wasi-pattern-match

Motivation

- Server-side Wasm functions scan network packets and HTTP requests/responses to detect abnormal traffic; they must return results quickly
- Fully utilize hardware features (new instructions, GPU offload, potential hardware accelerators), which are not available in Wasm runtimes.

Direction

- Goals
  - Support PCRE syntax (or subset)
  - API: iterate over matches, find the first match, check existence
  - Improve upon current WebAssembly performance
- Non-goals
  - All of the PCRE API; e.g., string replacement

PCRE = Perl Compatible Regular Expressions

APIs

Scanner:

```
create_scanner: function(list<pattern>) -> expected<scanner, error>

close_scanner: function(scanner) -> error
```

APIs

Scan a single block:

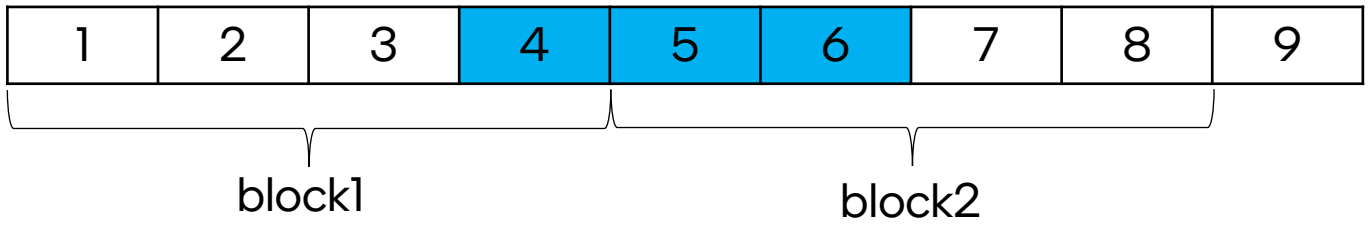
```
scan_block: function(scanner, buffer) -> expected<scan, error>
```

Scan consecutive blocks:

```
create_stream: function(scanner) -> expected<stream, error>

scan_stream: function(stream, buffer) -> expected<scan, error>

close_stream: function(stream) -> error
```



APIs

Retrieve results:

```
has_match: function(scan) -> bool

all_matches: function(scan) -> list<match>

first_match: function(scan) -> expected<match, error>
```

Code example

```
Scanner scanner = create_scanner([“dangerous url", "credit card", "phone number"]);
Scan scan = scan_block(scanner, buffer);
if (has_match(scan)) {
  // Detected dangerous traffic, drop the connection or return an error code.
} else {
  // Continue.
}
close_scanner(scanner);
```

Performance (TODO)

References

- Hyperscan: a high-performance multiple regex matching library
- RegEx Acceleration - NVIDIA Networking Docs
- Accelerating Regular-Expression Matching on FPGAs with High-Level Synthesis