

Test Suite Prioritization by Interaction Coverage

Jaymie Strecker

University of Maryland

strecker@cs.umd.edu

Renée Bryce

University of Nevada Las Vegas

reenebryce@cs.unlv.edu

Atif Memon

University of Maryland

atif@cs.umd.edu



Main contributions

- Extension of interaction testing for Event-Driven Software
 - Specific application to GUI-based testing
 - Empirical results

Overview

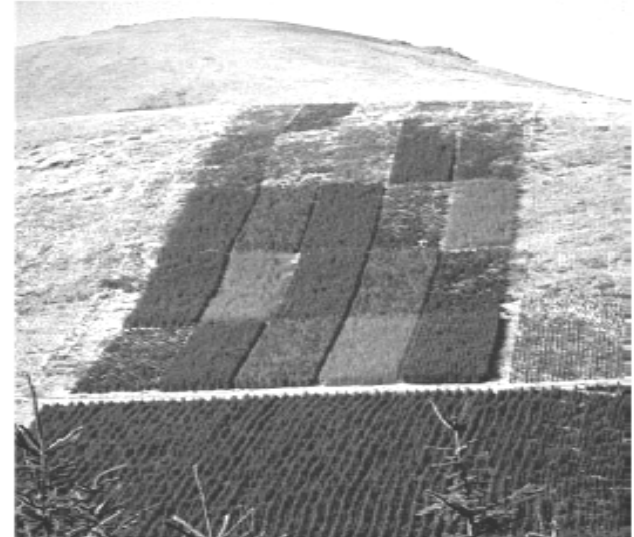
- Background
 - Event-Driven Software (EDS)
 - Interaction testing
- Prioritization by Interaction Coverage
 - Example
 - Experiments and Results using 4 GUI-based programs
- Conclusions and Future Work

Background – Event-Driven Software

- Examples:
 - Web applications, GUIs, network protocols, device drivers, embedded software, ...
- Challenge:
 - Large number of possible states require a large number of tests that cover different event sequences!
- Preliminary work-in-progress:
 - Understand the event-interaction adequacy of test suites
 - Prioritize pre-existing test suites by event-interactions

Background – Interaction Testing

- Interaction testing has been used in several fields:
 - Agriculture
 - Combinatorial chemistry
 - Genomics
 - **Software/hardware testing**



Study of Mozilla web browser found 70% of defects with 2-way coverage; ~90% with 3-way; and 95% with 4-way. [Kuhn *et. al.*, 2002]

Interaction testing of 109 software-controlled medical devices recalled by US FDA uncovered 97% of flaws with 2-way coverage; and only 3 required higher than 2. [Kuhn *et. al.*, 2004]

Software Interaction Test Example

Input:

Hardware	Operating System	Network Connection	Memory
PC	Windows XP	Dial-up	64MB
Laptop	Linux	DSL	128MB
PDA	FreeBSD	Cable	256MB

Sample test:

Test No.	Hardware	Operating System	Network Connection	Memory
1	PC	Windows XP	Dial-up	64MB

2-, 3-, and 4-way interactions covered

(PC, Windows XP)	(PC, Dial-up)	(PC, 64MB)
(Windows XP, Dial-up)	(Windows XP, 64MB)	(Dial-up, 64MB)
(PC, Windows XP, Dial-up)	(PC, Windows XP, 64MB)	(Windows XP, Dial-up, 64MB)
(PC, Windows XP, Dial-up, 64MB)		

Pair-wise Software Interaction Test Example

Hardware	Operating System	Network Connection	Memory
PC	Windows XP	Dial-up	64MB
Laptop	Linux	DSL	128MB
PDA	FreeBSD	Cable	256MB

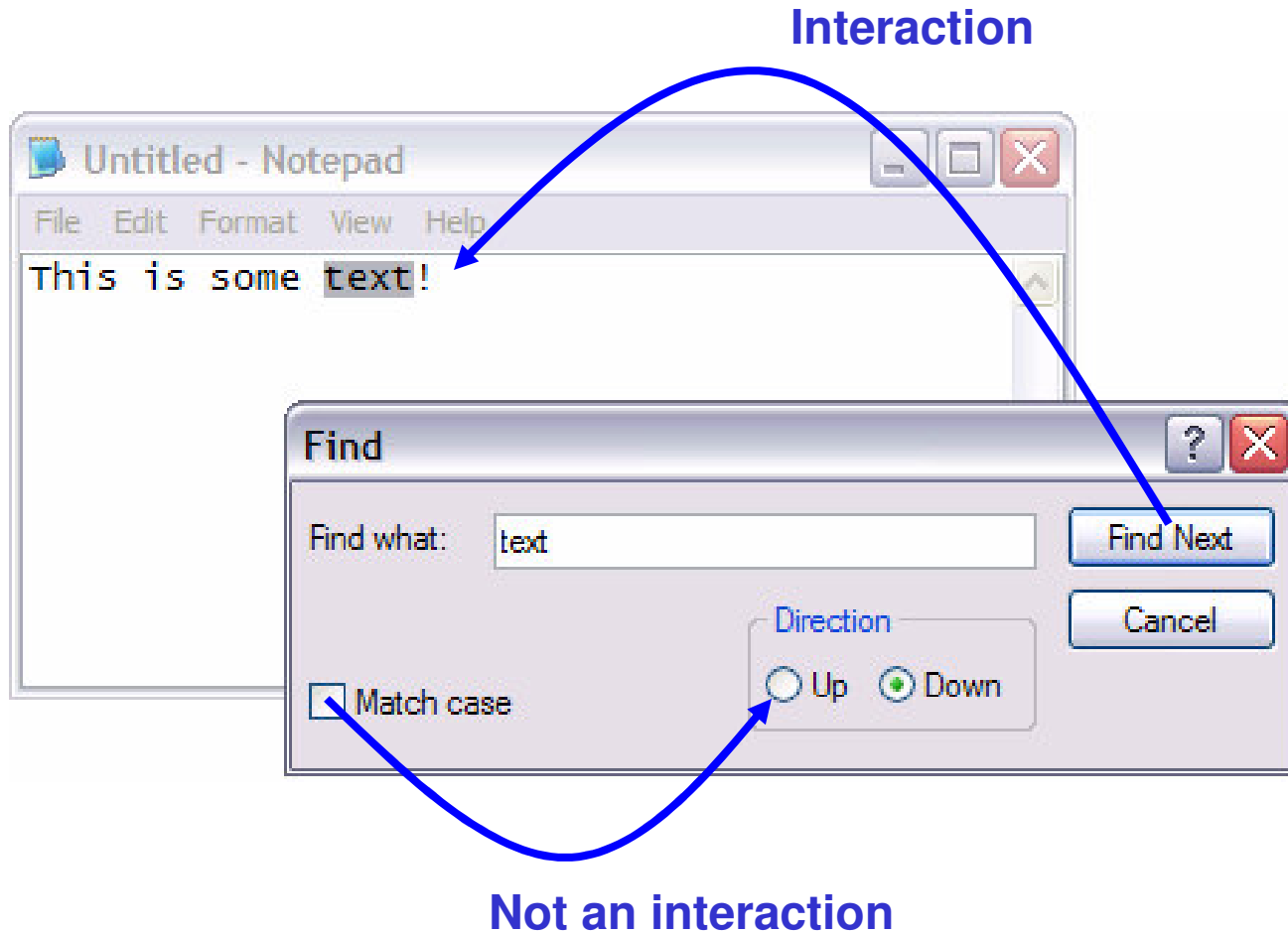
Four factors (components) have three levels (options) each

