



CONFIGURATION-AWARE TESTING

Myra Cohen, BSSw Fellow

<https://www.cs.iastate.edu/~mcohen>

mcohen@iastate.edu

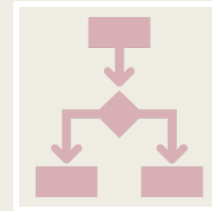
Overview



Motivation



Combinatorial
Testing

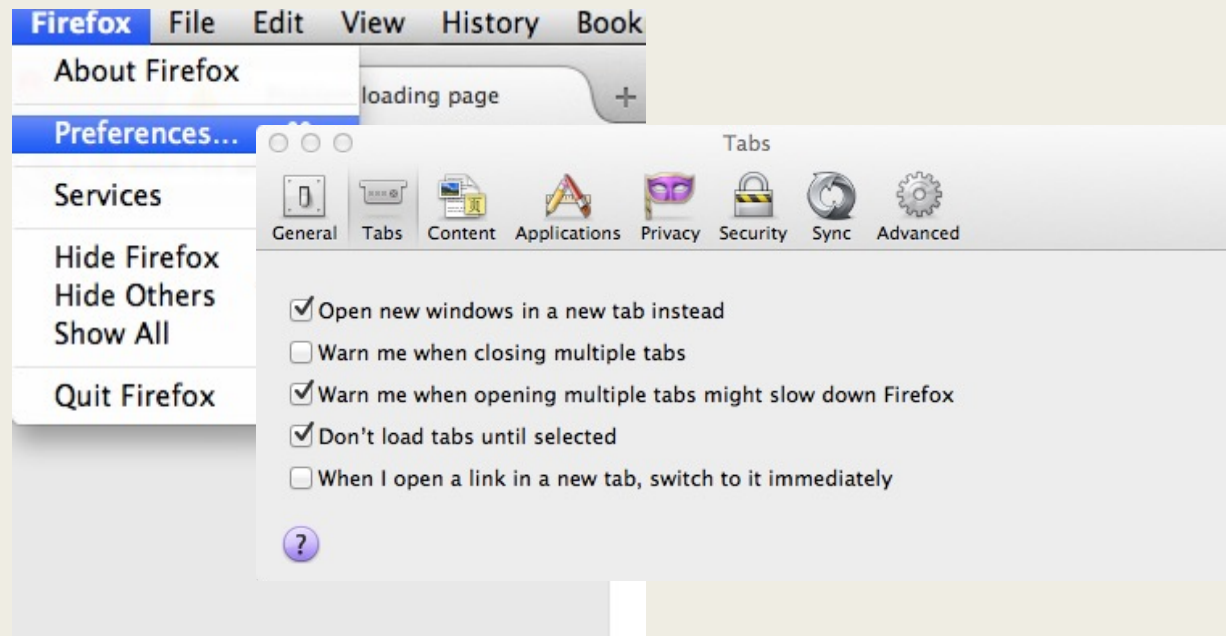


Fault Characterization

Configurability



Configurability



Configuration Interfaces

☐ Warn me when closing multiple tabs

Configuration Interfaces

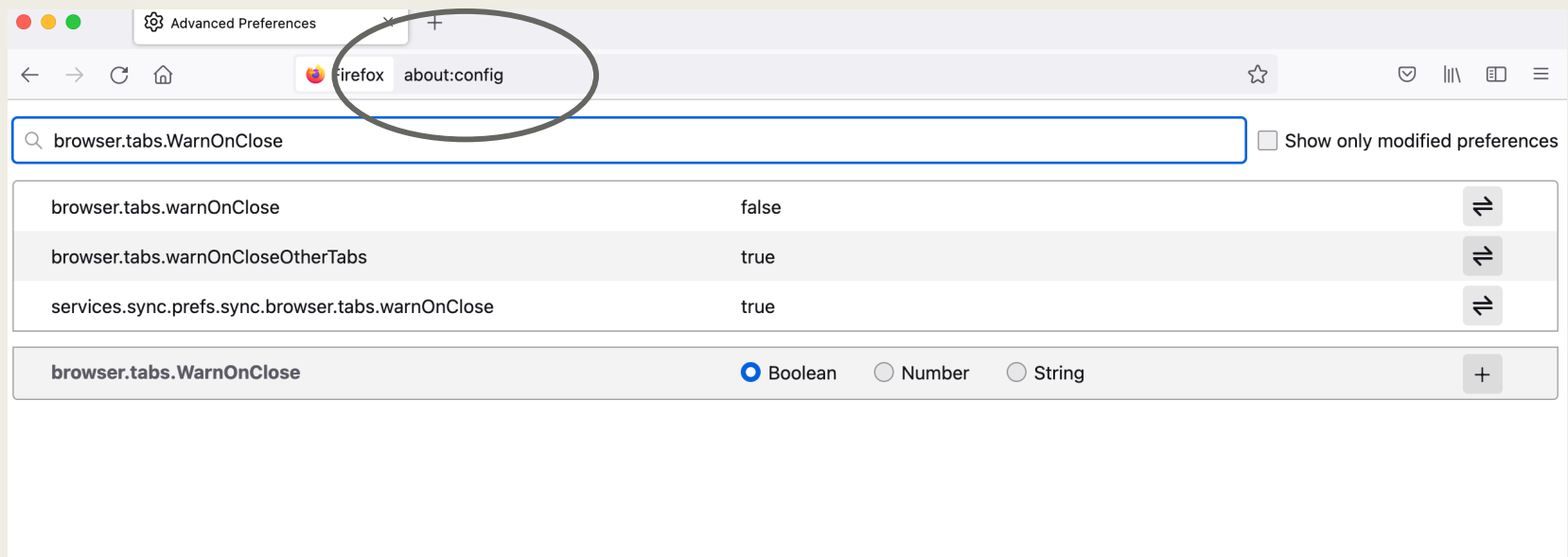
☐ Warn me when closing multiple tabs

Prefs.js



```
user_pref("browser.tabs.warnOnClose", false);
```

Configuration Interfaces



Configuration Interfaces



Proceed with Caution

Changing advanced configuration preferences can impact Firefox performance or security.

☒ Warn me when I attempt to access these preferences

Accept the Risk and Continue

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ?

gactacgatcgggc

[Clear](#)

Query subrange ?

From

To

BLAST

Search **database Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

[Algorithm parameters](#)

[Restore default search parameters](#)

General Parameters

Max target sequences

100 ?

Select the maximum number of aligned sequences to display ?

Short queries

☒ Automatically adjust parameters for short input sequences ?

Expect threshold

10 ?

Word size

28 ?

Max matches in a query range

0 ?

Scoring Parameters

Match/Mismatch Scores

1,-2 ?

Gap Costs

Linear ?

Filters and Masking

Filter

☒ Low complexity regions ?

☐ Species-specific repeats for: Homo sapiens (Human)

Mask

☒ Mask for lookup table only ?

☐ Mask lower case letters ?

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ?

gactacgatcgggc

[Clear](#)

Query subrange ?

From

To

BLAST

Search **database Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

 **Algorithm parameters**

[Restore default search parameters](#)

General Parameters

Max target sequences

100

Select the maximum number of aligned sequences to display ?

Short queries

☒ Automatically adjust parameters for short input sequences ?

Expect threshold

10

Word size

28

Max matches in a query range

0

Scoring Parameters

Match/Mismatch Scores

1,-2

Gap Costs

Linear

