

Trustworthiness Models to Categorize and Prioritize Code for Security Improvement

III. OVERALL APPROACH

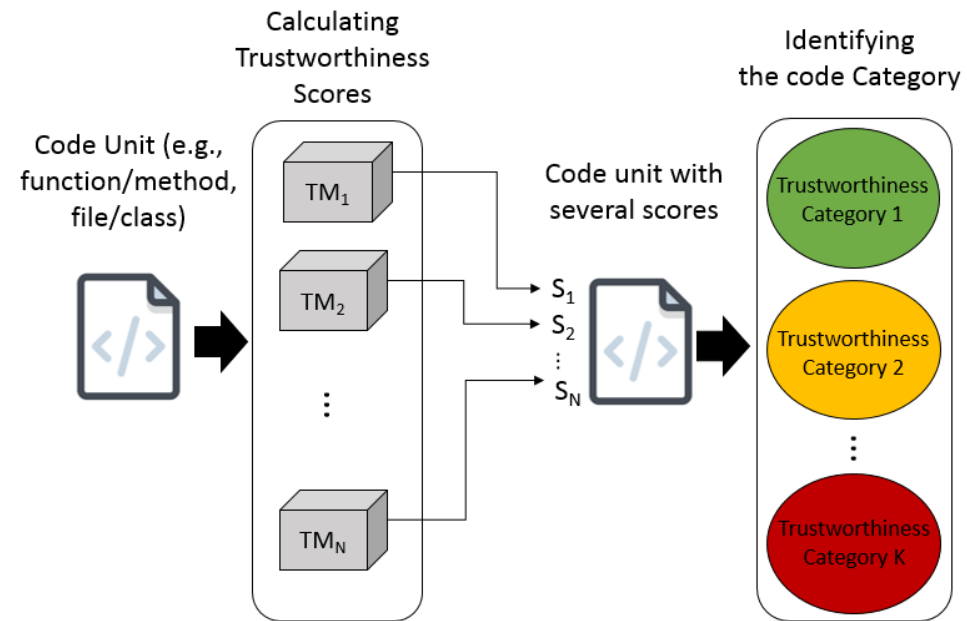


Figure 1: Approach proposed for code categorization from a security perspective.