Test No.	Hardware	Operating System	Network Connection	Memory
1	PC	Windows XP	Dial-up	64MB
2	Laptop	FreeBSD	DSL	64MB
3	PDA	Linux	Cable	64MB
4	Laptop	Windows XP	Cable	128MB
5	Laptop	Linux	Dial-up	256MB
6	PC	Linux	DSL	128MB
7	PDA	Windows XP	DSL	256MB
8	PC	FreeBSD	Cable	256MB
9	PDA	FreeBSD	Dial-up	128MB

Test Suite Prioritization by Interaction Coverage

- Modify previous application of *software interaction testing* to prioritize test suites, rather than generate test suites from scratch
- Compare to other prioritization techniques:
 - Test lengths (long to short, short to long)
 - Unique coverage of events
 - Random ordering

Example window with events



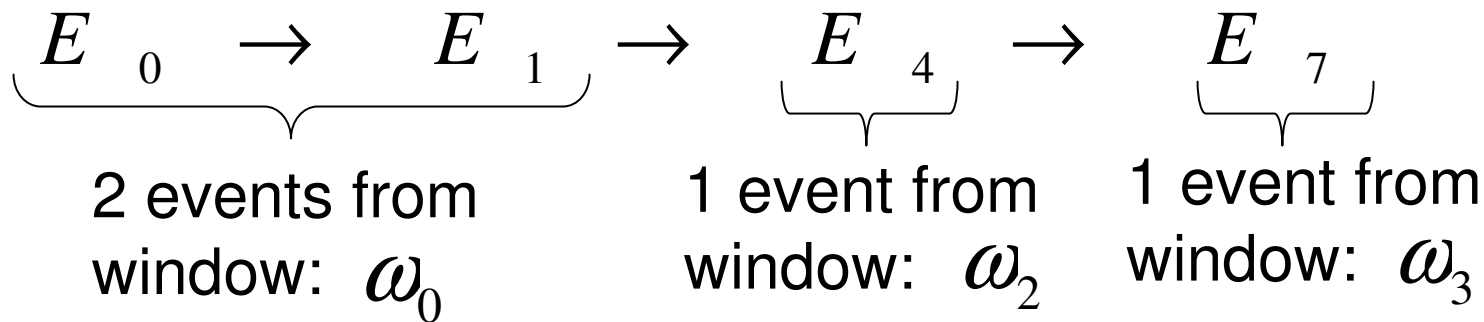
Example test

Input of 4 windows with 2 events each

ω_0	ω_1	ω_2	ω_3
E_0	E_2	E_4	E_6
E_1	E_3	E_5	E_7

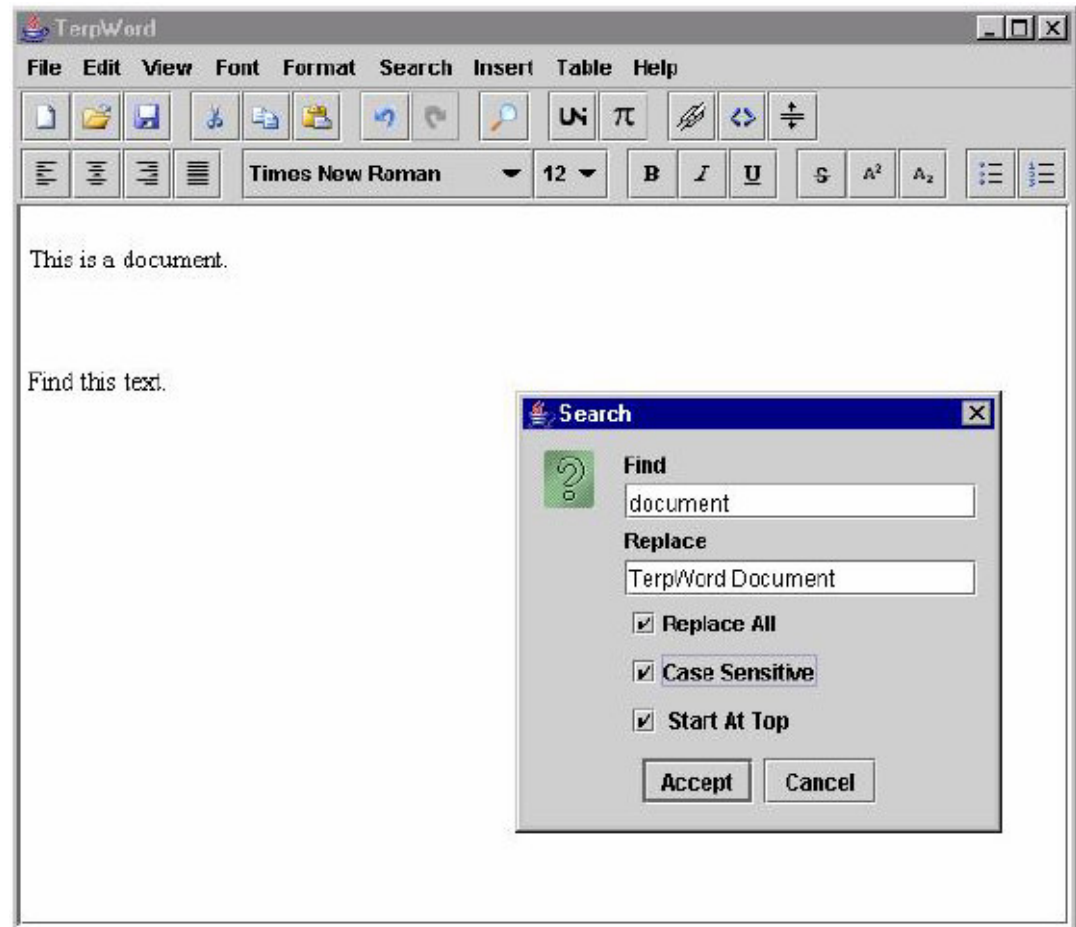
E_0, E_4
 E_0, E_7
 E_1, E_4
 E_1, E_7
 E_4, E_7
 E_0, E_4, E_7
 E_1, E_4, E_7