Filters and Masking

Filter

☒ Low complexity regions ?
☐ Species-specific repeats for: Homo sapiens (Human)

Mask

☒ Mask for lookup table only ?
☐ Mask lower case letters ?

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ?

gactacgatcgggc

[Clear](#)

Query subrange ?

From

To

BLAST

Search **database Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

Algorithm parameters

[Restore default search parameters](#)

General Parameters

Max target sequences

100

Select the maximum number of aligned sequences to display ?

Short queries

☒ Automatically adjust parameters for short input sequences ?

Expect threshold

10

Word size

28

Max matches in a query range

0

Scoring Parameters

Match/Mismatch Scores

1,-2

Gap Costs

Linear

Linear

Existence: 5 Extension: 2

Existence: 2 Extension: 2

Existence: 1 Extension: 2

Existence: 0 Extension: 2

Existence: 3 Extension: 1

Existence: 2 Extension: 1

Existence: 1 Extension: 1

Filters and Masking

Filter

Mask

☒ Mask for lookup table only ?

☐ Mask lower case letters ?

MEGAHIT (DNA Assembler)



Credit
M. Cashman

- DNA sequenced into small segments (reads)
- Assembly combines reads into longer *continuous sequences*
- Result is a certain *number of continuous sequences*



MEGAHIT (DNA Assembler)

- DNA sequenced into small segments (reads)
- Assembly combines reads into longer *continuous sequences*
- Result is a certain *number of continuous sequences*



Credit
M. Cashman





Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

--k-min

--k-max

--k-step

--k-list

--min-contig-len

Output Objects

Output Assembly name





Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

--k-min

--k-max

--k-step

--k-list

--min-contig-len

Output Objects

Number of continuous sequences

284

Configuration

Default





Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Job Status

```
56 [utils.h : 126] Real: 0.4915 user: 0.3962 sys: 0.0720 maxrss: 24688
57 --- [Sat Mar 24 04:35:57 2018] k-max reset to: 119 ---
58 --- [Sat Mar 24 04:35:57 2018] k list: 21,29,39,59,79,99,119 ---
59 --- [Sat Mar 24 04:35:57 2018] Extracting solid (k+1)-mers for k = 21 ---
```

