MNCaRT collects a diverse set of automata processing tools of tools for creating, manipulating, and executing automata. Instead, a general and extensible framework is needed to enable the development of platform-independent applications and to support experimental automata. Designs. Instead, a general and extensible framework is needed to enable the development of platform-independent applications and to support experimental automata paradigms. The tools allow researchers to easily evaluate designs across hardware platforms. Tool support includes manipulation of automata, execution of complex machines, high-speed processing of NFAs and DFAs, and compilation of regular expressions. We provide engines to execute automata on CPUs (with VASim and Intel Hyperscan), GPUs (with custom DFA and NFA engines), and FPGAs (with an HDL translator). We also introduce MNRL, an open-source, general-purpose and extensible state machine representation language developed to support MNCaRT. The representation is flexible enough to support traditional finite automata (NFAs, DFAs) while also supporting more complex machines, such as those which propagate multi-bit signals between processing elements. We hope that our ecosystem and representation language stimulates new efforts to develop efficient and specialized automata processing applications.

I. INTRODUCTION

Years of research and development have resulted in high-throughput automata processing architectures and software engines [1]–[3]. This has lead to the discovery of non-obvious use-cases and application domains for finite automata, such as natural language processing [4], network security [5], graph analytics [6], high-energy physics [7], bioinformatics [8]–[10], pseudo-random number generation and simulation [11], data-mining [12], [13], and machine learning [14].

Unfortunately, the software frameworks for the construction, manipulation, and translation of automata are frustratingly fractured (e.g. have inconsistent serialization formats) and restrictively licensed (e.g., Micron licenses a comprehensive SDK, but it is closed-source and specifically targets their D480 Automata Processor, or AP [2]). While these tools are useful for developing applications for the AP, the tools do not allow researchers to easily evaluate designs across hardware platforms, such as CPUs, GPUs, and FPGAs. The tools also cannot be easily extended to support new architectures and automata paradigms. Instead, a general and extensible framework is needed to enable the development of platform-independent applications and to support experimental automata designs.

We present MNCaRT (the MNRL Network Computation and Research Testbed, pronounced “mincaart”) [15], a suite of tools for creating, manipulating, and executing automata. MNCaRT collects a diverse set of automata processing tools and algorithms into a central location and will grow as new tools are developed. We currently provide support for compiling state machines from Perl compatible regular expressions (PCRE) [16] to MNRL, high-speed execution of NFAs and DFAs using Intel Hyperscan [3], and optimization and simulation of experimental automata designs with the Virtual Automata Simulator (VASim) [17]. Further, we provide back-ends for executing DFAs on GPUs (DFAGE) [18], FPGAs [19], and exploring routing constraints for experimental spatial architectures via the Automata-to-Routing (ATR) tool [20]. All tools in MNCaRT are publicly available (typically under BSD licenses), and available pre-installed in a Linux container, allowing both academics and industry experts to contribute to, and use, the ecosystem.

To support our ecosystem, we have created MNRL, the MNRL Network Representation Language (pronounced “mineral”), a JSON-based, open-source language to support the development of, and experimentation with, new automata-based applications and architectures. MNRL allows a user to define a network (or collection) of MNRL nodes, which represent the states within automata. Each node stores configuration information (such as node type, name, etc.) and connections to other nodes within the network. The language specification is general, allowing state machines other than finite automata to be represented. We provide initial definitions for traditional finite automata states, homogeneous states, up-counters, and Boolean logic in the MNRL specification; additional node types may be defined by the user for specific applications.

To summarize, this works presents the following:

- MNCaRT, an comprehensive repository of compatible tools for development, visualization, and analysis of automata processing on CPUs, GPUs, and FPGAs.
- MNRL, an extensible, open-source JSON specification for representing state machines.
- Extensions to Intel’s Hyperscan PCRE engine, supporting compilation to and execution of MNRL files.
- Updated versions of VASim, REAPR, DFAGE, and iNFAnt2, which support reading and writing of MNRL files.

II. BACKGROUND AND RELATED WORK

A finite automaton includes of a set of states and a set of transitions defining how the states become active based
on symbols observed in an input stream \([21]\). In a non-deterministic finite automaton (NFA), it is possible to transition to multiple states on the same input symbol. Automata are often represented as a graph, defining the topological layout of the computation. Computation is therefore decoupled from the definition of the state machine, allowing for a common execution engine to process arbitrary automata, improving code reuse and reducing sources for bugs. Automata can also be represented as a set of regular expressions, which define the search pattern the automata recognize.

In the remainder of this section, we briefly highlight some existing automata processing engines and discuss limitations of current automata representation languages.

**Automata Processing Engines.** Micron’s D480 AP \([2]\) is a custom hardware accelerator which directly executes homogeneous finite automata. Becchi et al. have developed a set of tools and algorithms for efficient CPU-based automata processing \([22]\), and other CPU engines include Intel’s Hyper- scan \([3]\) and Google’s RE2 \([23]\). Automata processing engines have also been developed for GPUs and FPGAs (e.g., \([24]\), \([25]\), and \([26]\)). Unfortunately, existing engines do not share a common automata representation, making cross-architecture comparison and development of automata-based algorithms challenging and time-consuming.