Sample test



Experiments

- Four programs and pre-existing test suites:
 - TerpCalc
 - TerpPaint
 - TerpSpreadsheet
 - TerpWord



Programs under test

	TerpCalc	TerpPaint	TerpSpread-sheet	TerpWord
Windows	2	11	9	12
Widgets	82	200	145	112
Events	85	247	188	156
LOC	9916	18376	12791	4893
Classes	141	219	125	104
Methods	446	644	579	236
Branches	1306	1277	1521	452

Interaction Coverage in Test Suites

2-way (pairs)	Calc	Paint	Spread-sheet	Word
No. of pairs	1065	26253	14721	10815
% pairs covered in test suite	99%	46%	51%	65%
Min # of tests required to cover unique pairs	67	199	103	146

3-way (triples)	Calc	Paint	Spread-sheet	Word
No. of triples	n/a	1,577,160	626,012	439,734
% triples covered in test suite	n/a	6%	10%	17%
Min # of tests required to cover unique triples	n/a	185	127	168

Terms

- Average percentage of faults detected (APFD)
 - Measures how rapidly a prioritized test suite detects faults
- Prioritization techniques
 - 2- and 3-way IC
 - Unique event
 - Long to short
 - Short to long
 - Random

APFD for TerpCalc

All 2-ways covered →

% of test suite	Unique event	2-way IC	Long to short	Short to long	Random
10	2.57%	4.67%	4.41%	0.67%	4.08%
20	7.64%	10.84%	10.56%	3.10%	9.56%
30	13.50%	17.41%	17.38%	6.52%	16.05%
40	19.71%	24.18%	24.35%	10.27%	23.06%
50	26.26%	31.27%	31.47%	15.68%	30.20%
60	33.15%	38.53%	38.63%	21.69%	37.46%
70	40.21%	45.89%	46.02%	28.06%	44.79%
80	47.48%	53.31%	53.48%	34.69%	52.16%
90	54.85%	60.76%	60.94%	41.77%	59.60%
100	62.32%	68.26%	68.44%	49.18%	67.10%

10% = 30 tests

APFD for TerpWord

% of test suite	Unique event	2-way IC	3-way IC	Long to short	Short to long	Random
10	7.48%	7.93%	7.99%	7.60%	0.30%	6.08%
20	16.50%	17.25%	17.28%	16.95%	4.26%	14.72%
30	25.90%	27.15%	27.12%	26.48%	9.48%	23.88%
40	35.54%	37.05%	37.02%	36.38%	15.96%	33.40%
50	45.54%	46.94%	46.91%	46.27%	24.18%	43.08%
60	55.54%	56.84%	56.81%	56.17%	33.35%	52.98%
70	65.54%	66.81%	66.71%	66.09%	42.99%	62.94%
80	75.54%	76.81%	76.67%	76.09%	52.86%	72.90%
90	85.54%	86.81%	86.66%	86.09%	62.83%	82.86%
100	95.54%	96.81%	96.66%	96.09%	72.83%	92.84%

10% = 25 tests

APFD for TerpSpreadsheet

% of test suite	Unique event	2-way IC	3-way IC	Long to short	Short to long	Random
10	6.98%	6.01%	6.01%	5.69%	0.01%	1.28%
20	16.48%	15.20%	14.58%	14.82%	0.16%	6.73%
30	25.97%	24.97%	24.35%	24.44%	0.54%	14.50%
40	35.46%	34.96%	34.34%	34.18%	1.25%	23.13%
50	44.96%	44.96%	44.34%	44.18%	2.30%	32.44%
60	54.70%	54.96%	54.34%	54.18%	4.04%	41.83%
70	64.57%	64.96%	64.34%	64.18%	11.22%	51.36%
80	74.44%	74.96%	74.34%	74.18%	19.63%	61.05%
90	84.37%	84.96%	84.34%	84.18%	28.66%	70.80%
100	94.37%	94.96%	94.34%	94.18%	38.66%	80.68%

10% = 30 tests

APFD for TerpPaint

% of test suite	Unique event	2-way IC	3-way IC	Long to short	Short to long	Random
10	4.62%	6.21%	6.46%	7.53%	0.79%	2.15%
20	12.14%	15.31%	15.55%	16.78%	2.99%	6.21%
30	20.48%	24.87%	24.96%	26.23%	6.02%	12.24%
40	29.02%	34.53%	34.58%	35.92%	9.56%	19.42%
50	37.90%	44.21%	44.38%	45.82%	13.78%	27.63%
60	47.31%	54.18%	54.29%	55.76%	18.33%	36.42%
70	57.03%	64.18%	64.29%	65.71%	23.39%	45.73%
80	66.94%	74.18%	74.29%	75.69%	29.31%	55.37%
90	76.92%	84.18%	84.29%	85.69%	36.70%	65.22%
100	86.92%	94.18%	94.29%	95.69%	45.54%	75.18%