--k-list



--min-contig-len

300 ≤

2000

Output Objects

Number of continuous
sequences

284

Configuration

Default



Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

--k-min

--k-max

--k-step

--k-list

--min-contig-len

Output Objects

Number of continuous sequences

1 ≤

1 ≤

1 ≤

300 ≤

284

119

285

≤ 127

≤ 255

≤ 28

2000

Configuration

Default

k-max=119





Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

2

--k-min

1 ≤

21

≤ 127

--k-max

1 ≤

119

141

≤ 255

--k-step

1 ≤

12

≤ 28

--k-list



--min-contig-len

300 ≤

2000

Output Objects

Number of continuous sequences

284

285

285

Configuration

Default

k-max=119

Default-Manual



Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

2

2

--k-min

1 ≤

21

21

≤ 127

--k-max

1 ≤

119

141

99

≤ 255

--k-step

1 ≤

12

12

≤ 28

--k-list



--min-contig-len

300 ≤

2000

Output Objects

Number of continuous sequences

284

285

285

289

Configuration

Default

k-max=119

Default-Manual

k-max=99



Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.



Run

Configure

Job Status

Result

Input Objects

Read Library

rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count

2

2

--k-min

1 ≤

21

21

≤ 127

--k-max

1 ≤

119

141

99

≤ 255

--k-step

1 ≤

12

12

≤ 28

--k-list



--min-contig-len

300 ≤

2000

Output Objects

Number of sequences

284

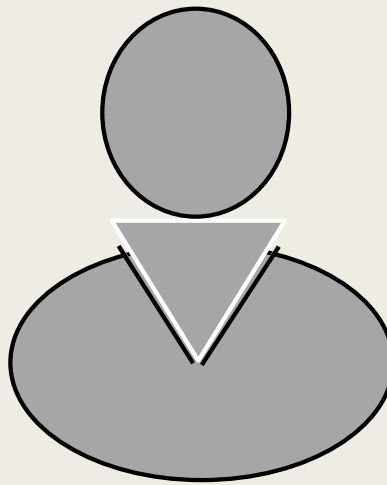
285

285

289

Configuration

Developer Feedback



MEGA
HIT

Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.

Run

Configure

Job Status

Result

Input Objects

Read Library

PE.reads

Parameters

hide advanced

Parameter preset

--min-count		2	2	
--k-min	1 ≤	21	21	≤ 127
--k-max	1 ≤	119	141	99 ≤ 255
--k-step	Default = 10 1 ≤	12	12	≤ 28
--k-list	+			
--min-contig-len	300 ≤			2000

Output Objects

Number of continuous sequences

Configuration

Default

k-max=119

Default-Manual

k-max=99



Assemble Reads with MEGAHIT v1.1.1

Assemble metagenomic reads using the MEGAHIT assembler.

Run

Input Objects

Read Library rhodo.art.q20.PE.reads

Parameters (5 advanced parameters showing) [hide advanced](#)

Parameter preset

--min-count		2	2
--k-min	1 ≤	21	21 ≤127
--k-max	1 ≤	119	141 99 ≤255
--k-step	1 ≤	12	12 ≤28

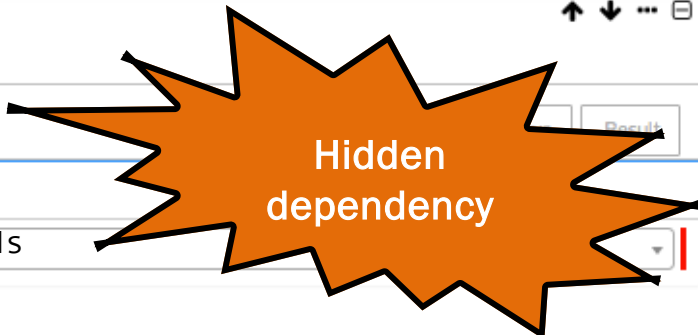
--k-list [21,29,39,59,79,99,119,141]

--min-contig-len 300 ≤ 2000

Output Objects

Number of continuous sequences

Configuration Default k-max=119 Default-Manual k-max=99



MEGA
HIT

Assemble Reads with MEGAHIT v1.1.1
Assemble metagenomic reads using the MEGAHIT assembler

↑ ↓ ⋮ ☰

Run

Configure

Job Status

Result

Input Objects

Read Library

rhods.art q20.1.2.1.aus

Parameters (5 advanced)

