博士請求論文審査要旨

情報セキュリティ大学院大学
情報セキュリティ研究科

論文題目　：ファイルフォーマットからの逸脱に着目した悪性文書ファイル検知方式の研究
申請者　：大坪　雄平
審査委員会　：主査　教授　田中 英彦
副査　教授　後藤　厚宏
副査　教授　湯淺　墾道
副査　教授　大久保隆夫
副査　准教授　橋本　正樹

Ⅰ. 論文内容の要旨

本論文は,「ファイルフォーマットからの逸脱に着目した悪性文書ファイル検知方式の研究」と題し，9 章と付録からなる．近年、特定の組織や個人を狙って情報摂取等をおこなう標的型攻撃の脅威が顕在化している．このような被害を抑止し情報システムのセキュリティを向上させることを目的に, メールに添付された悪性文書ファイルを研究対象とし，それが実行される前に高確率で検出を行う方式を提案したものである．

第 1 章は,「序論」で，本研究の背景と研究の目的，及び論文の構成についてまとめている．

第 2 章は「関連研究」で，メールに添付された悪性文書ファイルの検知を行う既存の研究についてまとめたもので，パターンマッチングを用いる方式，攻撃実行の最初に使われる閲覧ソフトの脆弱性を突く exploit と呼ばれるコードの検知手法，exploit 後にその端末の制御を奪う部分の検知手法，更に、悪性活動を行う実行コードファイルの検知手法, そして文書ファイルの構造に着目した検知手法の 5 つに分類し, それぞれの概要と、その問題点について分析している．その結果，それらは，毎年出現する様々な脆弱性個々への対応や，検出を阻害する難読化への対応, 更に学習を用いる場合はその学習用サンプル収集課題の存在, 閲覧ソフトが表示する内容と文書ファイルとの関係の流動性，等の限界があることを述べている．

第 3 章は「悪性文書ファイルに対する予備調査」で，標的型攻撃に使用された実際の悪性文書ファイルを対象に，その傾向について分析している．対象は，日本で出現した標的型攻撃から採取した悪性文書ファイル tar(09-12)と，世界で公開されているマルウエアダンプサイトから取得した悪性文書ファイルで，それらの分析から，多くの文書ファイルが RTF, CFB, 及び PDF の 3 種であること，標的型攻撃に用いられる悪性文書ファイルは殆どが，その中に実行コードを含む dropper であり，標的型攻撃以外に用いられる悪性文書ファイルは殆どが, その中に実行コードを含まず他所からコードを読み込む downloader であることを示している．また，dropper に埋め込まれた実行ファイルは，多くの場合難読化を目的にエンコードされているが，そのエンコード方式を分析し，換字による方式，転置による方式，ファイル形式固有の方式，等を明らかにするとともに，その解読手法を与え実行ファイルの検知方式を提案し，それを実装して評価している．その結果，合計 370 個の検体の内，実行ファイルが含まれていた検体は 362 個であり，この内の 96.1%を検知することに成功している．

第 4 章は「ファイルフォーマットからの逸脱に着目した検知方式の提案」で，まず悪性文書ファイルの構造と閲覧ソフトの動作について，具体的な一つの検体を例に示し，その分析から，閲覧ソフトが悪性文書ファイルを処理するためには, そのファイルがある程度ファイルフォーマットの仕様に沿ったファイルである必要が

あり，それが攻撃者に対する制約となるので，逆に検知が可能となり得ることを述べている．しかしながら，この点に関する既存の研究では学習を用いたものが多く，その場合は学習サンプルを集めることが必要であるが，その中に当該組織の情報や機微な情報が含まれることがあり，情報共有が困難で，実際上の実現は難しいと述べている．一方，文書ファイル tar(09-12)を分析すると，文書ファイル仕様から逸脱した構造を持つ悪性文書ファイルが殆どであり，RTF では 99.0%，CFB では 99.0%，PDF では 99.4%がそうであることを示している．その結果，その逸脱構造が悪性文書ファイル特有のものであれば，それをキーに悪性文書ファイルの検出が可能となると述べその方式提案を行っている．

第 5 章は「悪性文書ファイルの仕様からの逸脱と検知」で，対象とした 3 種の文書ファイル RTF, CFB, PDF それぞれのファイル仕様の概要を述べ，実行コードを埋め込むために使われる仕様からの逸脱を分類して，ファイルの終端やサイズの異常，管理不可能領域の利用等をベースに，RTF では 1 種類，CFB では 4 種類，PDF では 3 種類の逸脱構造を明らかにしている．

第 6 章は「検知ツール o-checker の開発」で，5 章の逸脱構造を検知するため作成した検知ツール o-checker の実装について述べている．プログラミング言語 Python を用いて実装し合計 2,588 行のプログラムで，公開サイトに載せている．

第 7 章は「実験」で，様々なデータセットを用いて o-checker を動かした結果を述べている．まず実験 1 は dropper に対する o-checker の検知率評価であって，検体に tar(09-12)以降の新しい検体を用い，98.4%の検知率であった．次に，無害な文書ファイルに対する誤検知率の評価実験 2 では，公開されている研究用の無害なファイルを対象として 0.3%の誤検知率を示し，実験 3 では一般の悪性文書ファイルに対する o-checker の検知率評価を行い，VirusTotal から収集した検体に対し 52.9%の検知率を示している．更に，実験 4 は downloader に対する o-checker の検知率を調べたもので，マルウエアダンプサイトから入手した検体を用いた結果では 0.9%の検知率であったと述べている。

第 8 章は「検討」で，7 章の結果を分析している．その結果，実験 2 の誤検知の原因が破損した文書ファイルであり，dropper の殆どが逸脱構造を含む文書ファイルであって，逆に逸脱構造を含む文書ファイルは殆どが悪性文書であった．更にこの章では，提案方式の効果を評価し，ファイル構造は一般に継続的に使われ実行コードの埋め込みはその構造の逸脱を引き起こすことが多く，結果として本方式は長期に有効であろうと述べている．また同時に，この方式の限界について触れるとともに，今後に残された課題について述べている．

第 9 章は，「結論」で，本論文の提案についてまとめている．

付録は，　参考文献と，この研究を外部発表したリストをまとめたものである．

Ⅱ. 論文審査結果の要旨

本論文は，標的型攻撃で用いられる添付文書ファイルが悪性の文書ファイルであるか否かを，文書ファイルの仕様逸脱チェックにより非常に効率良く検知する手法を提案し，その効果を様々な検体に適用して評価したもので，今後とも効果が期待でき，今後の情報システムのセキュリティ向上と情報学に貢献するところが少なくない．

よって，本論文は，博士（情報学）の論文として合格と認められる．

Ⅲ. 審査経過

本審査委員会は，平成 28 年 7 月 27 日に論文内容とこれに関連する事項について口述試問を行い，博士論文として合格と判断した．

情報セキュリティ大学院大学　情報セキュリティ研究科