IV. APPROACH INSTANTIATION

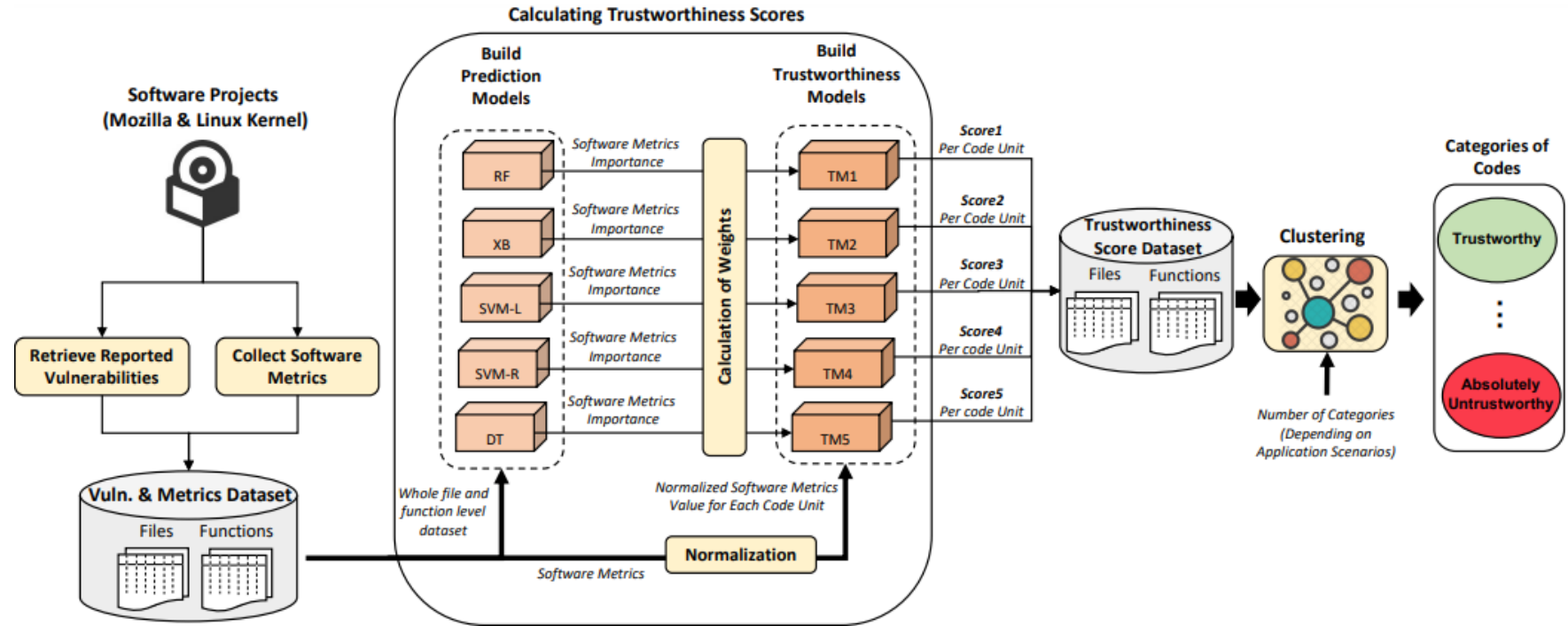


Figure 2: Methodology for approach instantiation.

Table 1: Summary of the dataset.

Software Project	# Files			# Functions		
	Total	Non vulnerable	Vulnerable	Total	Non vulnerable	Vulnerable
Linux Kernel (C)	95905	93727	2178	477342	476587	755
Mozilla Firefox (C++)	46444	45600	844	354479	353781	698

IV. RESULTS

Table 2: File-level metrics Weight.

