Reverse Engineering for Input Modeling
Input Parameter Model Inference from Network Traces

Manuel Leithner | MATRIS Group, SBA Research | FH St. Pölten

Combinatorial Testing

Testing is an essential task in any secure software development lifecycle. Combinatorial Testing combines
► mathematical coverage guarantees
► small test sets
► flexible extensions (constraints, budgeting, . . .)

Typical Workflow
1. Input modeling: Generate model of parameters & values
2. Test generation: Construct combinatorial test set (Covering Array [CA])
3. Test translation: Transform abstract test cases to concrete messages
4. Test execution: Submit messages to target, record response
5. Test oracle: Decide whether test was handled correctly

Combinatorial testing requires a model (IPM) of input parameters, their values, and potentially existing constraints.
► Additional effort to create and maintain
► Often not available in practice
► Must reverse engineer to test proprietary protocols

Thesis Contribution
► First work to combine protocol reverse engineering based on network traces with input parameter modeling
► Translates generated test cases to concrete protocol messages
► Open Source implementation based on Netzob
► Identifies avenues for future work, e.g. shortcomings of model definitions

Message Format

Netzob protocol message format (“Symbol”): Tree made up of fields, each containing
► Node variables, encapsulating other nodes
  ▪ Repeat child node
  ▪ Alternative between child nodes
  ▪ Concatenation of child nodes
► Leaf variables, contain concrete data
  ▪ Data variables, primitive data types
  ▪ Relation variables, based on other fields

Primitive data types
Integers, strings, IPs, . . .

Modeled using boundary values
1. Partition domain of parameter based on semantics
2. Identify values at boundaries of partitions, e.g. min, −1, 0, 1, max
3. Mark negative (invalid) values, e.g. larger than allowed

Node variables
Repetition, choice, concatenation

Modeled using metaparameters
► Number of repetitions
► Which alternative to select for a node

State of research: Coverage definition lacking
► Split metaparameter test set from value test set, combine later
► Nested node variables result in huge model or incomplete coverage
► Additional research required to solve identified shortcomings

Summary
► Combinatorial testing is an efficient & effective black-box testing method
► Offers mathematically guaranteed coverage and small test set sizes
► Requires input parameter model, often not available in practice
► Approach: Reverse engineering to infer input parameter models
► Pluggable mechanism allows choice of test set generator
► Translates generated test sets to concrete protocol messages