10% = 30 tests

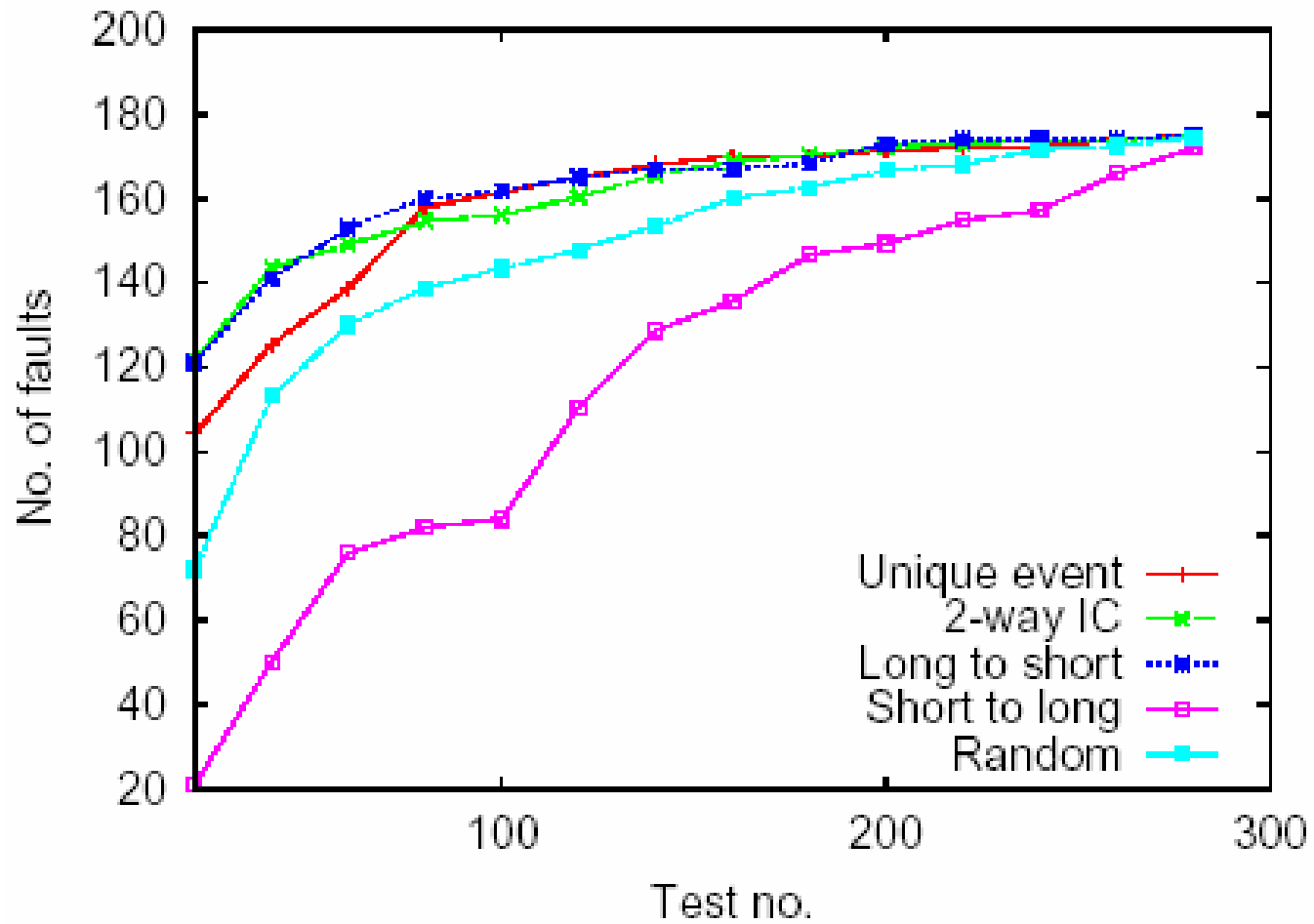
Summary of Results

- TerpCalc and TerpWord benefit the most from prioritization by interaction coverage
 - Their test suites have the highest percentage of interaction coverage
- TerpSpreadsheet and TerpPaint do not benefit
 - These are the largest programs and have test suites with low interaction coverage

Conclusions and Future Work

- Prioritization by interaction coverage improves the rate of fault detection for two of our programs and their pre-existing test suites
- Future work will include:
 - additional empirical studies
 - extensions to examine prioritization by interaction coverage in regards to test lengths
 - prioritization by incremental *t*-way coverage