Software Metrics	(a) Linux Kernel					(b) Mozilla Firefox			
	RF	DT	Xboost	SVM		RF	DT	Xboost	SVM
SumEssential	0,032	0,215	0,167	0,059		0,022	0,200	0,095	0,052
MaxEssential	0,030	0,219	0,120	0,058		0,022	0,000	0,000	0,045
SumCyclomaticStrict	0,023	0,181	0,065	0,054		0,020	0,000	0,000	0,051
CountStmtExe	0,021	0,168	0,029	0,053		0,018	0,000	0,000	0,049
SumCyclomatic	0,018	0,169	0,000	0,053		0,018	0,000	0,000	0,050
CountLineCodeExe	0,022	0,000	0,054	0,048		0,017	0,000	0,026	0,050
AltCountLineComment	0,025	0,000	0,051	0,044		0,021	0,200	0,058	0,050
CountLineCode	0,022	0,000	0,040	0,048		0,020	0,000	0,052	0,051
AltCountLineBlank	0,022	0,000	0,037	0,047		0,020	0,201	0,091	0,052
CountLineBlank	0,020	0,000	0,038	0,046		0,018	0,195	0,000	0,051
AvgEssential	0,024	0,027	0,000	0,049		0,015	0,000	0,000	0,000
CountLine	0,020	0,000	0,029	0,050		0,019	0,000	0,000	0,052
MaxCyclomaticModified	0,021	0,000	0,031	0,046		0,021	0,000	0,000	0,000
CountStmt	0,017	0,000	0,028	0,051		0,015	0,000	0,000	0,050
CountLinePreprocessor	0,022	0,000	0,061	0,000		0,017	0,000	0,044	0,048
MaxCyclomaticStrict	0,021	0,011	0,000	0,050		0,022	0,000	0,030	0,000
AltCountLineCode	0,022	0,000	0,000	0,050		0,018	0,000	0,029	0,052
SumCyclomaticModified	0,019	0,000	0,000	0,052		0,021	0,000	0,031	0,051
CountDeclFunction	0,026	0,000	0,000	0,044		0,020	0,000	0,030	0,047
CountLineInactive	0,024	0,000	0,045	0,000		0,027	0,000	0,079	0,000
CountSemicolon	0,019	0,000	0,000	0,050		0,014	0,000	0,031	0,050
CountLineComment	0,027	0,000	0,041	0,000		0,024	0,204	0,134	0,050
MaxCyclomatic	0,018	0,000	0,000	0,047		0,020	0,000	0,000	0,000
CountLineCodeDecl	0,026	0,000	0,040	0,000		0,016	0,000	0,026	0,051
RatioCommentToCode	0,026	0,000	0,033	0,000		0,028	0,000	0,047	0,000
CountStmtDecl	0,019	0,000	0,039	0,000		0,020	0,000	0,055	0,049
AvgLine	0,025	0,000	0,026	0,000		0,024	0,000	0,000	0,000
CountStmtEmpty	0,021	0,000	0,027	0,000		0,025	0,000	0,000	0,000
AvgCyclomaticStrict	0,016	0,010	0,000	0,000		0,015	0,000	0,000	0,000
AvgLineCode	0,024	0,000	0,000	0,000		0,023	0,000	0,028	0,000
MaxFanIn	0,023	0,000	0,000	0,000		0,019	0,000	0,000	0,000
AltAvgLineCode	0,023	0,000	0,000	0,000		0,024	0,000	0,000	0,000
AvgFanIn	0,022	0,000	0,000	0,000		0,014	0,000	0,000	0,000
MaxFanOut	0,020	0,000	0,000	0,000		0,020	0,000	0,000	0,000
CountPath	0,019	0,000	0,000	0,000		0,024	0,000	0,000	0,000
AltAvgLineComment	0,019	0,000	0,000	0,000		0,016	0,000	0,000	0,000

AvgLineBlank	0,018	0,000	0,000	0,000	0,013	0,000	0,000	0,000
AvgLineComment	0,018	0,000	0,000	0,000	0,017	0,000	0,000	0,000
HK	0,018	0,000	0,000	0,000	0,020	0,000	0,000	0,000
AltAvgLineBlank	0,018	0,000	0,000	0,000	0,013	0,000	0,000	0,000
MaxNesting	0,017	0,000	0,000	0,000	0,018	0,000	0,000	0,000
FanIn	0,016	0,000	0,000	0,000	0,023	0,000	0,030	0,000
FanOut	0,014	0,000	0,000	0,000	0,024	0,000	0,058	0,000
AvgFanOut	0,013	0,000	0,000	0,000	0,020	0,000	0,000	0,000
AvgMaxNesting	0,013	0,000	0,000	0,000	0,017	0,000	0,000	0,000
SumMaxNesting	0,013	0,000	0,000	0,000	0,017	0,000	0,000	0,000
AvgCyclomaticModified	0,012	0,000	0,000	0,000	0,013	0,000	0,000	0,000
AvgCyclomatic	0,011	0,000	0,000	0,000	0,013	0,000	0,000	0,000
MaxMaxNesting	0,011	0,000	0,000	0,000	0,014	0,000	0,000	0,000
CBO	0,010	0,000	0,000	0,000	0,034	0,000	0,000	0,000
LCOM	0,000	0,000	0,000	0,000	0,027	0,000	0,026	0,000

Table 3: Function-level Weight.

