

Social Interactions around Cross-System Bug Fixings: The Case of FreeBSD and OpenBSD



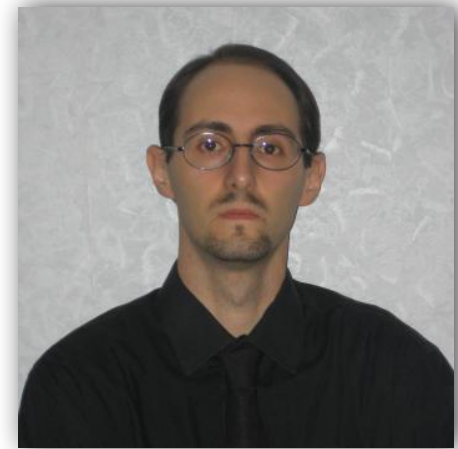
Gerardo Canfora
University of
Sannio



Luigi Cerulo
Unitelma
Sapienza



Marta Cimitile
University of
Sannio



Massimiliano Di
Penta University of
Sannio

Cross-System Bug Fixings (CSBFs)

- 他のシステムで使われたパッチを他のシステムに再利用すること
 - Ex. NeoOfficeがOpenOfficeのパッチを再利用

Use OpenOffice.org patch from the following URL to prevent regcomp from crashing on PowerPC

Fix bug 3434 by using newer version of this OOo file obtained from OOo's svn

背景

- クローンを変更する際は、その変更をクローンしたソフトウェアに適切に普及させる必要性がある
- 一つのソフトウェアシステム内であれば、クローンの変更を上手く管理することができることがわかっている
- CSBFsについての研究はあまりされていない

目的とアプローチ

- 目的
 - FreeBSD-OpenBSD間のCSBFsの活動を認識すること
 - CSBFを行なっている開発者の社会的役割を理解すること
- アプローチ
 - コミットノートからCSBFsをマイニングする手法を提案
 - 開発者コミュニティに対してソーシャルネットワーク分析を行う

リサーチクエスチョン

- RQ1
 - FreeBSD と OpenBSD で共通しているコミッター/メール貢献者はどれぐらいいるのか？
- RQ2
 - CSBFs はどれほどの頻度で起こるのか？
- RQ3
 - CSBFs に関係している開発者はだれか？
- RQ4
 - CSBFs に関係しているメール貢献者は、他の開発者よりも活動しているか？