**Limitations of Automata Representation Languages.** The Automata Network Markup Language (ANML) is a proprietary description language developed for the Micron D480 AP. Licensing restrictions make the language challenging to use for prototyping new automata elements, and additional annotations cannot be added to elements in ANML while maintaining support for current tools. Therefore it is not a good choice for unifying automata processing engines.

Becchi et al.’s tools use a simple NFA representation based on the theoretic definition of NFAs and cannot be easily extended to support more complex state machines. The language is custom, and there is no support in general-purpose programming languages for reading and manipulating these files.

Regular expressions are commonly used to generate automata, but are difficult to develop and maintain. Many applications (e.g., particle tracking, motif searchers, and rule mining) would be represented by non-intuitive regular expressions that are often exhaustive enumerations of all possible matches. Additionally, programming of regular expressions can be extremely error-prone due to variations in regular expression syntax, which leads to high rates of runtime exceptions \([27]\).

While other automata representation languages exist (e.g., Dot and JFLAP), these present similar licensing, generalizability, and maintainability challenges.

### III. MNRL: A JSON-BASED AUTOMATA LANGUAGE

We have developed MNRL, an extensible, open-source automata representation language, which allows for the topological specification of a collection of finite state machines using JSON syntax. While JSON is supported by most common general-purpose programming languages, we provide C++ and Python bindings to support additional validation checks.

It is important to note that the MNRL format specifies the layout of a machine but does not specify how elements behave, allowing many types of state machines to be represented, including traditional NFAs \([21]\) and homogeneous NFAs \([28]\). \(^3\) Behavior is left for the execution engine to specify and implement (allowing MNRL to be an extremely flexible file format). Therefore, MNRL is similar in intent to the Unified Modeling Language (UML), in which developers describe and design software systems while eliding implementation details \([29]\).

#### A. MNRL Format

A MNRL file contains a single MNRL network—a collection of one or more state machines that are executed in parallel using the same input. The file contains an array of MNRL nodes, which define each element in the network. An example node is given in Figure 1. A node consists of:

- A unique identifier
- A node type (state, homogeneous state, up counter, boolean, etc.)
- How the node is enabled (e.g. on the start input processing or when an incoming edge is active)

\(^3\)MNRL is general enough to represent more powerful machines (e.g. push-down automata, cellular automata, and Turing machines).
• Whether the node reports (generates an output signal) when activated
• An array of input ports, each with a unique ID and specified width (number of wires)
• An array of output ports, each with a unique ID, specified width, and list of connected nodes
• Custom attributes, specific to each element type

A developer can encode the topological layout of the state machines within the network and to specify the sort of behavior the underlying execution engine should assign to each node. The implementation of behavior is not defined in the MNRL file; instead, the computation engine that processes a MNRL network is responsible for specifying the semantics for each node type. Therefore, node types and execution engines are typically co-designed. If an engine needs information (e.g., symbol sets for matching against an input stream) to process a node, this configuration can be embedded in a MNRL node’s attributes. For the standard node types, we have specified additional attributes to support their respective expected behaviors.

B. Extending the MNRL Schema

MNRL is designed to be extensible, enabling research on new, custom automata functionality and allows researchers to quickly define custom attributes for new node types. We provide the specification of MNRL as a JSON schema, which allows for validation of file syntax. Because custom node types become part of the JSON schema, prototype extensions to the MNRL format can still be statically checked with minimal effort from the developer. The MNRL file format could easily be extended to support additional node types such as non-deterministic counters and stacks (to support push-down automata). Because MNRL supports variable-width ports, it is also possible to represent elements that share more than a single bit of data with elements downstream.

IV. THE MNCaRT ECOSYSTEM

Our goal with this work is to enable the development of a rich, vibrant ecosystem of compatible tools for manipulating and executing automata. We are collecting these tools in an umbrella repository, the MNRRL Network Computation and Research Testbed (or MNCaRT). By keeping tools catalogued in a single location, we hope to maintain the interoperability of tools and reduce fracturing in the ecosystem. We also provide a Linux container configured to use all of the MNCaRT tools.

Figure 2 describes the interaction between tools provided with MNCaRT. Our ecosystem supports workflows beginning with high-level languages, such as PCRE, and ending with execution on CPUs, GPUs, and FPGAs. We also support execution on Micron’s Automata Processor via conversion to Micron’s Automata Network Markup Language (ANML). Additionally, we provide compatible benchmarks for testing and experimenting with tools in MNCaRT. In this section, we briefly describe the tools that make up the initial release of MNCaRT.

A. High-Level Languages

Regular Expressions. Our framework supports programming models that represent pattern searches at higher levels of abstraction. We compile PCRE to MNRL files using Intel Hyperscan’s parsing and compilation routines. Hyperscan is an open-source, industry-standard regular expression processing library supported by Intel. The tool returns a graph representation of the compiled state machine, which we traverse to generate a MNRL file.

RAPID. RAPID is a high-level programming language for execution of sequential pattern-matching applications. This C-like language is extended with three keywords to support parallel matching of patterns against a single data stream as well as sliding window pattern recognition. We have extended the RAPID compiler to emit MNRL files, allowing for high-level programming within the MNCaRT ecosystem.

