Vegeta HTTP load testing

CONCEPTS

Vegeta sends a constant request rate, regardless of how slow the server is. Latency grows instead of throughput dropping.

Targets stdin or file
↓ attack
Results binary stream
↓ report / plot / encode
Output text, JSON, HTML, CSV

Results are a binary stream: write it to a file with > or tee, since the report cannot be regenerated otherwise.

ATTACK

vegeta attack
Fire requests from stdin targets

-rate=*<n>*/*<unit>*
Requests per time unit (default 50/1s)

-duration=*<t>*
How long to attack (0 = forever)

-targets=*<file>*
Read targets from a file, not stdin

-output=*<file>*
Write binary results to a file

-rate=0 -max-workers=*<n>*
Go as fast as possible

-header="*<Name: value>*"
Add a request header (repeatable)

-body=*<file>*
Body file for every request

-timeout=*<t>*
Per-request timeout (default 30s)

-workers=*<n>*
Initial workers (default 10, grows)

-max-workers=*<n>*
Cap the worker pool

-insecure
Ignore invalid TLS certificates

ATTACK (cont.)

-format=json
Targets are JSON lines

TARGETS

HTTP targets file

```
# one request per block, blank line
between
GET http://localhost:8080/items?x=1
X-Account-ID: 99

POST http://localhost:8080/things
Content-Type: application/json
@./newthing.json
```

In the HTTP format a body must be a file reference with @path; inline bodies are not supported. JSON targets take base64 bodies.

REPORT

vegeta report *<file>*.bin
Text summary of an attack

cat *<file>*.bin | vegeta report
Read results from stdin

-type=json
Metrics as JSON

-type=hist[0,100ms,200ms]
Latency histogram with buckets

-every=*<t>*
Print a report at each interval

Reading the text report

Field	Meaning
Requests	Total, rate sent, throughput
Latencies	Min, mean, p50/p90/p95/p99, max
Success	Share of 200-399 responses
Status Codes	Count per code; 0 = send failed
Error Set	Distinct errors seen

PLOT AND ENCODE

vegeta plot *<file>*.bin > *<file>*.html
Interactive latency-over-time chart