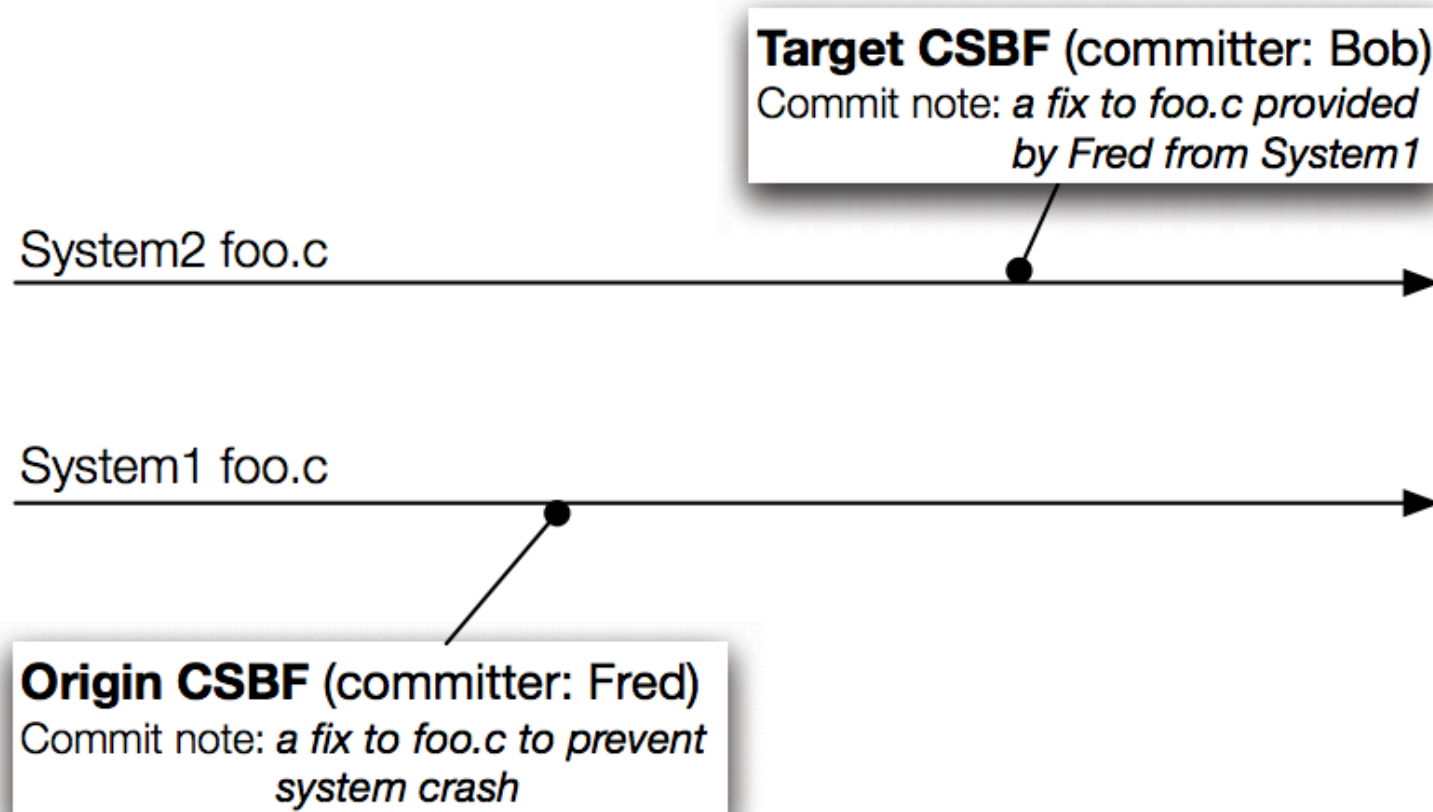
CSBFコミットの検出

1. 各プロジェクトのコミット情報を取得
 - 変更されたファイル名, 日時, 変更行数
 - コミッターID, コミットノート
2. 他方のシステムを参照しているコミットノートをフィルター
 - 表記ゆれによるfalse positiveを制限するため
 - Ex. "Inspired from FreeBSD" - "From FreeBSD"
3. コードクローン検出手法を用いてCSBFを検出

CSBFコミットの検出

1. 両プロジェクトで共通する名前のファイルを取得
2. CCFinderを用いて共通するコードをチェック
 - 最低10%のクローンがあれば共通とみなす
3. 過去のリビジョンから変更があるかを調べる
4. 本文の類似度を調べる
 - tf法を適応（閾値0.2）
5. 前のリビジョンのものをOrigin CSBF,後のリビジョンのものをTarget CSBFとする

CSBFコミットの検出



開発者/メール貢献者の特定

- 名寄せ：Birdらのアプローチを参考に行う[1]
 - 「.」,ミドルネームを削除
 - 完全一位した名前はマージ
 - ファーストネームまたはラストネームのイニシャルが一致したらマージ
 - ユーザーIDの一致
- メールアドレスのマージ
 - @以下を削除
 - 名寄せの結果を利用

[1]C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan. Mining email social networks. In Proceedings of the 2006 international workshop on Mining software repositories, MSR '06, pages 137–143

SNA分析

- メーリングリスト内でメール交換を行なっている場合, ノードを繋げる
- 用いるメトリクス
 - Degree:繋がっているノードの数
 - Betweenness:全てのパス中のあるノードを通る割合
 - Brokerage metrics
 - Coordinator:同じグループ内のメールを仲介
 - Gatekeeper:異なるグループのメールを同じグループに仲介
 - Representative:異なるグループへメールを仲介

データセット

- 対象プロジェクト

Table 1: System history characteristics

	FreeBSD	OpenBSD
Time range	1993–2009	1998–2009
KLOC (min–max)	76–3,298	1,268–2,298
# of Source files (min–max)	211–6,797	4,434–6,440
# of Commits	119,259	70,895
# of Change sets	45,979	27,939
# of Mailing list messages	66,352	22,518