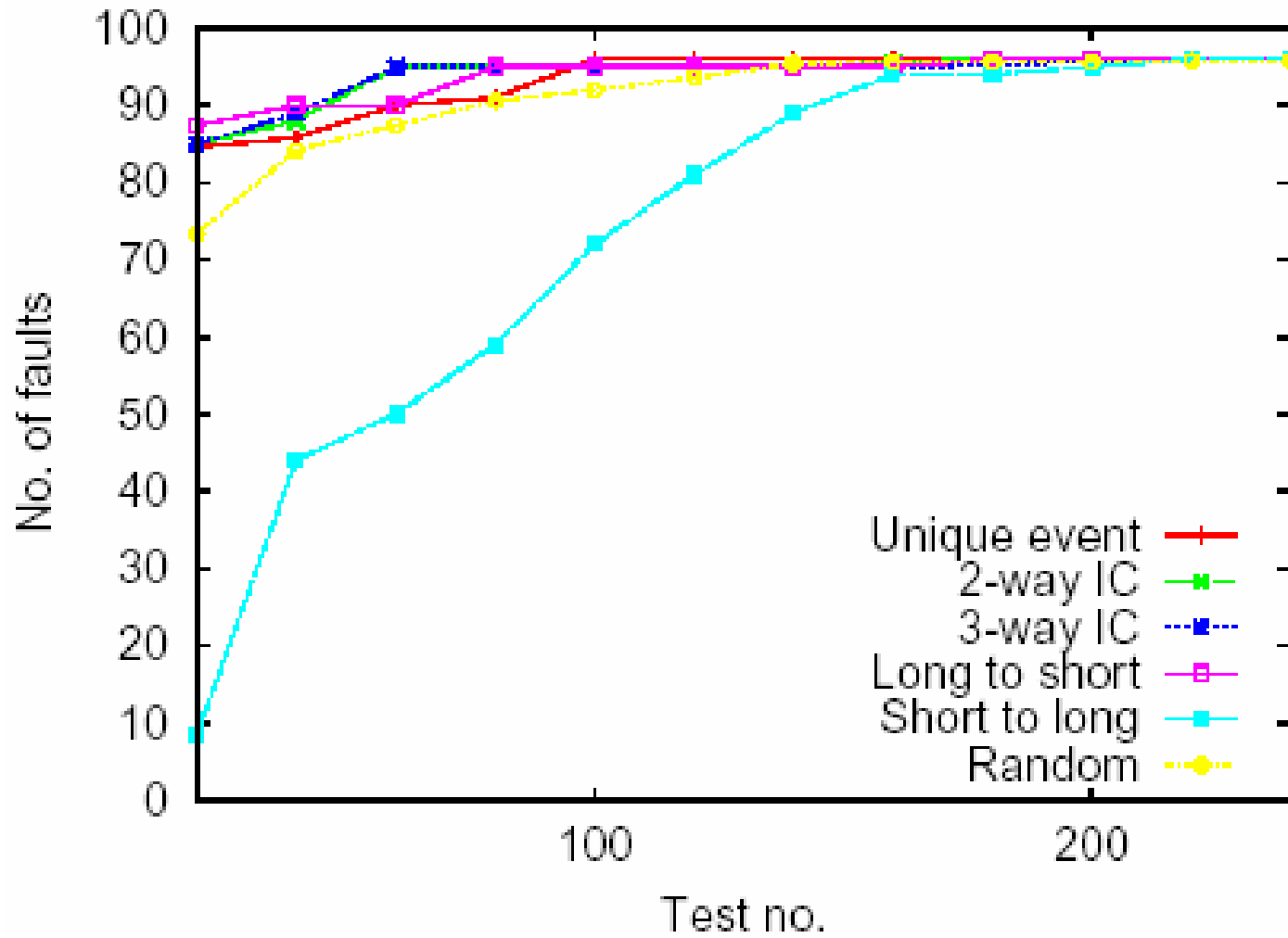
TerpCalc



APFD for TerpCalc: First 10 Tests

Test no.	Uniq. event	2-way IC	Long to short	Short to long	Rand.
1	0.02%	0.02%	0.02%	0.00%	0.01%
2	0.09%	0.11%	0.07%	0.01%	0.03%
3	0.16%	0.18%	0.11%	0.01%	0.04%
4	0.31%	0.33%	0.21%	0.02%	0.08%
5	0.39%	0.41%	0.26%	0.02%	0.12%
6	0.56%	0.58%	0.38%	0.03%	0.18%
7	0.64%	0.67%	0.45%	0.03%	0.22%
8	0.81%	0.85%	0.61%	0.04%	0.29%
9	0.90%	0.95%	0.71%	0.05%	0.33%
10	1.07%	1.16%	0.91%	0.06%	0.41%

