

タイトル案 (5 つ以上)

- 機械学習を用いたファズデータのチェックサム及びハッシュ値の推定 (決定)

著者

藤本高史 (Takafumi Fujimoto)

研究の位置づけ (1. 背景; 2. それに対する取り組み; 3. 課題; 4. 本研究)

- 近年のファジングツールにおいてファズデータ作成に用いられる変異ベース手法は, チェックサムやハッシュにより入力データは殆ど通過しないため, 網羅率が低いことがあげられる.
- [難波 2018] は, 機械学習を用いて入力データに対するチェックサムの推定を行い, 68%の正答率となった.
- 8 byte データに対するチェックサムのみの学習しか行っていないため, 汎用性に欠ける.
- 機械学習を用いて, 8 byte 以上の入力データに対するチェックサム及びハッシュ値の推定を行う.
 - ハッシュ値は CRC16, CRC32, MD5, SHA1.
 - 可変長データに対応.

結果

- ランダム文字列と英文に対するチェックサムや各ハッシュ値において, 高精度で推定.

表 1: 英文の学習精度

	学習データ	テストデータ	学習誤差	推定誤差
checksum	72 %	72 %	0.66	0.66
CRC16	57%	57%	0.57	0.57
CRC32	53%	53%	0.65	0.65
MD5	15%	15%	2.5	2.5
SHA1	5%	5%	2.9	2.9

表 2: ランダム文字列の学習精度

	学習データ	テストデータ	学習誤差	推定誤差
checksum	63%	63%	0.95	0.95
CRC16	55%	55%	0.63	0.63
CRC32	52%	52%	0.66	0.66
MD5	5%	5%	3.09	3.09
SHA1	5%	5%	3.07	3.07