priority order.

The modified priority order preserves each rule's priority by changing the action's address with respect to the proposed modified order. The 8-bit input and its corresponding address of the action for conventional and proposed modified order is shown in Table 2 for better understanding. The input * * 011111 shows an address of 0101 (5) in the conventional method, but the proposed method requires an address of 001 (1). Hence the proposed modified priority encoder revises the address of each action based on the modified order to eliminate the additional logic needed to find the binary order.

Input	Conventional	Modified	Priority	
bits	Order	Order	Availability	
**** ***0	0 00	1 11	0	
**** **01	0 01	1 01	0	
**** *011	0 10	1 00	0	
**** 0111	0 11	1 10	0	
***0 1111	1 00	0 11	0	
**01 1111	1 01	0 01	0	
*011 1111	1 10	0 00	0	
0111 1111	1 11	0 10	0	
1111 1111	-	-	1	

Table 2: Conventional and Modified priority order

3.3 Proposed Priority Selection Methodology

The proposed algorithm adopts a decomposed approach without employing any folding structure. For an N-bit input, its priority determination is done by dividing it into N/K chunks, each chunk being K-bits in size. The priority of each K-bit chunk is determined simultaneously in parallel during Stage-1. In Stage-2, the priority of K^2 -bits is determined, and this process continues until the $log_k N^{th}$ stage. This approach allows for efficient and parallel processing of the input's priority at different stages. A priority encoder converts multiple input signals into a binary code representing the highest priority input. Priority selection will be categorized into two types as follows

- (a) **TYPE 1 N to N bit Priority Selection:** Priority selection with *n*-bit to *n*-bit refers to a process where all other bits are the inversion of high priority bit except for the high priority bit. This form of priority selection is beneficial in hardware applications where the address decoder of memory can be substituted.
- (b) **TYPE 2 N to** $\log_2 N$ bit Priority Selection: The priority encoder examines the *N*-bit inputs in order of priority and detects a high priority bit position as a log_2N -bit binary code. An application like packet classification uses this priority encoder to compute the memory address from the mult-match output. The address used to choose the required action from the memory.

3.4 Performance Analysys of Proposed Priority Selction

The proposed decomposition algorithm for the 4096-bit priority encoder is compared with existing literature in Table 3, revealing a significant improvement. The new approach achieves an impressive 90% reduction in Energy Per Operation (EPO) and a minimum of 70% efficiency improvement in the Area Delay Product (ADP) for both TYPE 1 and TYPE 2 design. This indicates that the proposed algorithm is highly efficient and advantageous over existing methods, making it a promising solution for priority encoder designs in large-scale applications.

Design	Delay	Area	Power	PDP	ADP
	(nSec)	μm^2	$\mu \mathbf{W}$	(fJ)	
Mux based Bi et al. (2005b)	2.7914	87	1.546	4.0642	242.8
Hybrid	1.1643	93	2.321	2.7032	108.2
Veeramachaneni et al. (2007)					
OR based	0.6898	102	2.213	1.5265	70.35
ChetanKumar et al. (2011)					
Conventional	3.1276	118	1.936	6.055	369.05
Proposed TYPE 1 PE4	0.2575	124	2.311	0.595	31.93
Proposed TYPE 2 PE4	0.2526	197	2.314	0.5854	49.76

Table 3: Performance	comparison	of 32-bit Decom	posed PE using	; 4-bit PE
	1		1 U	/

4 FPGA based Range Enhanced Reconfigurable Packet Classification Engine (RER PCE)

The solution introduces a new deterministic approach for designing a Range Enhanced Reconfigurable Packet Classification Engine (RER PCE) tailored for FPGAs. The proposed framework utilizes the RAM-based (LUT-RAM / BRAM) "Ternary Match" technique to handle prefix and range prefix representations efficiently. Additionally, employs rule-reordering for priority selection, enabling simultaneous best-match and multi-match capabilities within the same architecture. Reconfigurable hardware is a promising technology for implementing firewalls, routing mechanisms, and new protocols in high-performance network systems, and the approach aims to leverage its potential effectively.