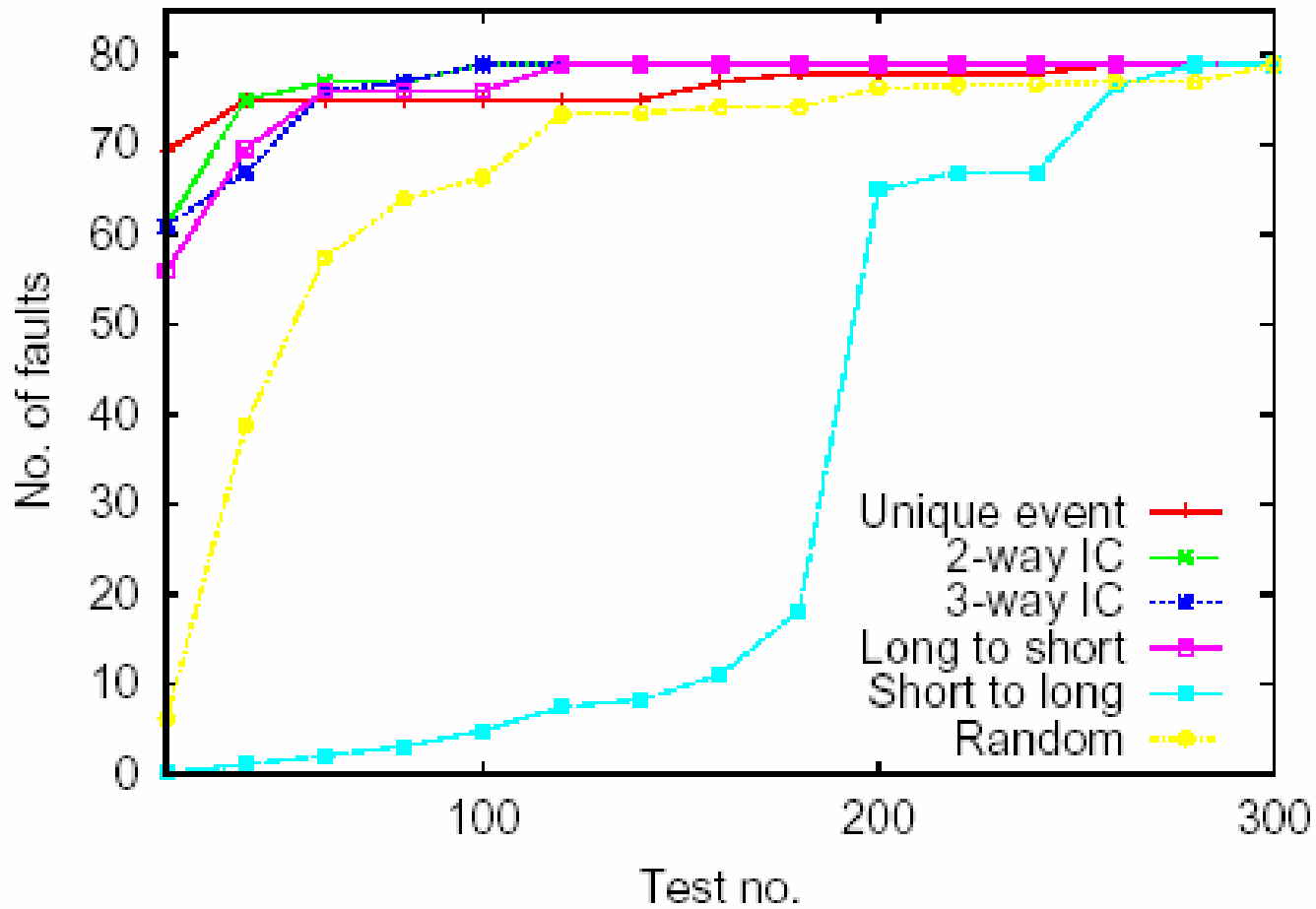
TerpWord



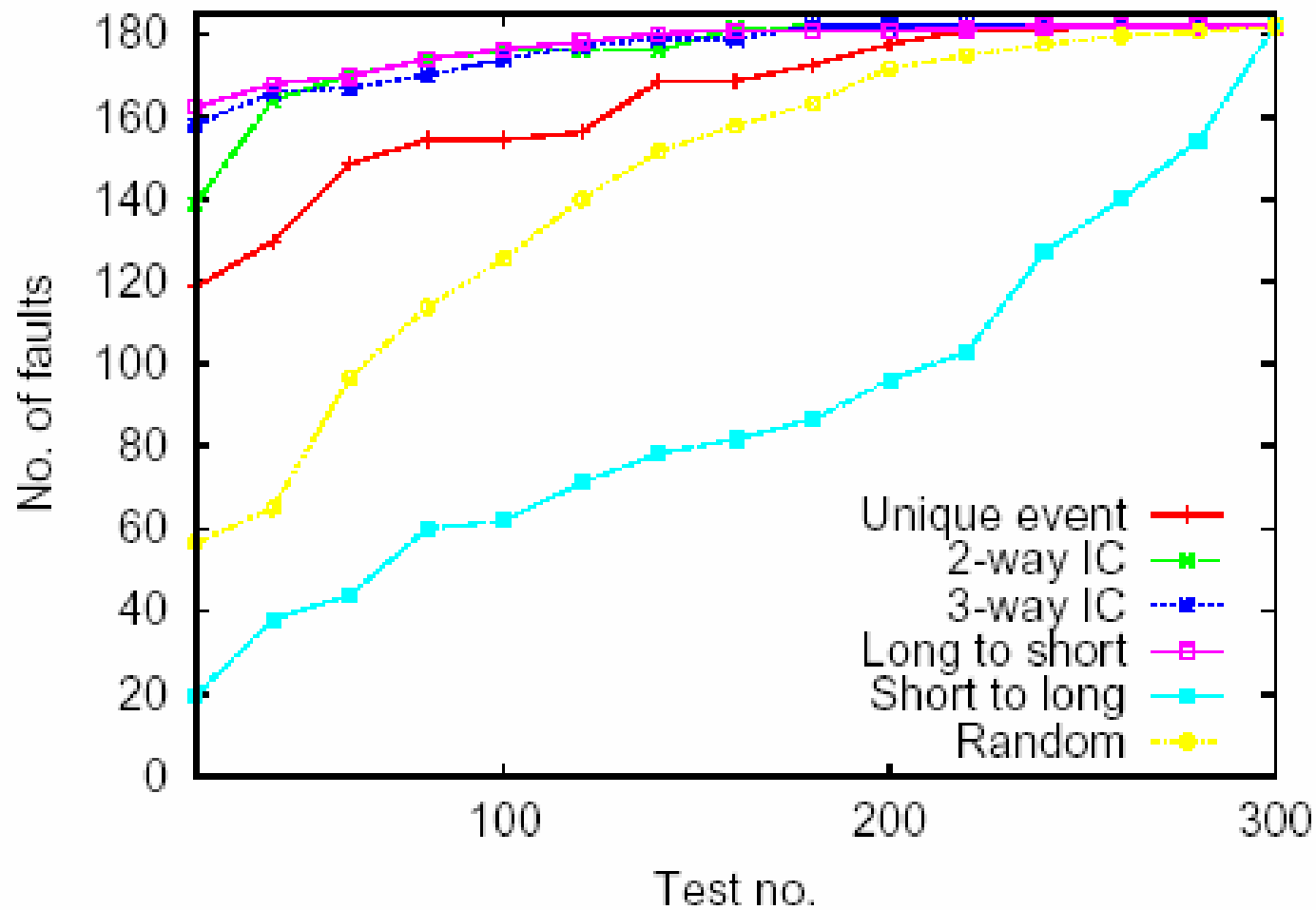
APFD for TerpWord: First 10 Tests

Test no.	Uniq. event	2-way IC	3-way IC	Long to short	Short to long	Rand.
1	0.02%	0.09%	0.09%	0.06%	0.00%	0.03%
2	0.09%	0.32%	0.32%	0.22%	0.00%	0.11%
3	0.24%	0.59%	0.60%	0.41%	0.00%	0.22%
4	0.47%	0.88%	0.89%	0.64%	0.00%	0.34%
5	0.71%	1.19%	1.18%	0.88%	0.00%	0.50%
6	0.98%	1.50%	1.47%	1.16%	0.00%	0.69%
7	1.27%	1.81%	1.77%	1.44%	0.00%	0.92%
8	1.56%	2.13%	2.07%	1.73%	0.00%	1.16%
9	1.87%	2.45%	2.38%	2.04%	0.00%	1.41%
10	2.21%	2.77%	2.71%	2.35%	0.00%	1.66%