- データソース
 - CVS
 - Bug-fixingメーリングリスト

RQ1:結果

- FreeBSDとOpenBSDで共通しているコミッター/メール貢献者はどれぐらいいるのか？

	FreeBSD	OpenBSD	Both
# of committers	383	211	26
# of mailing list contribs	8,035	3,843	359
# of committers $\cap$ mailing list contribs	213	122	17

両プロジェクトに共通した
コミッター/メール貢献は10%に満たないほどだった

RQ2:結果

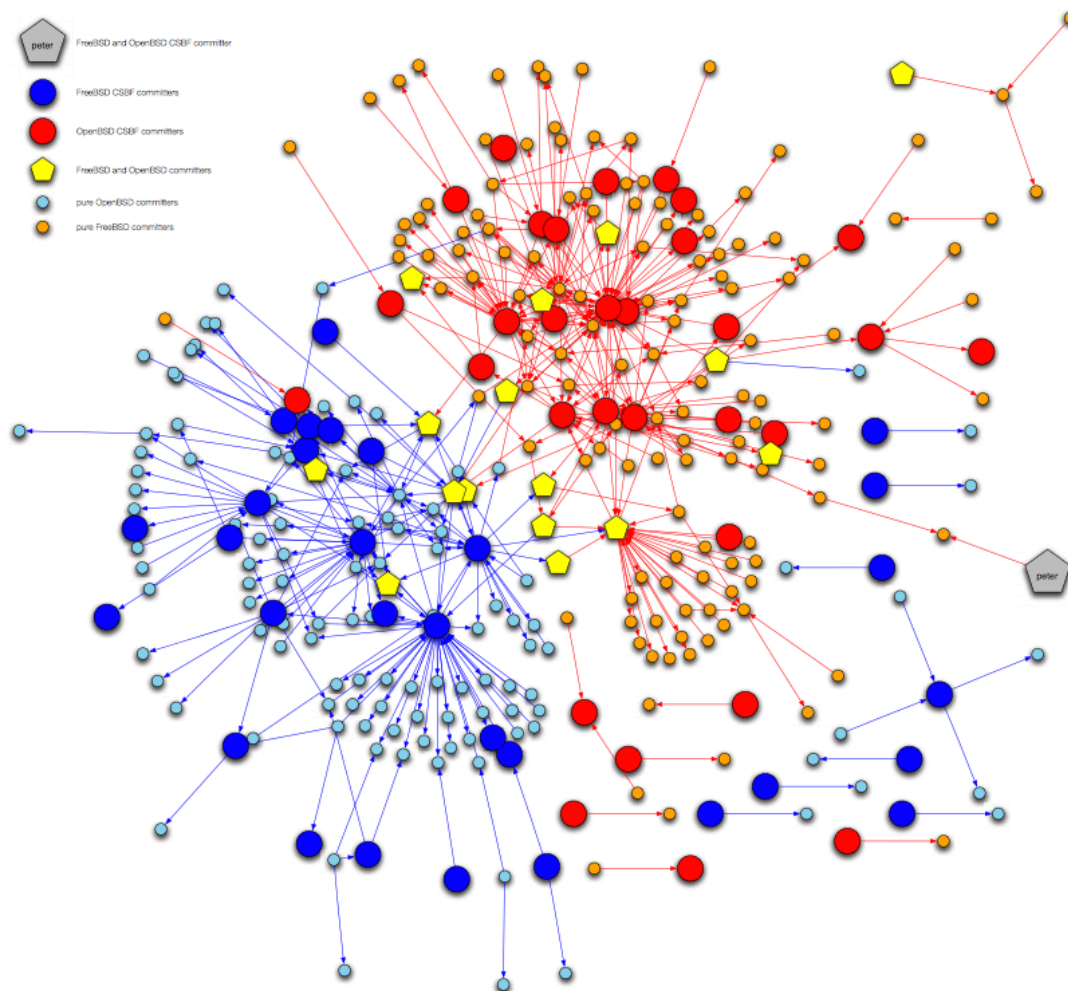
- CSBFsはどれほどの頻度で起こるのか？

	FreeBSD	OpenBSD
# of referring commits	439	933
# of referring change sets	200	476
# of referring commits between cloned files	133	296
# of linked commits (total)	59	161
# of linked commits (.c files)	37	93
# of linked commits (.h files)	22	68
# of linked change sets	46	120