(a) Linux Kernel					(b) Mozilla Firefox			
Software Metrics	RF	DT	Xboost	SVM	RF	DT	Xboost	SVM
CountOutput	0,071	0,044	0,076	0,056	0,055	0,000	0,089	0,053
CountLineCodeDecl	0,062	0,000	0,054	0,048	0,044	0,000	0,063	0,052
MaxNesting	0,043	0,024	0,000	0,000	0,050	0,000	0,034	0,052
CountInput	0,042	0,039	0,068	0,050	0,042	0,000	0,077	0,000
AltCountLineBlank	0,040	0,004	0,050	0,049	0,045	0,207	0,065	0,055
Knots	0,040	0,000	0,047	0,051	0,025	0,000	0,026	0,040
CountLineBlank	0,039	0,005	0,000	0,049	0,052	0,209	0,086	0,055
CountLineCode	0,036	0,163	0,133	0,056	0,039	0,196	0,056	0,053
MinEssentialKnots	0,036	0,182	0,041	0,049	0,024	0,000	0,000	0,000
AltCountLineComment	0,036	0,005	0,031	0,000	0,037	0,000	0,029	0,045
MaxEssentialKnots	0,036	0,180	0,038	0,049	0,025	0,000	0,000	0,000
CyclomaticStrict	0,035	0,007	0,027	0,044	0,026	0,000	0,025	0,050
CountSemicolon	0,034	0,000	0,044	0,047	0,031	0,000	0,038	0,047
CountLineComment	0,034	0,000	0,000	0,000	0,037	0,000	0,000	0,045
CountStmtDecl	0,034	0,000	0,035	0,000	0,042	0,000	0,032	0,048
Cyclomatic	0,034	0,000	0,000	0,043	0,033	0,000	0,000	0,051
CountLine	0,033	0,156	0,059	0,056	0,042	0,196	0,103	0,055
CountLineCodeExe	0,033	0,000	0,045	0,056	0,035	0,000	0,050	0,051
CyclomaticModified	0,033	0,000	0,025	0,044	0,034	0,191	0,020	0,052
RatioCommentToCode	0,033	0,005	0,033	0,000	0,045	0,000	0,042	0,000
CountPath	0,033	0,000	0,049	0,050	0,024	0,000	0,045	0,049

AltCountLineCode	0,033	0,160	0,065	0,056	0,040	0,000	0,045	0,052
CountStmtExe	0,032	0,000	0,039	0,048	0,040	0,000	0,040	0,047
CountStmt	0,031	0,000	0,038	0,048	0,039	0,000	0,033	0,048
Essential	0,031	0,026	0,000	0,049	0,028	0,000	0,000	0,000
CountLinePreprocessor	0,021	0,000	0,000	0,000	0,007	0,000	0,000	0,000
CountLineInactive	0,019	0,000	0,000	0,000	0,006	0,000	0,000	0,000
CountStmtEmpty	0,018	0,000	0,000	0,000	0,052	0,000	0,000	0,000

Table 4: Example of the real (a) and normalized (b) values of four software metrics for five Linux Kernel's files.

(a) File-level dataset of Linux Kernel project

#	ID File	Aff.	Count Line Inactive	Sum Cyclomatic	Fan Out	HK
1	1407302	v	1	4	12	150
2	6081702	v	0	46	164	113172372
3	15866802	nv	141	113	22	13821104
4	2627502	v	7	203	270	23115237
5	32375002	nv	302	385	148	1933014887

(b) Normalized File-level dataset of Linux Kernel project

#	ID File	Aff.	Count Line Inactive	Sum Cyclomatic	Fan Out	HK
1	1407302	v	0,993	0,991	0,900	1,000
2	6081702	v	1,000	0,898	0,000	0,000
3	15866802	nv	0,000	0,750	0,817	0,126
4	2627502	v	0,948	0,550	0,000	0,000
5	32375002	nv	0,000	0,147	0,000	0,000

Table 5: Example of trustworthiness scores for five files of Linux Kernel.

#	ID File	Trustworthiness Score			
		RF	DT	Xboost	SVM
1	1407302	0,904	0,975	0,903	0,960
2	6081702	0,548	0,800	0,812	0,811
3	15866802	0,570	0,600	0,585	0,650
4	2627502	0,500	0,518	0,538	0,533
5	32375002	0,295	0,134	0,161	0,242

Table 6: Clustering of Linux Kernel files and functions.