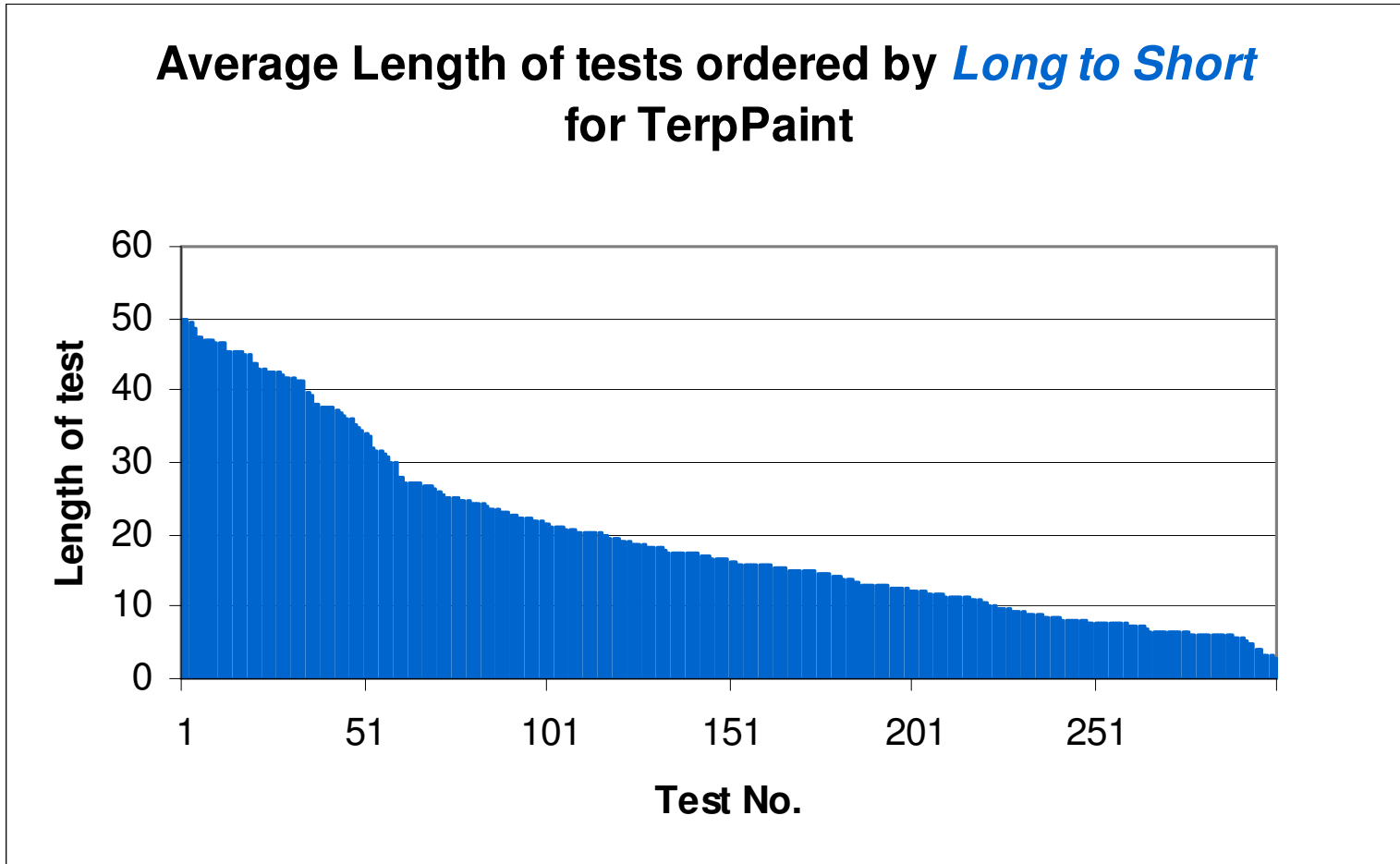
TerpSpreadsheet



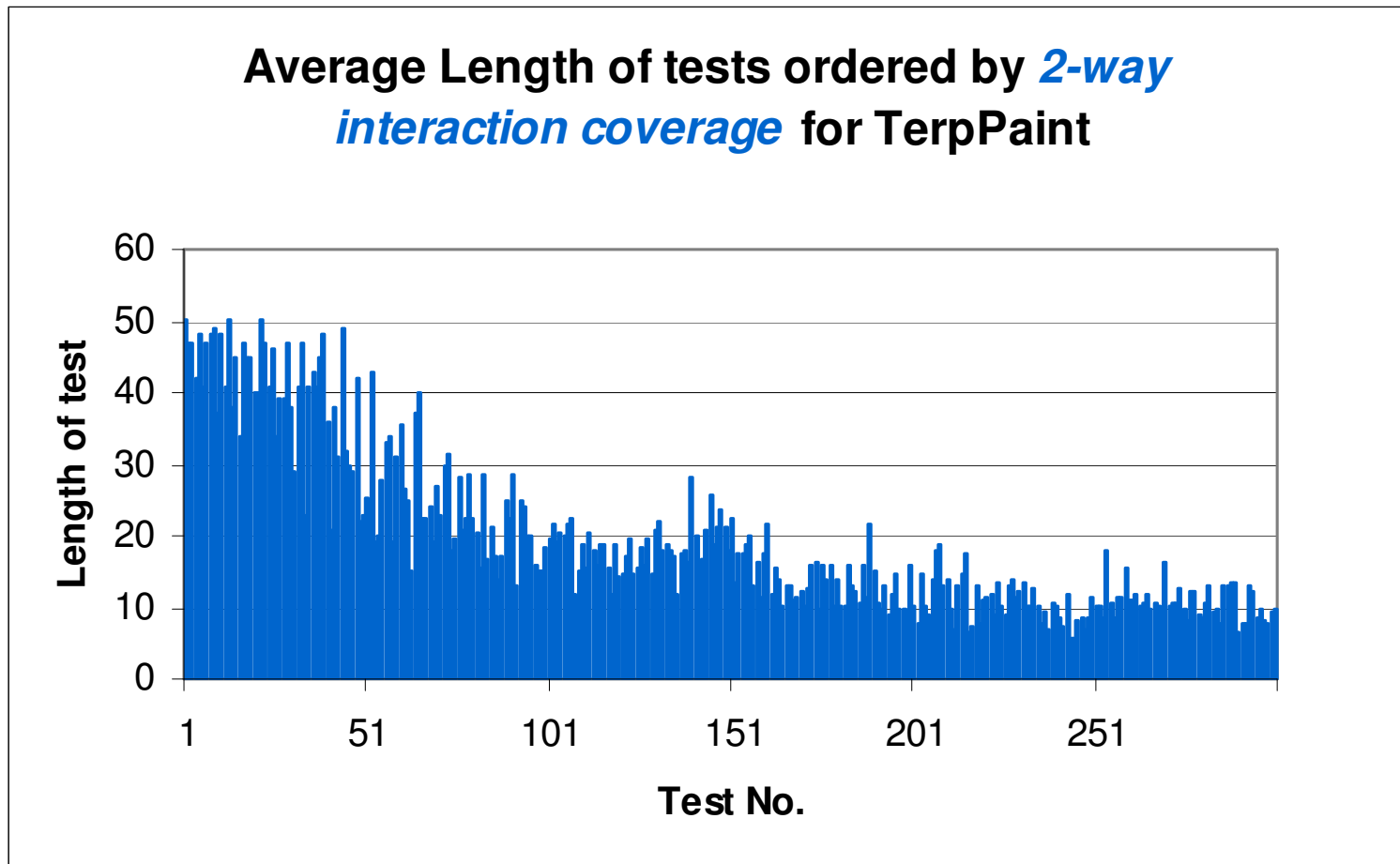
TerpPaint



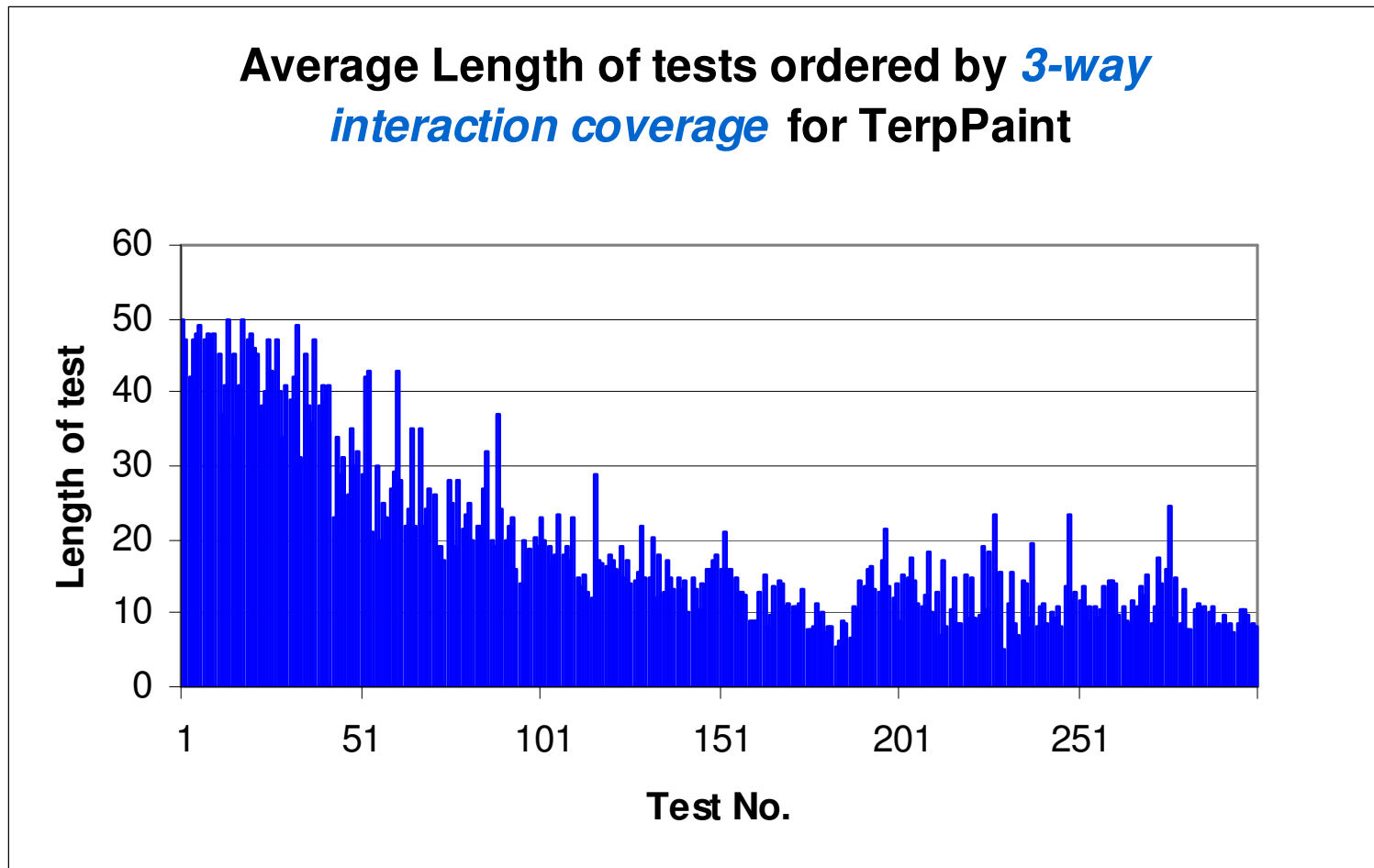
TerpPaint Test Lengths: Long to Short



TerpPaint Test Lengths: 2-way interaction coverage



TerpPaint Test Lengths: 3-way interaction coverage



Future Work: Time-based test suite prioritization

Test No.	Time to execute test (in secs.)	No. of 2-way interactions covered
1	10	6
2	5	4
3	5	4
4	15	8