

CCFinder & Gemini Code Clone Detection

Osaka University

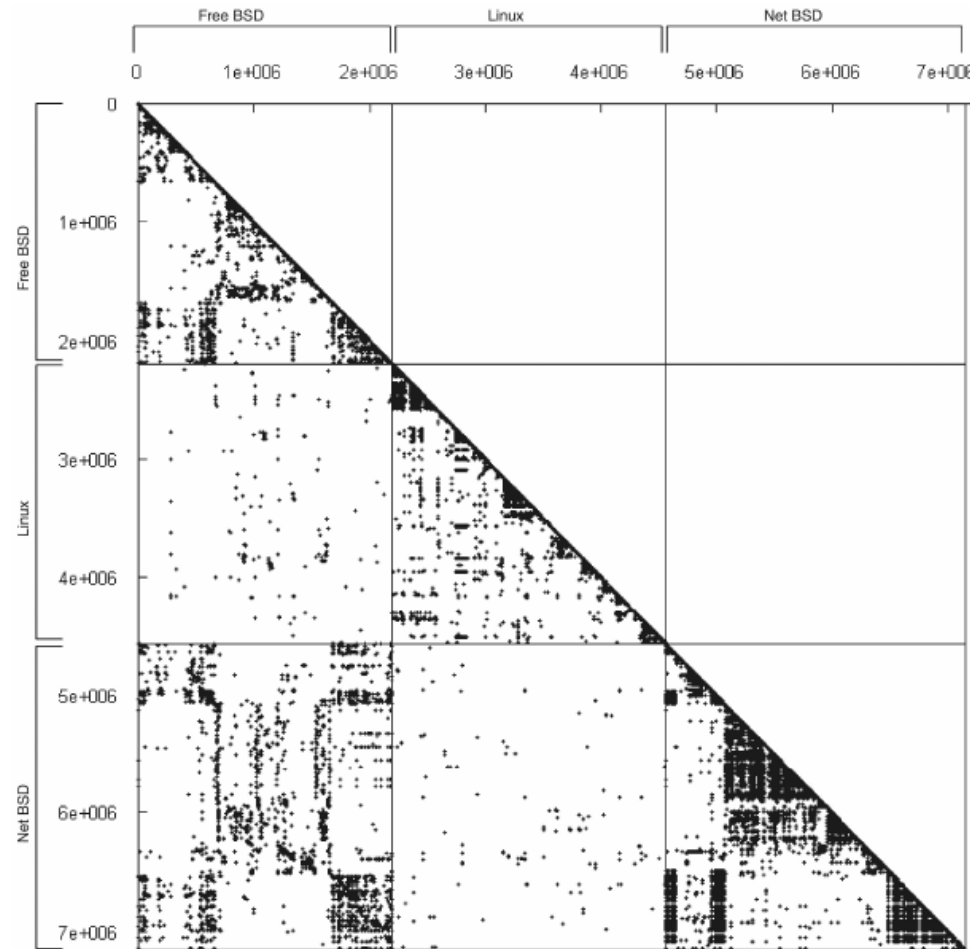
Toshihiro Kamiya et al

Presented by: Eddy Parkinson

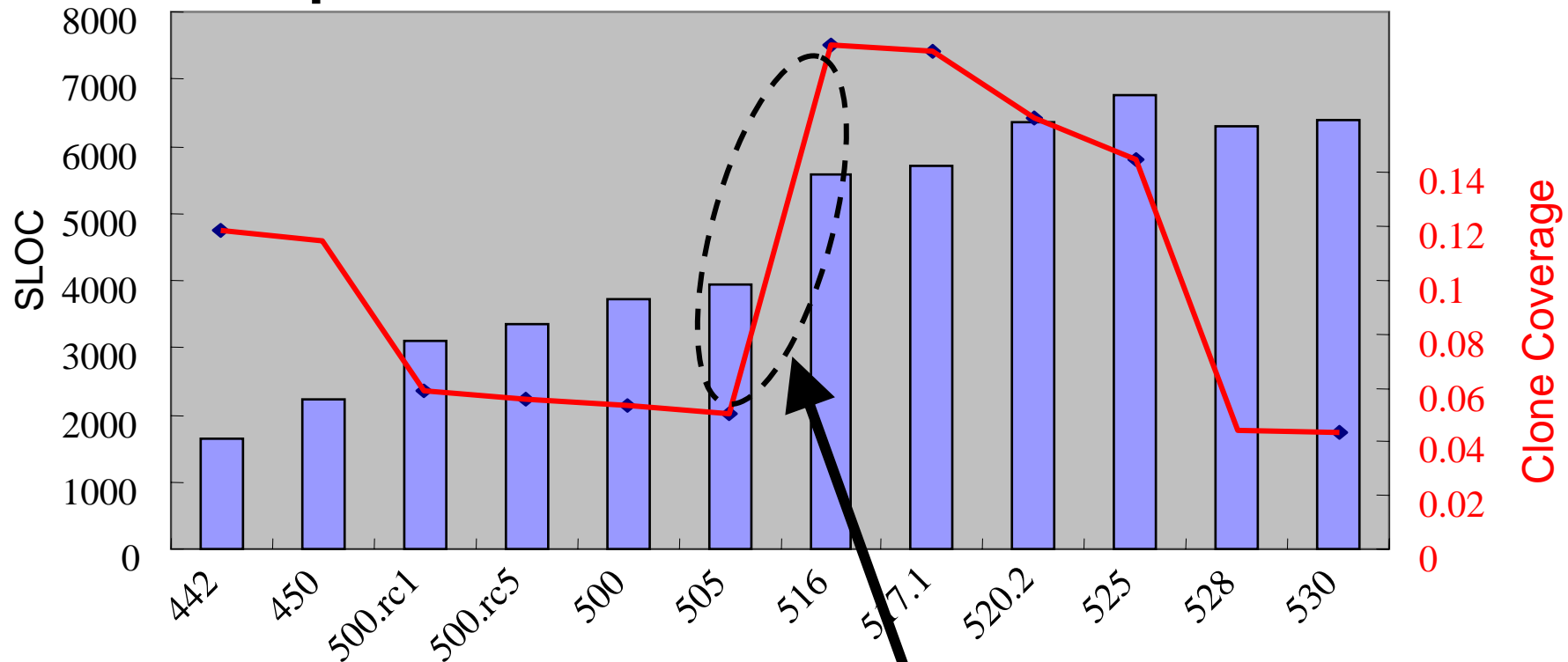
Business problem: Large scale maintenance of cloned code