(a) File-level results

Clusters	Score RF		Score DT		Score Xboost		Score SVM		Score RF, Score DT, Score Xboost, Score SVM				
	# Files	Medoid	# Files	Medoid	# Files	Medoid	# Files	Medoid	# Files	Med. RF	Med. DT	Med. Xboost	Med. SVM
Cluster 1	22813	0,85	28581	0,933	24769	0,883	26508	0,928	27156	0,830	0,927	0,884	0,927
Cluster 2	22793	0,758	25838	0,801	22591	0,795	25721	0,801	25098	0,738	0,815	0,759	0,802
Cluster 3	21940	0,637	18195	0,631	20851	0,67	18675	0,65	17429	0,578	0,657	0,644	0,65
Cluster 4	15651	0,478	12645	0,404	15400	0,464	13369	0,435	13842	0,478	0,449	0,475	0,478
Cluster 5	12708	0,304	10646	0,076	12294	0,182	11632	0,136	12380	0,337	0,117	0,155	0,148
Total	95905	-	95905	-	95905	-	95905	-	95905	-	-	-	-

(b) Function-level results

Clusters	Score RF		Score DT		Score Xboost		Score SVM		Score RF, Score DT, Score Xboost, Score SVM				
	# Func.	Medoid	# Func.	Medoid	# Func.	Medoid	# Func.	Medoid	# Func.	Med. RF	Med. DT	Med. Xboost	Med. SVM
Cluster 1	135767	0,885	189464	0,928	152339	0,930	173559	0,921	166562	0,876	0,935	0,923	0,924
Cluster 2	115832	0,782	101008	0,817	124464	0,798	113869	0,788	132604	0,756	0,809	0,778	0,797
Cluster 3	95082	0,639	79247	0,650	88195	0,642	78533	0,631	83694	0,571	0,566	0,570	0,583
Cluster 4	73553	0,448	62324	0,351	65060	0,405	62298	0,405	51796	0,386	0,286	0,326	0,343
Cluster 5	57108	0,161	45299	0,045	47284	0,088	49083	0,086	42686	0,133	0,043	0,120	0,084
Total	477342	-	477342	-	477342	-	477342	-	477342	-	-	-	-

Table 7: Validation of clustering results using files of Linux Kernel project.

(a) Score RF

Clusters	# Files	% Files	# vuln.	% vuln.
Cluster 1	22813	23,8%	79	0,3%
Cluster 2	22793	23,8%	282	1,2%
Cluster 3	21940	22,9%	643	2,9%
Cluster 4	15651	16,3%	855	5,5%
Cluster 5	12708	13,3%	1264	9,9%
Total	95905	-	3123	-

(b) Score DT

Clusters	# Files	% Files	# vuln.	% vuln.
Cluster 1	28581	29,8%	102	0,4%
Cluster 2	25838	26,9%	466	1,8%
Cluster 3	18195	19,0%	519	2,9%
Cluster 4	12645	13,2%	850	6,7%
Cluster 5	10646	11,1%	1186	11,1%
Total	95905	-	3123	-

(c) Score Xboost

Clusters	# Files	% Files	# vuln.	% vuln.
Cluster 1	24769	25,8%	82	0,3%
Cluster 2	22591	23,6%	279	1,2%
Cluster 3	20851	21,7%	603	2,9%
Cluster 4	15400	16,1%	816	5,3%
Cluster 5	12294	12,8%	1343	10,9%
Total	95905	-	3123	-

(d) Score SVM

Clusters	# Files	% Files	# vuln.	% vuln.
Cluster 1	26508	27,6%	66	0,2%
Cluster 2	25721	26,8%	424	1,6%
Cluster 3	18675	19,5%	592	3,2%
Cluster 4	13369	13,9%	820	6,1%
Cluster 5	11632	12,1%	1221	10,5%
Total	95905	-	3123	-

(e) Score RF, Score DT, Score Xboost, Score SVM

Clusters	# Files	% Files	# vuln.	% vuln.
Cluster 1	27156	28,3%	79	0,29%
Cluster 2	25098	26,2%	389	1,55%
Cluster 3	17429	18,2%	555	3,18%
Cluster 4	13842	14,4%	775	5,60%
Cluster 5	12380	12,9%	1325	10,70%
Total	95905	-	3123	-

Table 8: Validation of clustering results using functions of Linux Kernel project.