Parameter preset

--min-count

2

2

--k-min

1 ≤

21

21

≤127

--k-max

1 ≤

119

141

99

≤255

--k-step

1 ≤

12

12

≤28

--k-list

+

--min-contig-len

300 ≤

2000

Output Objects

Number of continuous sequences

Configuration

Default

k-max=119

Default-Manual

k-max=99

Hidden
constraint

$k\text{-max} \leq \text{MAX}(\text{read length}) + 20$ && be an odd number

Challenge for Testing

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Media Player

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Media Player

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Media Player

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Testing the Player

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Test Case: Open video and play to completion

Interaction Fault

Encoding	Format	Cache Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		



crash

Test Case: Open video and play to completion

Configuration-Dependent Security Bugs (CERT)

Overview

Apache Struts 2 framework, versions 2.5 to 2.5.12, with REST plugin insecurely deserializes untrusted XML data. A remote, unauthenticated attacker can leverage this vulnerability to execute arbitrary code in the context of the Struts application.

Description

CWE-502: Deserialization of Untrusted Data - CVE-2017-9805

In Apache Struts 2 framework, versions 2.5 to 2.5.12, the REST plugin uses `xStreamHandler` with an instance of `XStream` to deserialize XML data. Because there is no type filtering, a remote, unauthenticated attacker may send a specially crafted XML payload to execute arbitrary code in the context of the Struts application.

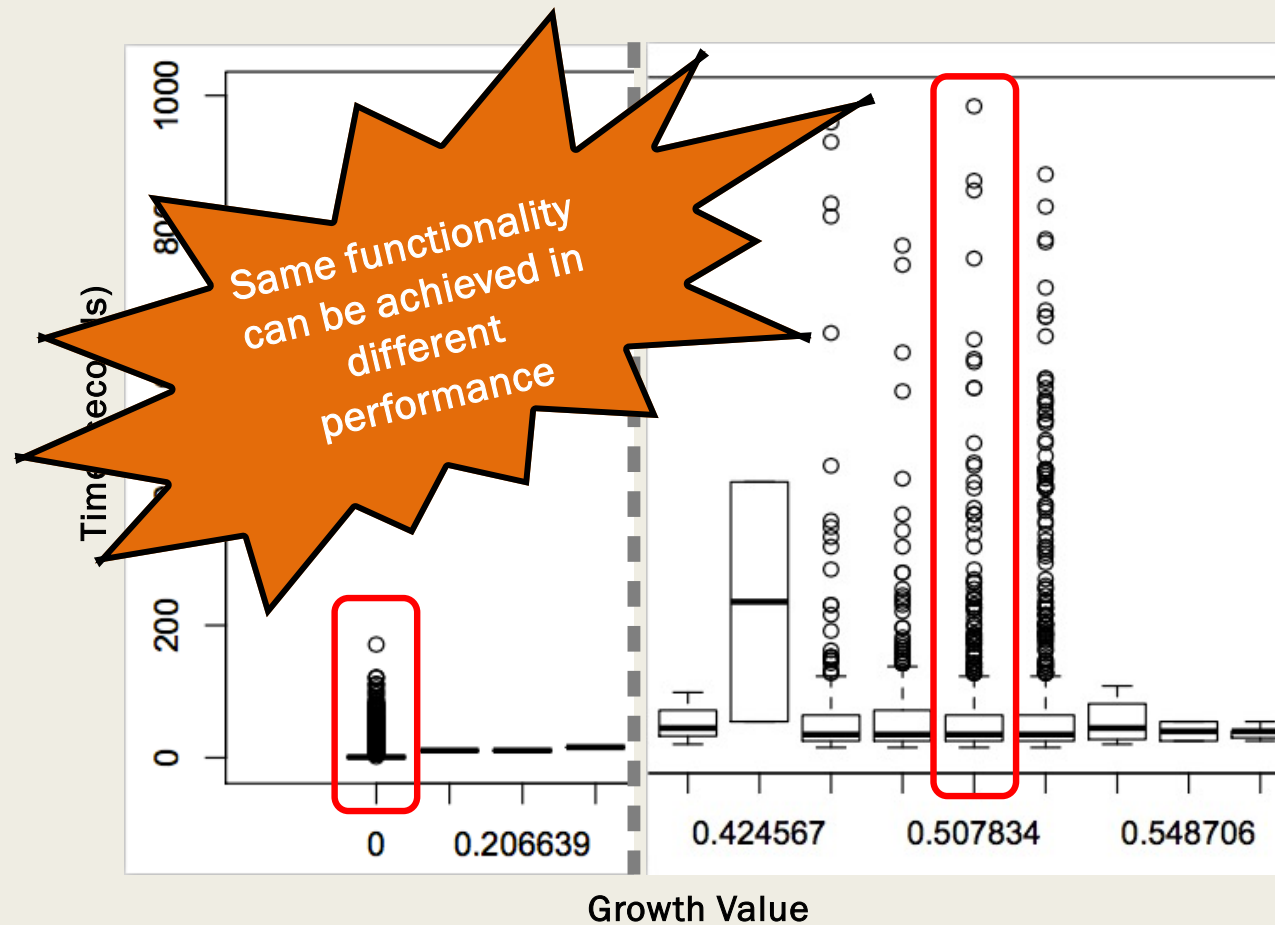
Refer to the researcher's [blog post](#) for more information about this vulnerability. A [Metasploit module](#) with exploit code is publicly available.