B. Benchmarks

ANMLZoo. The ANMLZoo benchmark suite contains a diverse set of automata applications and associated input stimuli. Applications range from configurable, synthetic benchmarks to algorithms that are not easily represented by regular expressions and can therefore have vastly different execution characteristics. We have generated MNRL representations for all benchmarks in the suite.

---

*For additional details, please see Angstadt et al. [30].
https://hub.docker.com/r/kevinaangstadt/mncart
C. State Machine Representations

MNRL. Our JSON-based automata representation language provides the glue for MNCaRT ecosystem. This allows the various tools to function seamlessly with each other. Further details regarding MNRL may be found in Section III.

ANML. MNCaRT also supports Micron’s Automata Network Markup Language (ANML) and therefore allows performance comparisons with the D480 AP. We provide this support via a translation to MNRL. This translation is supported for MNRL node types provided with our default distribution of the language. Because element types in ANML are dictated by the design of the AP, our translator does not support custom node types.

D. Analysis, Transformation, and Compilation

Hyperscan Compilation. We provide an extension to Hyperscan (hscompile) that parses MNRL files and compiles finite automata to a serialized Hyperscan pattern database, allowing offline compilation. Hyperscan is an industry standard automata processing toolchain that offers a state-of-the-art, high-performance CPU automata processing engine.

VASim. We have extended VASim to support parsing of MNRL files. VASim is a general-purpose framework for automata simulation, optimization, transformation, and performance modeling. The tool enables prototyping, debugging, simulation, and analysis of automata-based applications and architectures. Additionally, VASim can parse Micron ANML files, allowing for conversion with MNRL.

Automata Lab. Automata Lab is a web-based graphical environment for visualizing, editing, and simulating finite automata. The tool uses VASim to manipulate automata, and the resulting state machines are displayed graphically, allowing for user interaction. Users may upload MNRL files or choose from applications in the ANMLZoo benchmark suite.

REAPR. REAPR is a design automation tool for generating highly-efficient FPGA-based automata kernels. REAPR generates RTL representations of homogeneous finite automata, a class of automata also implemented by Micron’s AP. Additionally, REAPR generates a reporting architecture, which allow for pattern matches discovered by the executing automata to be communicated back to the host system.

Automata-to-Routing. ATR utilizes the Versatile Place and Route (VPR) tool to model spatial automata processing architectures. We use VASim to emit VPR-readable circuits of MNRL networks and provide guidance to construct custom, parameterizable, spatial architecture description files to accept these custom state machine circuits. ATR is thus capable of modeling spatial architectures that are purpose-built to accept MNRL state machines.

E. Execution Engines

Hyperscan CPU Engine. We provide a tool (hsrun) for processing MNRL files against an input stream using the Hyperscan execution core. This tool deserializes the Hyperscan pattern database and node mapping produced by hscompile. The tool then scans the given input file against the database and prints out human-readable reporting information (e.g. MNRL ID and input stream offset). If multiple compiled MNRL files and/or input files are passed to hsrun, the tool will execute all pairings of the files using a supplied number of threads.

VASim CPU Engine. In addition to support for transformation and analysis of finite automata, VASim supports simulation of a diverse set of finite automata models. While Hyperscan achieves higher throughput, VASim’s modular design allows for quick prototyping to test new automata elements and designs, such as those including custom compute units.

FPGA Engine. In addition to generating hardware NFA kernels, REAPR can also generate a full platform execution environment for certain automata applications. The REAPR platform has been demonstrated to offer up to $183\times$ speedup over best-effort CPU implementations.

GPU Engines (DFAGE and iNFAnt2). MNCaRT contains both a GPU-based DFA engine (DFAGE) and NFA engine (iNFAnt2). The NFA engine was described previously by Wadden et al.; we therefore briefly describe DFAGE. Use of DFAGE first requires automata to be transformed into one or more DFAs using VASim. Note that this process is performed offline by the CPU. Each DFA consists of a state transition table and an acceptance vector, both of which are stored in the GPU’s global memory. A transition table is represented by a 2-D array containing the next state identifiers for every pair of current state identifier and input symbol. Similar to previous implementations, our DFA matching engine supports multipacket processing to take advantage of the extreme parallelism of GPU architectures.

V. Conclusions

We present MNCaRT, a suite of tools for analyzing, executing, and transforming automata networks. We support execution of MNRL networks on CPUs, GPUs, and FPGAs, and we provide a workflow for execution on Micron’s AP. Support for high-level pattern-matching languages, such as PCRE and RAPID is also provided as part of MNCaRT. Finally, we allow for design space exploration through analysis functionality in the VASim and ATR tools. We hope that this suite of tools will enable innovation and in this emerging, and important application domain.

Acknowledgment

This work was supported in part by grants from the NSF (CCF-1116673, CCF-1629450, CCF-1619123, CNS-1619098), AFRL (FA8750-15-2-0075), Jefferson Scholars Foundation, Achievement Rewards for College Scientists.
REFERENCES