(a) Score RF

Clusters	# Funcs.	% Funcs.	# vuln.	% vuln.
Cluster 1	135767	28,4%	37	0,03%
Cluster 2	115832	24,3%	129	0,11%
Cluster 3	95082	19,9%	108	0,11%
Cluster 4	73553	15,4%	153	0,21%
Cluster 5	57108	12,0%	175	0,31%
Total	477342	-	602	-

(b) Score DT

Clusters	# Funcs.	% Funcs.	# vuln.	% vuln.
Cluster 1	189464	39,7%	65	0,03%
Cluster 2	79247	16,6%	108	0,14%
Cluster 3	101008	21,2%	160	0,16%
Cluster 4	62324	13,1%	108	0,17%
Cluster 5	45299	9,5%	161	0,36%
Total	477342	-	602	-

(c) Score Xboost

Clusters	# Funcs.	% Funcs.	# vuln.	% vuln.
Cluster 1	152339	31,9%	45	0,03%
Cluster 2	124464	26,1%	152	0,12%
Cluster 3	88195	18,5%	113	0,13%
Cluster 4	65060	13,6%	132	0,20%
Cluster 5	47284	9,9%	160	0,34%
Total	477342	-	602	-

(d) Score SVM

Clusters	# Funcs.	% Funcs.	# vuln.	% vuln.
Cluster 1	173559	36,4%	61	0,04%
Cluster 2	78533	16,5%	89	0,11%
Cluster 3	113869	23,9%	156	0,14%
Cluster 4	62298	13,1%	131	0,21%
Cluster 5	49083	10,3%	165	0,34%
Total	477342	-	602	-

(e) Score RF, Score DT, Score Xboost, Score SVM

Clusters	# Funcs.	% Funcs.	# vuln.	% vuln.
Cluster 1	166562	34,9%	57	0,03%
Cluster 2	132604	27,8%	170	0,13%
Cluster 3	83694	17,5%	127	0,15%
Cluster 4	51796	10,9%	94	0,18%
Cluster 5	42686	8,9%	154	0,36%
Total	477342	-	602	-

Table 9: Clustering results of Mozilla Firefox project.

(a) File-level results

Clusters	Score RF		Score DT		Score Xboost		Score SVM		Score RF, Score DT, Score Xboost, Score SVM				
	# Files	Medoid	# Files	Medoid	# Files	Medoid	# Files	Medoid	# Files	Med. RF	Med. DT	Med. Xboost	Med. SVM
Cluster 1	14094	0,834	19037	0,932	16220	0,886	16185	0,949	15580	0,834	0,939	0,893	0,948
Cluster 2	10991	0,724	10314	0,804	11638	0,778	12069	0,833	11595	0,716	0,848	0,795	0,827
Cluster 3	9068	0,622	6661	0,635	7386	0,640	7270	0,646	8344	0,615	0,689	0,640	0,686
Cluster 4	7404	0,455	4928	0,367	5568	0,445	5014	0,417	5336	0,455	0,342	0,447	0,380
Cluster 5	4887	0,256	5504	0,000	5632	0,148	5906	0,038	5589	0,307	0,000	0,147	0,043
Total	46444	-	46444	-	46444	-	46444	-	46444	-	-	-	-