4.1 Deterministic Properties of FPGA Resource Utilization

The stage-wise analysis gives the generalized function to compute the required FPGA resources for different rule dimensions. The five-field classification with the 104-bit rule can give three rule dimensions: 1) Prefix Rule, 2) 1D-Rule, and 3) 2D-Rule. The proposed solution uses 22 LUT-RAM cells (11 cells per prefix rule) to represent two prefix rules or 13 BRAM (0.18 per prefix rule) to represent 72 prefix rules. Similarly, the 1D-Rule needs 82 LUT-RAM cells (41 cells per 1D-Rule) for two 1D-Rules or 43 BRAM (0.59 per 1D-Rule) to represent 72 1D-Rules. BRAM-based 1D-Rule gives less number of match lines compared to LUT-RAM-based 1D-Rule. Hence the BRAM has opted for the representation of the 1D-Rules than the LUT-RAM-based 1D-Rules. The Virtex-7 (XC7V2000T) FPGA consists of 1292 BRAM cells can able to represent 72 * |1292/43| = 2163 1D-Rules or 72 * |1292/13| = 7155 prefix rules. Similarly 344800 **LUT-RAM** can able to process 2 * |344800/82| = 8408 1D-Rules or 2 * |344800/22|= 31344 prefix rules. This implementation leads to a maximum of 10.5K 1D-Rules or 38.5K prefix rules in the Virtex-7 FPGA for the five-field classification. The FPGA resource requirement for a particular dataset can be determined based on the number of fields, the number of bits in each field, and the number of range fields in a rule.

The Scalable TCAM given by Jiang (2013) claims 2.4 Mbits of TCAM with a throughput of 150 MPPS for the 150-bit rule size in Virtex-7 FPGA, but it failed to maintain the TCAM size with a varying number of rules size. None of the existing techniques in the literature determines the FPGA resource utilization for the complete PCA to implement particular data set. The suggested framework shows the deterministic hardware utilization in each stage to find the required FPGA version of the specific data set. Since the complete architecutre is pipelined with respect to each SLICE with single level of LUT-RAM or BRAM, the throughput of the proposed design meets with basic element speed.

The proposed work can give 400 MPPS throughput with 4 Mbits of TCAM to accommodate 38.5 K prefix rules in the Virtex-7 FPGA for BRAM, which gives 166.6% more throughput and 66.6% more TCAM space compared to scalable TCAM proposed by Jiang (2013). LUT-RAM-based TCAM framework shows the throughput of 520 MPPS with 3.2 Mbits TCAM space to accommodate any size of rules.

5 ASIC solution for Ternary and Range (TeRa) based Packet Classification Engine

The research introduces two distinct ASIC architectures with a modified priority selection. The TYPE 1 design is suitable for ASIC-based architecture where a customizable address decoder of the memory is applicable. The TYPE 2 design methodology supports both ASIC designs and FPGA implementations since it uses a conventional memory for action lookup. This modified priority selection significantly contributes to achieving a throughput of 6.6 BPPS for TYPE 2 and 9.9 BPPS for TYPE 1 based packet classification architectures. Furthermore, the research presents an innovative encoding method that reduces memory usage by 45% in each field, along with an update logic to facilitate this encoding approach. Integrating a carry-based range comparator enhances hardware optimization for range processing without the need for prefix conversion. Additionally, the match inversion logic streamlines the processing of exceptional or inverse fields without incurring extra hardware overhead.

5.1 Specialized Range Match Circuit

The range should follow the mathematical comparison $L \le S \le U$, where L represents the lower range, U stands for the upper range, and S denotes the search key. Subtraction between the values shows a positive outcome when a smaller value from a larger one. Conversely, subtracting a larger value from a smaller one yields a negative result, while subtracting two equal values results in zero. In digital systems, subtraction is executed using techniques like 1's complement and 2's complement. These methods demonstrate an advantageous property when assessing the "less than or equal" condition based on the carry bit, which is as follows:

(a) Comparison using 2's complement

The subtraction of X - Y using 2's complement creates a carry (1) if the result is positive. The carry indicates X is greater than or equal to Y. Since 2's complement does not have the negative zero, it produces carry for equal values of X and Y during subtraction.

(b) Comparison using 1's complement

The subtraction of X - Y using 1's complement creates a carry (1) if the result is positive except zero. The carry indicates X is greater than Y. Since 1's complement has independent representation for positive and negative zero, it will not produce carry for equal values of X and Y during subtraction.