Cloned code
in FreeBSD
Linux and
NETBSD

Each black dot
indicates
cloned code

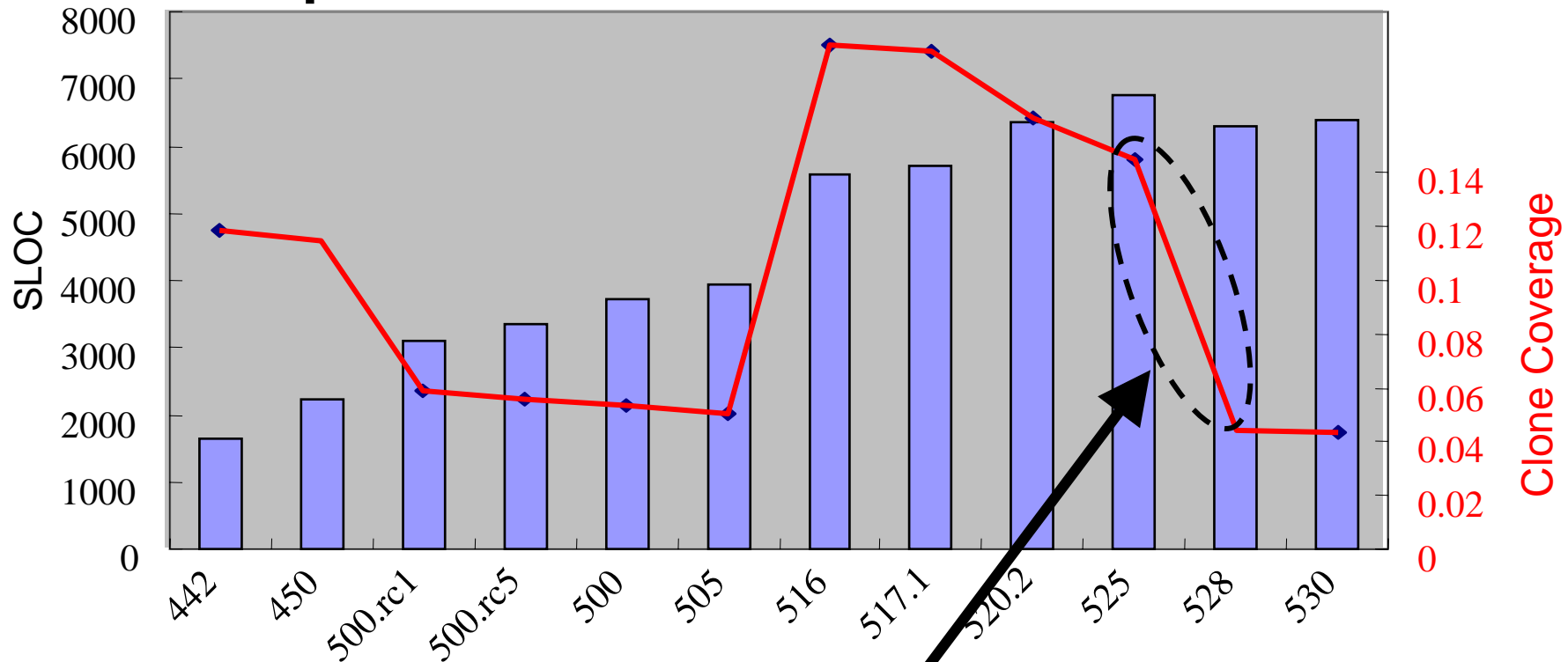


Example: Clone Inserted



FILE	Ver. 505		Ver. 516	
	SLOC	CCvrg	SLOC	CCvrg
charproc.c	566	0	792	7.58%
mml2mid.c	681	0	891	8.31%
mmlproc.c	1585	9.3%	2261	12.47%
note.c	805	4.78%	1348	29.97%