Remove or limit the REST plugin

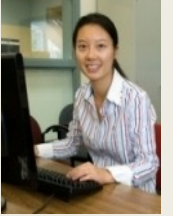
If it is not used, consider removing the REST plugin. Per the vendor, it is also possible to limit its functionality to normal server pages or JSON with the following configuration change in `struts.xml`:

```
<constant name="struts.action.extension" value="xhtml,,json" />
```

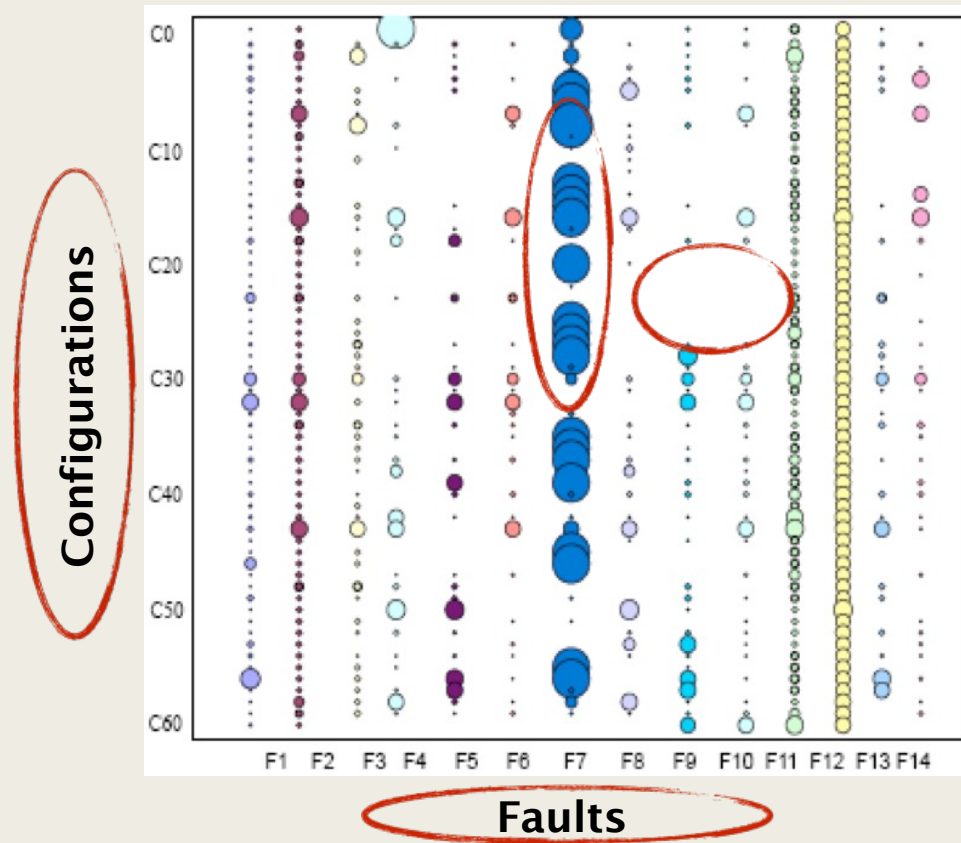
Also Impacts Program Performance



Configuration Dependence



Credit
X. Qu



Real Configuration Spaces

Encoding	Format	Caching Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Real Configuration Spaces

$3 \times 3 \times 3 \times 2 \times 2 = 108$ configurations

Encoding	Format	Caching Level	Closed-Captioning	Network Access
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		



Real Configuration Spaces

$3 \times 3 \times 3 \times 2 \times 2 = 108$ configurations

10 features with 5 options = 9,765,625 configs

Encoder	Encoding	Losses	Losses	Network
MPEG	Audio	Low	Yes	Yes
RAW	Video	Medium	No	No
WAV	Stream	High		