5.2 Results and Discussions

The suggested prefix-based PCE architecture showcases an impressive throughput of 9.9 Billion Packets Per Second (BPPS) for the packet classification architecture with TYPE 1 based priority encoder design. In contrast, the proposed architecture with TYPE 2-based priority selection achieve throughputs of 6.6 BPPS. In terms of power efficiency, the proposed design outperforms the optimal work previously presented by Murugesan and Sk (2017), exhibiting almost eight times reduction in Area-Delay Product (ADP) and nine times better for Power-Delay Product (PDP).

6 Sedinary CAM Based Transfigure Architecture

The research introduces the transfigure architecture, a novel approach for efficient and scalable packet classification. This architecture incorporates a sixteen-state Content Addressable Memory (CAM) and a programmable logic block to handle both prefix and range processing. One key advantage of this architecture is its ability to process ranges directly without converting them into prefixes, resulting in reduced hardware complexity. The proposed architecture optimizes the priority grouping and reordering method, effectively eliminating the need for priority selection overhead. This ensures efficient rule processing and significantly improves the overall performance. Furthermore, the architecture can seamlessly scale rule size or field size without introducing delays with a well-designed pipelining stage.

6.1 Results and Discussions

The comparison of the transfigure architecture for five-field classifications with both existing literature and the proposed ASIC solution is displayed in Table 4. The table demonstrates that the suggested configurable solution achieves a significant reduction of at least eleven times and seven times in terms of ADP (Area-Delay Product) and PDP (Power-Delay Product), respectively, compared to the proposed TYPE 2 and TYPE 1 based ASIC solution. Furthermore, this architecture attains a notable throughput of 11.9 BPPS (Billion Packets Per Second) while maintaining scalability to accommodate varying sizes of rules and fields.

Technique	Delay	Area	Power	# bits	ADP	PDP
	(pSec)	$(\mathbf{mm^2})$	(\mathbf{W})	per Rule	$ imes 10^{-12}$	(pJ)
PPRC						
Chang and Lu (2013)	1233	1893.6	67.61	208	2334.8	83360.7
RMA						
Murugesan and Sk (2017)	401	448.08	1.1	208	179.7	423.456
Proposed (Type 1)	101	149.11	0.21	149	15.9	21.1
Proposed (TYPE 2)	151	211.2	0.31	149	31.9	46.81
Proposed (Transfigure)	84	16.73	0.033	149	1.41	2.77

Table 4: Performance comparision of 104 bit rule in 45nm TSMC node

In summary, the transfigure architecture demonstrates a revolutionary approach to packet classification, providing enhanced performance, scalability, and efficiency compared to existing solutions. This design streamlines range processing and optimize rule management, resulting in a cutting-edge solution that meets the demands of modern packet processing applications.

7 Conclusion

This thesis presents a comprehensive analysis that proposes scalable hardware solutions catering to various stages of packet classification within ASIC and FPGA-based applications. The research analysis suggests that the priority selection overhead in packet

classification can be minimized using the two proposed recursive types of priority selection logic: (1) N to N Priority Encoder of Type 1 and, (2) N to log₂N Priority Encoder of Type 2. The proposed FPGA-based methodology, the Range Enhanced Reconfigurable (RER) Packet Classification Engine (PCE), demonstrates a deterministic approach with a throughput of 520 MPPS. The ASIC-based Ternary and Range (TeRA) PCE introduces a novel subtraction-based range processing circuit, achieving an impressive throughput of 9.9 BPPS. The proposed novel Sedinary (Sixteen State) CAM-based Configurable Transfigure Architecture showcases an exceptional throughput of 11.9 BPPS. This innovation highlights the capacity to efficiently tailor hardware mechanisms to process both prefix and range fields.

In essence, this thesis bridges the gap between analysis and practical implementation, culminating in a diverse range of hardware solutions that effectively address the challenges of packet classification within the realms of ASIC and FPGA-based applications. These contributions collectively enrich the high speed network processing landscape, offering enhanced efficiency, scalability, and innovation across various networking scenarios.