両プロジェクト共に
わずかな頻度で起こっているとは言えない

RQ3:結果

CSBFsに関係している開発者はだれか？



RQ3:結果

CSBFsに関係している開発者はだれか？

Metric	Contributors involved in CSBFs			Other contributors			<i>p</i> -value	Cohen <i>d</i>
	Mean	Median	σ	Mean	Median	σ		
Degree	22.52	5.00	39.47	2.47	1.00	7.26	< 0.01	1.79
In degree	9.42	2.00	16.72	1.31	1.00	3.46	< 0.01	1.63
Out degree	13.10	2.50	23.46	1.16	1.00	4.38	< 0.01	1.78
Betweenness	20,248.50	1,938.90	45,880.18	1,220.75	< 0.01	9,363.52	< 0.01	1.42
Coordinator	455.92	4.50	1,481.99	8.33	0.00	186.66	< 0.01	1.26
Gatekeeper	11.45	0.00	34.22	0.28	0.00	5.06	< 0.01	1.30
Representative	8.61	0.00	37.03	0.18	0.00	3.72	< 0.01	1.02

CSBFsに関係している貢献者は
プロジェクト内・プロジェクト間のコミュニケーション
の重要な役割を担っている

RQ3:結果

CSBFsに関係している開発者はだれか？

Degree	In Degree	Out Degree	Betweenness	Coordinator	Gatekeeper	Representative
todd c miller	remko lodder	todd c miller	poul henning	poul henning	poul henning	otto moerbeek
otto moerbeek	poul henning	otto moerbeek	john mark gurney	otto moerbeek	todd c miller	poul henning
poul henning	otto moerbeek	poul henning	kenneth r westerback	todd c miller	henning brauer	todd c miller
remko lodder	todd c miller	miod vallat	todd c miller	bill fenner	bill fenner	henning brauer
bill fenner	bill fenner	henning brauer	remko lodder	miod vallat	remko lodder	theo de raadt
miod vallat	robert watson	bill fenner	otto moerbeek	henning brauer	otto moerbeek	brooks davis
henning brauer	miod vallat	remko lodder	henning brauer	mark kettenis	miod vallat	john mark gurney
mark kettenis	henning brauer	mark kettenis	bill fenner	gary palmer	robert watson	robert watson
gary palmer	mark murray	brooks davis	miod vallat	robert watson	angelos d keromytis	alex
robert watson	bill paul	gary palmer	theo de raadt	mark murray	kenneth r westerback	reyk floeter

各メトリクスの上位10名のほとんどはCSBFsの貢献者

RQ4:結果

- CSBFsに関係しているメール貢献者は、他の開発者よりも活動しているか？

Project	Metric	Contributors involved in CSBFs			Other contributors			<i>p</i> -value	Cohen <i>d</i>
		Mean	Median	σ	Mean	Median	σ		
FreeBSD	# of Commits	1,111.55	408.00	1,827.39	135.32	25.50	287.25	< 0.01	1.06
	LOC add/del	35,662.18	15,904.00	48,327.41	4,351.08	380.00	10,266.49	< 0.01	1.20
OpenBSD	# of Commits	1,102.71	404.50	1,987.68	86.85	22.50	172.31	< 0.01	0.93
	LOC add/del	30,170.52	13,082.00	47,175.23	3,534.24	433.00	11,448.81	< 0.01	0.95

CSBFsに関係のあるメール貢献者の方が他の開発者よりも活動量が多い

まとめと所感

- まとめ
 - FreeBSDとOpenBSD間のCSBFsについて分析
 - CSBFsはわずかであるとは言えない
 - CSBFsに関係しているコミッターはプロジェクト内・プロジェクト間において重要な役割を担っている
- 所感
 - 結果がきれいにでたなあ
 - 他のプロジェクト間のCSBFはどうなんだろう？

妥当性の評価

- 内的妥当性
 - CSBFの貢献者の役割とSNAのメトリクスとの関係は明らかにできていない（他の要因があるかも...）
- 外的妥当性
 - FreeBSD-OpenBSD以外のケースについては明らかにできていないため、他のケースでも同じ結果が得られるかはわからない