Real Configuration Spaces

$3 \times 3 \times 3 \times 2 \times 2 = 108$ configurations

10 features with 5 options = 9,765,625 configs

4 hours to run test suite \rightarrow 4,459 years to run

En	Learning	Unseen	Network
MP3	1	1	1
RA			
WAV	Stream	High	

Real Configuration Spaces

$3 \times 3 \times 3 \times 2 \times 2 = 108$ configurations

10 features with 5 options = 9,765,625 configs

4 hours to run test suite \rightarrow 4,459 years to run

gcc optimizer (199 options) $\sim 10^6$

Real Configuration Spaces

$3 \times 3 \times 3 \times 2 \times 2 = 108$ configurations

10 features with 5 options = 9,765,625 configs

4 hours to run test suite \rightarrow 4,459 years to run

gcc optimizer (199 options) — 10^{61}

Linux > 10,000 features



Combinatorial Interaction Testing (CIT)

- Sample the space so that all t -way combinations of values occur **AT LEAST** once
- t is defined as strength of testing

Combinatorial Interaction Testing

	Encoding	Format	Caching Level	Closed-Captioning	
1	MPEG	Stream	Medium	Yes	Yes
2	RAW	Video	High	No	No
3	MPEG	Video	Low	No	Yes
4	WAV	Stream	High	No	Yes
5	RAW	Stream	Low	Yes	No
6	MPEG	Audio	High	Yes	No
7	WAV	Video	Medium	Yes	No
8	RAW	Audio	Medium	No	Yes
9	WAV	Audio	Low	Yes	Yes

Combinatorial Interaction Testing (CIT)

	Encoding	Format	Caching Level	Closed-Captioning	
1	MPEG	Stream	Medium	Yes	Yes
2	RAW	Video	High	No	No
3	MPEG	Video	Low	No	Yes
4	WAV	Stream	High	No	Yes
5	RAW	Stream	Low	Yes	No
6	MPEG	Audio	High	Yes	No
7	WAV	Video	Medium	Yes	No
8	RAW	Audio	Medium	No	Yes
9	WAV	Audio	Low	Yes	Yes

Combinatorial Interaction Testing (CIT)

	Encoding	Format	Caching Level	Closed-Captioning	
1	MPEG	Stream	Medium	Yes	Yes
2	RAW	Video	High	No	No
3	MPEG	Video	Low	No	Yes
4	WAV	Stream	High	No	Yes
5	RAW	Stream	Low	Yes	No
6	MPEG	Audio	High	Yes	No
7	WAV	Video	Medium	Yes	No
8	RAW	Audio	Medium	No	Yes
9	WAV	Audio	Low	Yes	Yes

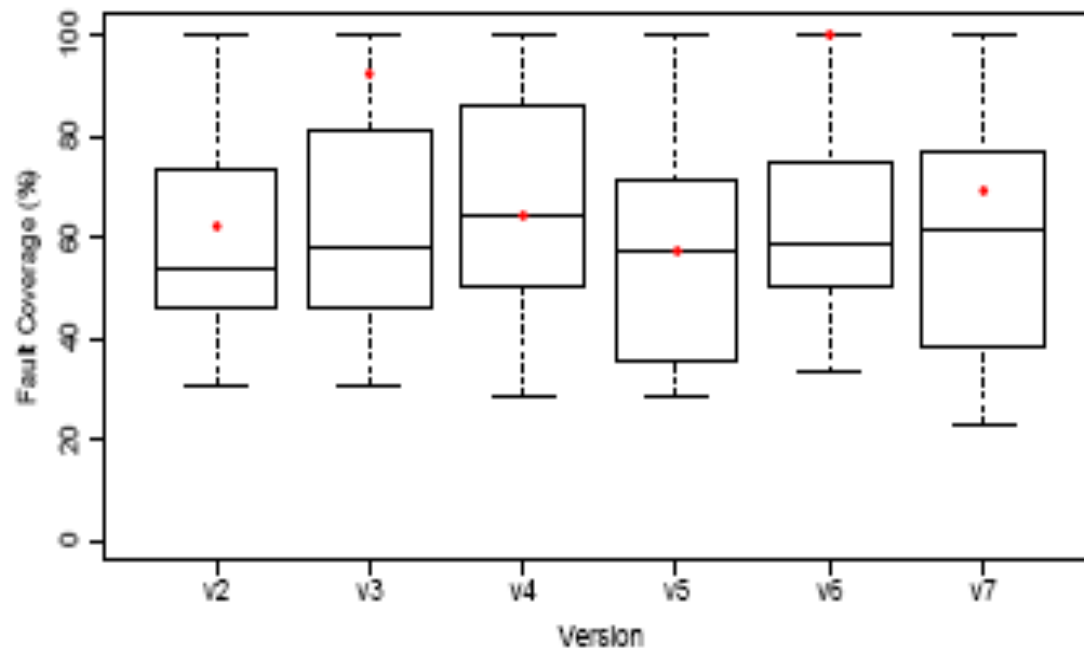
Combinatorial Interaction Testing (CIT)