8 Organization of the Thesis

The thesis organized into 7 chapters as follows

- Chapter 1: Introduction to Packet Classification
- Chapter 2: Scholarly Investigation on Literature
- Chapter 3: Effect of Modified Priority Selection in PCA
- Chapter 4: Proposed FPGA-based Hardware: RER-PCE
- Chapter 5: Proposed ASIC-based Hardware: TeRa-PCE
- Chapter 6: Proposed Configurable Sedinary CAM: Transfigure Architecture
- Chapter 7: Conclusion and Future Work

9 List of Publications

- Dhayalakumar M and Noor Mahammad Sk. 2023. Deterministic Approach for Range-enhanced Reconfigurable Packet Classification Engine. ACM Trans. Reconfigurable Technol. Syst. 16, 2, Article 29 (June 2023), 26 pages. https://doi.org/10.1145/3586577
- Skandha Deepsita S, Dhayala Kumar M, and Noor Mahammad SK. 2021. Energy Efficient Error Resilient Multiplier Using Low-power Compressors. ACM Trans. Des. Autom. Electron. Syst. 27, 3, Article 21 (May 2022), 26 pages. https://doi.org/10.1145/3488837
- S. Murugesan, M. Dhayalakumar and S. Noor Mahammad, "A Novel High Performance Universal Measurement Logic Element," 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, 2020, pp. 1-6, doi: 10.1109/CICT51604.2020.9312059.

10 Chip Tapeouts

Successfully taped out three Caravel SoC with a custom design using Skywater 130nm technology library in Efabless MPW3 Shuttle in 2022. The IC lists are as follows.

- (a) High Speed Signed Adder
- (b) Fixed to Single Precision Floating Point Converter
- (c) Approximate Multiplier

References

- 1. Abdel-Hafeez, S. and S. Harb (2006). A vlsi high-performance priority encoder using standard cmos library. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **53**(8), 597–601, doi:10.1109/TCSII.2006.876412.
- Agrawal, B. and T. Sherwood (2008). Ternary cam power and delay model: Extensions and uses. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 16(5), 554–564, doi:10.1109/TVLSI.2008.917538.
- Baboescu, F. and G. Varghese (2001). Scalable packet classification. In Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '01. Association for Computing Machinery, New York, NY, USA. ISBN 1581134118, doi:10.1145/383059.383075.
- Bi, S., W. Wang, and A. Al-Khalili (2005a). Multiplexer-based binary incrementer/decrementers. volume 2005. ISBN 0-7803-8934-4, doi:10.1109/NEWCAS.2005.1496662.
- Bi, S., W. Wang, and A. J. Al-Khalili (2005b). Multiplexer-based binary incrementer/decrementers. *The* 3rd International IEEE-NEWCAS Conference, 2005., 219–222.
- 6. Bremler-Barr, A. and D. Hendler (2012). Space-efficient tcam-based classification using gray coding. *Computers, IEEE Transactions on*, **61**(1), 18–30.
- Chang, Y.-J. and H.-Y. Lu (2013). Improving the performance of port range check for network packet filtering. ACM Trans. Des. Autom. Electron. Syst., 19(1). ISSN 1084-4309, doi:10.1145/2523069.
- 8. Cheng, Y.-C. and P.-C. Wang (2016*a*). Scalable multi-match packet classification using tcam and sram. *IEEE Transactions on Computers*, **65**(7), 2257–2269, doi:10.1109/TC.2015.2470242.
- 9. Cheng, Y.-C. and P.-C. Wang (2016b). Scalable multi-match packet classification using tcam and sram. *IEEE Transactions on Computers*, 65(7), 2257–2269, doi:10.1109/TC.2015.2470242.
- ChetanKumar, V., P. S. Phaneendra, S. E. Ahmed, S. Veeramachaneni, N. M. Muthukrishnan, and M. B. Srinivas (2011). A reconfigurable inc/dec/2's complement/priority encoder circuit with improved decision block. 2011 International Symposium on Electronic System Design, 100–105.
- 11. Delgado-Frias, J. and J. Nyathi (2000). A high-performance encoder with priority lookahead. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, **47**(9), 1390–1393, doi:10.1109/81.883335.
- 12. Fang Yu and R. H. Katz (2004). Efficient multi-match packet classification with tcam. In Proceedings. 12th Annual IEEE Symposium on High Performance Interconnects. doi:10.1109/CONECT.2004.1375197.
- 13. Gupta, P. and N. McKeown (1999). Packet classification on multiple fields. *ACM SIGCOMM Computer Communication Review*, **29**(4), 147–160.