# CompactDFA: Scalable Pattern Matching Using Longest Prefix Match Solutions

## Abstract:

A central component in all contemporary intrusion detection systems (IDSs) is their pattern matching algorithms, which are often based on constructing and traversing a deterministic finite automaton (DFA) that represents the patterns. While this approach ensures deterministic time guarantees, modern IDSs need to deal with hundreds of patterns, thus requiring to store very large DFAs, which usually do not fit in fast memory. This results in a major bottleneck on the throughput of the IDS, as well as its power consumption and cost. We propose a novel method to compress DFAs by observing that the name used by common DFA encoding is meaningless. While regular DFAs store separately each transition between two states, we use this degree of freedom and encode states in such a way that all transitions to a specific state are represented by a single prefix that defines a set of current states. Our technique applies to a large class of automata, which can be categorized by simple properties. Then, the problem of pattern matching is reduced to the well-studied problem of Longest Prefix Match (LPM), which can be solved either in ternary content-addressable memory (TCAM), in commercially available IP-lookup chips, or in software. Specifically, we show that with a TCAM our scheme can reach a throughput of 10 Gb/s with low power consumption.