	Encoding	Format	Caching Level	Closed-Captioning	
1	MPEG	Stream	Medium	Yes	Yes
2	RAW	Video	High	No	No
3	MPEG	Video	Low	No	Yes
4	WAV	Stream	High	No	Yes
5	RAW	Stream	Low	Yes	No
6	MPEG	Audio	High	Yes	No
7	WAV	Video	Medium	Yes	No
8	RAW	Audio	Medium	No	Yes
9	WAV	Audio	Low	Yes	Yes

Higher Strength CIT

- Strength (t) = 3 - all three-way combinations
- Needs *at least 27* combinations

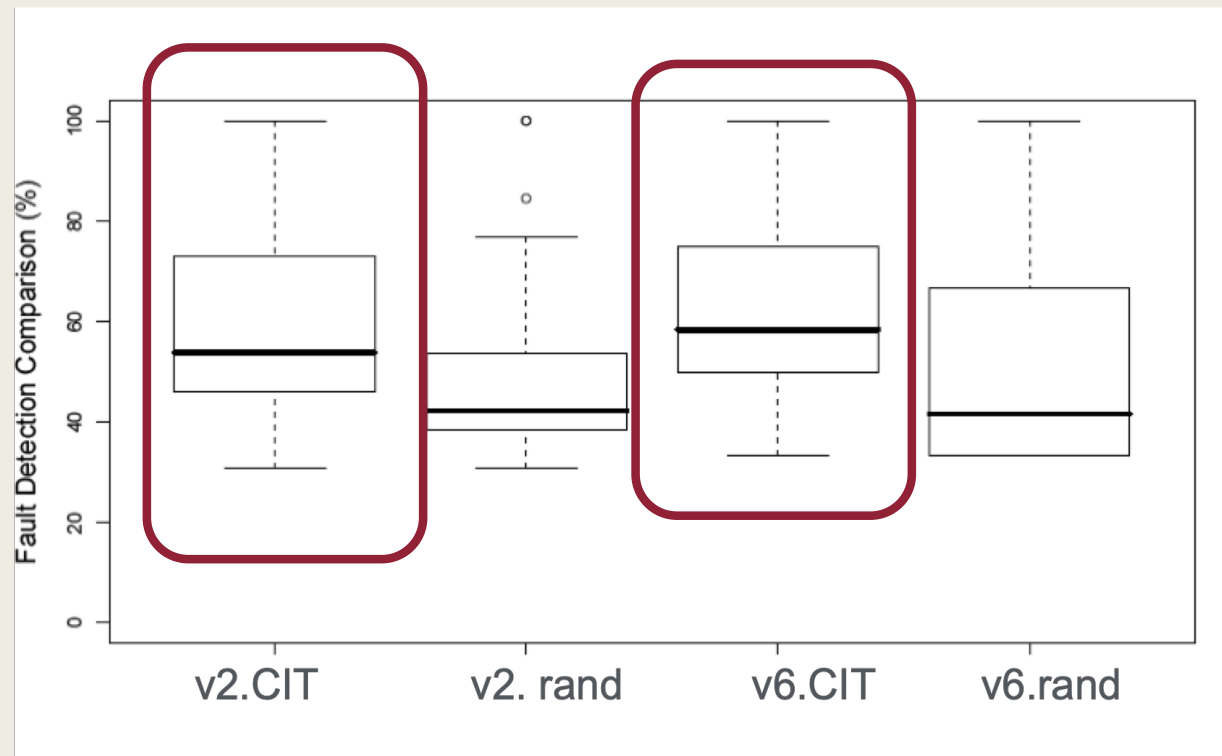
CIT vs. Default Configuration of vim



[Qu 08]

