OSS におけるセキュリティパターンの特定

早稲田大学 山本美聡, 鷲崎弘宜, 深澤良彰, 大久保隆夫, 海谷治彦, 吉岡信和

ソフトウェア開発において,どのようなセキュリティパターンがどこに使用されたのかは明確に記述されておらず,また開発者が独自の手法で適用している場合が多い．そのため第三者によるセキュリティパターンやその使用箇所を特定する事が難しく,セキュリティパターンとそれによって改善される脅威・脆弱性を関連づけることが難しい．そこで我々はオープンソースソフトウェアに適用されているセキュリティパターンを特定するため,まず各セキュリティパターンの文書を,TF-IDF 値を算出する事で重要単語により特徴付けた．この特徴をもってオープンソースソフトウェアの関連文書からパターンの使用を特定する．特定したセキュリティパターンと脆弱性データベースの情報を関連づけることで,エンピリカルな手法によりセキュアなソフトウェアの開発に役立てる．

1. 背景

ソフトウェア開発において,その関連文書にはセキュリティ項目についての言及が少なく適応されている手法が明確でないため,セキュリティパターンの適用とそれによって改善される脅威・脆弱性との関連付けが難しい．多くの OSS からそのデータを得る事が出来れば,エンピリカルな手法をもってよりセキュアな開発に役立てることが出来る．そのために各セキュリティパターンを特徴付ける事が必要となる．

2. 提案手法

本稿の Research Question を以下に示す.

RQ 各セキュリティパターンを特徴付ける.

我々は各セキュリティパターンの文書から TF-IDF 値により重要単語を算出し，その上で冠詞・代名詞・前置詞・副詞を削除した．これにより得られた 10 個の重要単語でパターンを特徴付けることで上記の RQ に応える. セキュリティパターンは[1]による定義に基づき,文書についても[1]を引用した．また TF-IDF 値の算出に際し使用したツールは自身で作成した.

3. 結果

上記の手法により得た結果のうち Execution Domain の結果を表 1 に示す. 表 1 はパターンの重要単語 10 個と，その TF-IDF 値を示している．Execution Domain の場合,domain や right 等重要と思われる単語が入っており,上位に特徴づけられる execution が入っている.

	単語	TF-IDF 値
1	execution	0.601771924
2	domain	0.582438609
3	nested	0.54161004
4	collect	0.4513417
5	resources	0.356206479
6	process	0.337117835
7	right	0.331688647
8	set	0.325771596
9	access	0.280923979
10	implement	0.246186382

表 1 Execution Domain の重要単語と TF-IDF 値

4. 結語

この調査によりセキュリティパターンを,自然言語による単語で特徴付けた．ソフトウェアの自然言語による関連文書等でこれらの特徴が見られるとき,特徴付けられたセキュリティパターンが使用されている事が期待出来る．今後はこの結果の妥当性を検証した上で,オープンソースソフトウェアに適用されているセキュリティパターンを特定する．その脆弱性データベースの情報から,各セキュリティパターンにより改善される脆弱性やバグを検証する.

5. 参考文献

[1] Christopher Nagappan, Ramesh Lai, Ray Steel：Core Security Patterns: Best Practices and Strategies for J2EE™, Web Services, and Identity Management (Sun Core Series)(200