

第 1 章

提案手法

本章では ReDoS 脆弱な正規表現の修正方法を述べる。

まず、REMEDY は、文字列例を正しく分類し、RWS1U を満たし、入力 of 正規表現に構文的に近い正規表現を出力として生成するのだった。ここで出力の正規表現は少なくとも入力として与えられた文字列例だけは正しく分類する能力はあるが、その他の文字列については入力の正規表現と同じように分類するかは分からないことに注意されたい。つまり、REMEDY の入力の正規表現と出力の正規表現は意味的に等価でない可能性がある。

正規表現を意味的に等しくかつ非脆弱なものへと修正する方法は既に存在する。[1] では、正規表現と等価な NFA を構成、そして DFA に変換し、その DFA を等価で strongly unambiguous な正規表現に戻すという手法を用いている。しかし、この手法は [1] でも言及されているように、修正後の正規表現のサイズが入力の正規表現のサイズに対して二重指数関数的になる場合もある。正規表現のサイズは文字列のパターンマッチングの処理時間に大きく関係するので、この手法は計算量の観点から望ましくない。

従って、正規表現のサイズをできるだけ小さく保ったまま非脆弱化すべく、以下のような手法を考案した。

参考文献

- [1] Brink van der Merwe, Nicolaas Weideman, and Martin Berglund. Turning evil regexes harmless. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists*, SAICSIT '17, New York, NY, USA, 2017. Association for Computing Machinery.