● default configuration

Fault Detection

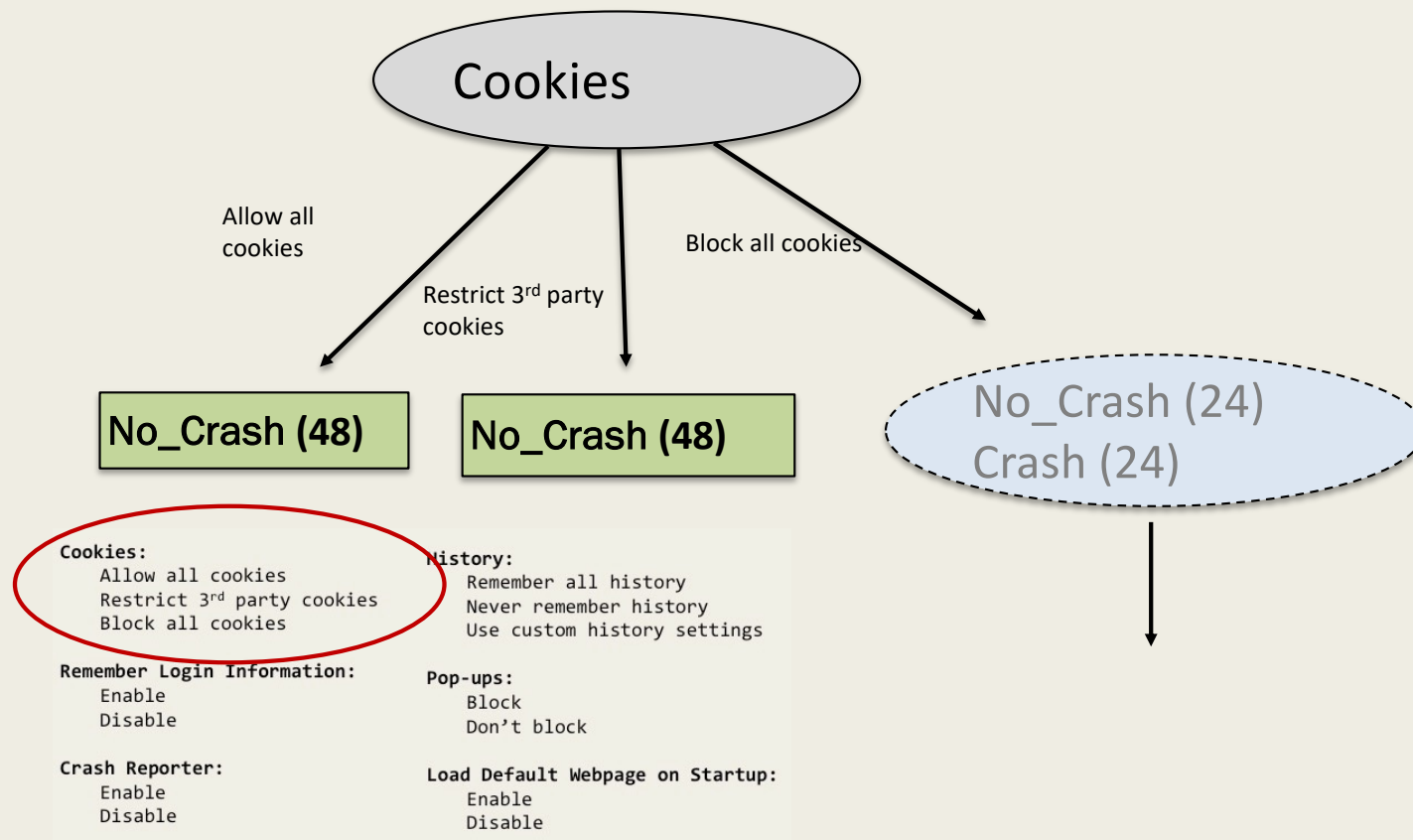




Add Tools for CIT

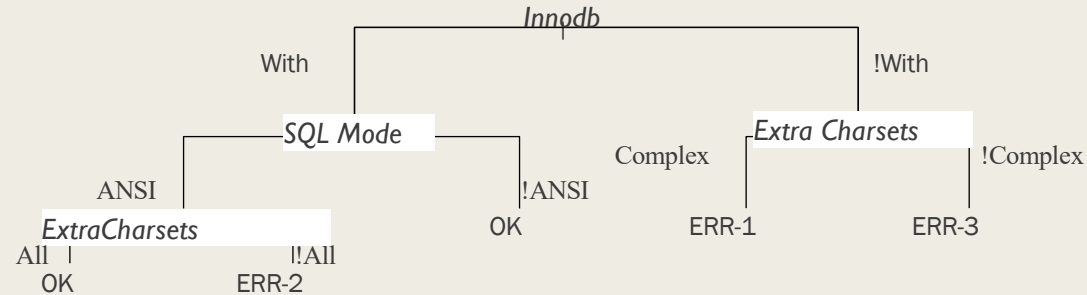


Classification: Decision Tree



Characterizing Test Failures

- Use machine learning techniques (classification trees) to model option setting patterns that explain test failures



Add more on Fault Characterization

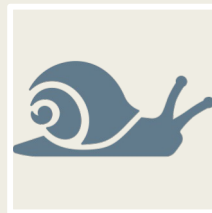


Summary

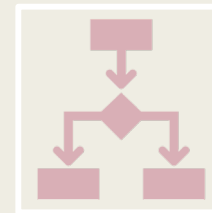
Overview



Motivation



Combinatorial
Testing



Fault Characterization

This work was supported by the Better Scientific Software Fellowship Program, funded by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy (DOE) Office of Science and the National Nuclear Security Administration; and by the National Science Foundation (NSF) under Grant No. 2154495.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the DOE or NSF.