(b) Function-level results

Clusters	Score RF		Score DT		Score Xboost		Score SVM		Score RF, Score DT, Score Xboost, Score SVM				
	# Func.	Medoid	# Func.	Medoid	# Func.	Medoid	# Func.	Medoid	# Func.	Med. RF	Med. DT	Med. Xboost	Med. SVM
Cluster 1	122468	0,833	134553	0,954	139925	0,884	131761	0,940	122019	0,844	0,954	0,900	0,949
Cluster 2	86253	0,748	78150	0,841	80185	0,755	83875	0,819	80506	0,750	0,853	0,788	0,829
Cluster 3	59840	0,580	60635	0,663	54522	0,609	56223	0,645	54447	0,630	0,709	0,651	0,682
Cluster 4	42818	0,389	40298	0,369	37772	0,374	40862	0,403	48038	0,424	0,507	0,462	0,516
Cluster 5	43101	0,092	40844	0,000	42076	0,101	41759	0,067	49470	0,112	0,036	0,133	0,114
Total	354480	-	354480	-	354480	-	354480	-	354480	-	-	-	-

Table 10: Validation of clustering Results using Mozilla Firefox files and functions.

ID File	Exp. based results	Average Medoids	ID Function	Exp. based results	Average Medoids
60426601	0,429	0,912	361442101	0,462	0,919
263601	0,408	0,912	20082901	0,458	0,919
1701201	0,31	0,912	304227601	0,272	0,919
51355601	0,233	0,912	121633001	0,303	0,810
52292101	0,165	0,803	311237201	0,153	0,671
1418301	0,171	0,663	113275001	0,082	0,671
2201801	0,103	0,663	119075401	0,094	0,485
45301201	0,07	0,108	314376601	0,057	0,485
22256501	0,065	0,108	328555501	0,055	0,485
19960501	0,046	0,108	53225001	0,063	0,102

Table 11: Results using different number of clusters.

(a) File-level					(b) Function-level			
Number of clusters = 3					Number of clusters = 3			
Clusters	# Files	% Files	# vuln	% vuln	# Func.	% Func.	# vuln	% vuln
Cluster 1	43697	45,6%	288	0,7%	165580	34,7%	79	0,0%
Cluster 2	33324	34,7%	1060	3,2%	138389	29,0%	75	0,1%
Cluster 3	18884	19,7%	1775	9,4%	173373	36,3%	448	0,3%
Total	95905	-	3123	-	477342	-	602	-
Number of clusters = 5					Number of clusters = 5			
Clusters	# Files	% Files	# vuln	% vuln	# Func.	% Func.	# vuln	% vuln
Cluster 1	27156	28,3%	79	0,29%	166562	34,9%	57	0,03%
Cluster 2	25098	26,2%	389	1,55%	132604	27,8%	170	0,13%
Cluster 3	17429	18,2%	555	3,18%	83694	17,5%	127	0,15%
Cluster 4	13842	14,4%	775	5,60%	51796	10,9%	94	0,18%
Cluster 5	12380	12,9%	1325	10,70%	42686	8,9%	154	0,36%
Total	95905	-	3123	-	477342	-	602	-
Number of clusters = 7					Number of clusters = 7			
Clusters	# Files	% Files	# vuln	% vuln	# Func.	% Func.	# vuln	% vuln
Cluster 1	19190	20,0%	46	0,2%	82066	17,2%	30	0,0%
Cluster 2	18751	19,6%	168	0,90%	45003	9,4%	22	0,0%
Cluster 3	19002	19,8%	397	2,09%	72538	15,2%	37	0,1%
Cluster 4	13906	14,5%	459	3,3%	54978	11,5%	30	0,1%
Cluster 5	9195	9,6%	459	4,99%	63074	13,2%	35	0,1%
Cluster 6	8367	8,7%	722	8,63%	80528	16,9%	166	0,2%
Cluster 7	7494	7,8%	872	11,6%	79155	16,6%	282	0,4%
Total	95905	-	3123	-	477342	-	602	-