Example: Clone removed



	Ver. 525		Ver. 528	
FILE	SLOC	CCvrg	SLOC	CCvrg
charproc.c	970	4.62%	979	4.51%
mml2mid.c	1318	0	1342	0
mmlproc.c	2506	19.58%	2360	9.06%
note.c	1591	26.36%	1129	0



Key advantages of CCFinder & Gemini

- CCFinder

- Token based clone detection (rather than line based)
- Parses the code to sanitize variables, method names etc
- Converts code into tokens for fast matching
- Supports: C/C++/C#, Java, COBOL, FORTRAN, LISP, Visual Basic

- Gemini - GUI-based clone analysis environment

- Scatter plot
- Metrics graph
- Source code view

GemX (Gemini)

The screenshot displays the GemX application window. The main window has a menu bar with 'File', 'Scope', 'Metrics', 'Settings', and 'Help'. Below the menu bar, there are tabs for 'Sc...', 'File Table', 'Clone-Set Table', 'Scatter Plot', and 'Source Text'. The 'Clone-Set Table' is active, showing a table with columns: Clone-Set ID, Length, POP, RAD, RNR(%), and TKS. The table contains 31 rows of data. A dialog box titled 'Specify Clone-Set Metric Ranges - GemX' is open in the foreground, allowing users to set minimum and maximum values for the metrics: Length, POP, RAD, RNR(%), and TKS. The 'Source Text' tab shows Java code for the application's settings dialog.

Clone-Set ID	Length	POP	RAD	RNR(%)	TKS
1	50	3	0	94.00	11
2	287	2	1	67.94	13
3	56	2	1	94.64	12
4	55	2	1	96.36	12
5	82	2	1	95.12	12
6	86	2	0	93.02	11
7	70	2	0	97.14	12
8	128	2	0	43.75	12
9	79	2	0	46.84	11
10	58	8	1	96.55	12
11	112	2	0	84.82	11
12	130	2	1	29.23	14
13	72	2	0	86.11	11
14	112	2	0	88.39	11
15	100	3	1	95.00	15
16	105	2	1	95.24	15
17	97	2	1	79.38	12
18	69	2	0	84.06	11
19	54	2	0	83.33	10
20	114	2	0	75.44	12
21	118	2	0	67.80	20
22	188	2	1	57.98	13
23	69	2	1	94.20	13
24	172	2	1	40.12	15
25	116	2	0	87.07	22
26	69	2	1	50.72	13
27	74	2	1	63.51	13
28	136	2	0	90.44	12
29	176	2	1	22.73	12
30	75	3	1	73.33	12
31	114	2	1	68.42	12

Specify Clone-Set Metric Ranges - GemX

Metric	Minimum	Maximum
Length	50	332
POP	2	8
RAD	0	2
RNR(%)	50.0	98.42
TKS	10	32

```
17 C:\kamiya\prog\neo\GemX\gemx\dialogs\GemXSettingsDialog.java
60
61 ckResizeScatterPlot = new Button(shellC, SWT.CHECK);
62 ckResizeScatterPlot.setText(Messages.getString("gemx.G
63 gridData = new GridData(GridData.FILL_HORIZONTAL);
64 gridData.horizontalSpan = 3;
65 ckResizeScatterPlot.setLayoutData(gridData);
66 ckResizeScatterPlot.addSelectionListener(new SelectionA
67 public void widgetSelected(SelectionEvent e) {
68 valueResizeScatterPlot = ckResizeScatterPlot.getSe
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87 ckCalcCloneMetricAlways = new Button(shellC, SWT.CHE
88 ckCalcCloneMetricAlways.setText(Messages.getString("g
89 gridData = new GridData(GridData.FILL_HORIZONTAL);
90 gridData.horizontalSpan = 3;
91 ckCalcCloneMetricAlways.setLayoutData(gridData);
92 ckCalcCloneMetricAlways.addSelectionListener(new Sele
```



Usage: CCFinder & Gemini

- Commercial Usage (about 30 companies)
 - NTT Data Corp., Hitachi Ltd., Hitachi GP, NEC soft Ltd., ASTEC Inc., SRA Inc., NASDA, Daiwa Computer, etc
- Detection of Clones in Student Course Work at Osaka University
- Used in a software copyright law suit



CCFinder – more information

- <http://www.ccfinder.net/>
- Some papers:
 - Toshihiro Kamiya, Shinji Kusumoto, and Katsuro Inoue, "**CCFinder: A Multi-Linguistic Token-based Code Clone Detection System for Large Scale Source Code**," IEEE Trans. Software Engineering, vol. 28, no. 7, pp. 654-670, (2002-7).
 - Norihiro Yoshida, Yoshiki Higo, Toshihiro Kamiya, Shinji Kusumoto, Katsuro Inoue, "**On Refactoring Support Based on Code Clone Dependency Relation**", Proceedings of 11th IEEE International Software Metrics Symposium, Como, Italy (2005-9).