

明 細 書

発明の名称：

情報処理プログラム、情報処理方法および情報処理装置

技術分野

[0001] 本発明は、情報処理プログラム、情報処理方法および情報処理装置に関する。

背景技術

[0002] インターネットに代表されるネットワークの普及により様々な情報が電子化されネットワークを通じて通信される状況下において、ネットワーク脅威に対するセキュリティの重要性が高まっている。

[0003] セキュリティ関連の技術としては、例えば、通信ログを記録可能なアプリケーションが出力した分析対象ログファイルにもとづいて不正アクセスの有無を判断する技術が提案されている。また、アクセスの重要度を示す指標値およびアクセスが不正である可能性の高さを示す指標値を導出してアクセスの危険度を判定する技術が提案されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2002-318734号公報

特許文献2：特開2018-041316号公報

発明の概要

発明が解決しようとする課題

[0005] ネットワークセキュリティにおいて例えば、サイバー攻撃を受けたような場合、損失を最小限にするためには攻撃全容を素早くつかむことが重要である。攻撃者が使用したマルウェア（不正かつ有害に動作させる意図で作成された悪意のあるソフトウェア）を早い段階で明らかにできれば攻撃への迅速な対処が可能になる。このため通信の危険度に応じたマルウェアに関連するデータを効率よく抽出することが可能な技術が要望されている。

- [0006] 1つの側面では、本発明は、通信の危険度に応じたデータを効率よく抽出することが可能な情報処理プログラム、情報処理方法および情報処理装置を提供することを目的とする。

課題を解決するための手段

- [0007] 上記課題を解決するために、情報処理プログラムが提供される。情報処理プログラムは、コンピュータに、通信データをセッション毎に分割して分割データ群を生成し、分割データ群のうちから、同一の識別情報を有し、かつセッション間隔が閾値以下である分割データを抽出し、抽出した分割データを連結して連結データを生成し、連結データに含まれる特定情報にもとづいて通信の危険度を決定し、連結データに含まれるファイルから、危険度にもとづくファイル情報を収集する処理を実行させる。

- [0008] また、上記課題を解決するために、コンピュータが上記情報処理プログラムと同様の制御を実行する情報処理方法が提供される。

さらに、上記課題を解決するために、上記情報処理プログラムと同様の制御を実行する情報処理装置が提供される。

発明の効果

- [0009] 1側面によれば、通信の危険度に応じたデータを効率よく抽出することができる。

本発明の上記および他の目的、特徴および利点は本発明の例として好ましい実施の形態を表す添付の図面と関連した以下の説明により明らかになるであろう。

図面の簡単な説明

- [0010] [図1]第1の実施の形態の情報処理装置の一例を説明するための図である。
[図2]第2の実施の形態の情報処理システムの一例を示す図である。
[図3]サーバ装置の機能ブロックの一例を示す図である。
[図4]サーバ装置のハードウェア構成の一例を示す図である。
[図5]コマンドと危険度レベルとの対応関係を示すテーブル情報の一例を示す図である。

[図6]危険度レベルとファイル情報との対応関係を示すテーブル情報の一例を示す図である。

[図7]通信ログをセッション分割してからファイル情報を収集するまでの動作の一例を説明するための図である。

[図8]通信ログをセッション分割してからファイル情報を収集するまでの動作の一例を説明するための図である。

[図9]通信ログの一例を示す図である。

[図10]分割セッションデータの一例を示す図である。

[図11]セッション間隔の一例を説明するための図である。

[図12]疑似セッションデータの生成の一例を示す図である。

[図13]コマンダー危険度レベル対応テーブルの一例を示す図である。

[図14]危険度レベルーファイル情報対応テーブルの一例を示す図である。

[図15]危険度レベルの決定の一例を示す図である。

[図16]ファイル情報の収集の一例を示す図である。

[図17]疑似セッションデータの生成動作の一例を示すフローチャートである。

[図18]危険度レベルにもとづくファイル情報収集の動作の一例を示すフローチャートである。

[図19]危険度レベルにもとづくファイル情報収集の動作の一例を示すフローチャートである。

発明を実施するための形態

[0011] 以下、本実施の形態について図面を参照して説明する。

[第1の実施の形態]

第1の実施の形態について図1を用いて説明する。図1は第1の実施の形態の情報処理装置の一例を説明するための図である。情報処理装置1は、制御部1aおよび記憶部1bを備える。

[0012] 制御部1aは、通信データをセッション毎に分割して分割データ群を生成し、分割データ群のうちから、同一の識別情報を有し、かつセッション間隔

が閾値以下である分割データを抽出する。また、制御部 1 a は、抽出した分割データを連結して連結データを生成し、連結データに含まれる特定情報にもとづいて通信の危険度を決定する。そして、制御部 1 a は、連結データに含まれるファイルから、危険度にもとづくファイル情報を収集する。

[0013] 記憶部 1 b は、通信データ、特定情報と危険度との対応関係、および収集されたファイル情報等を格納する。なお、制御部 1 a の機能は、情報処理装置 1 が備える図示しないプロセッサが、所定のプログラムを実行することによって実現される。

[0014] 図 1 に示す例を用いて動作について説明する。

〔ステップ S 1〕 制御部 1 a は、通信データ D 0 を取得する。通信データ D 0 は、セッション I D (Identity)、コマンド、識別情報およびオプションを有している。識別情報は、例えば、アカウント情報や、ユーザ認証に用いられるクレデンシャル情報（ユーザ I D やパスワード等）に相当する。

[0015] 通信データ D 0 において、コマンド c d 1、c d 2 によるセッションは、セッション I D が 1 であり、コマンド c d 3 によるセッションは、セッション I D が 2 であり、コマンド c d 4 によるセッションは、セッション I D が 3 になっている。また、セッション I D が 1、2 のセッションは、識別情報が A であり、コマンド c d 2 により 1 MB のファイル書き込みが行われたことが示されている。セッション I D が 3 のセッションは、識別情報が B になっている。

[0016] 〔ステップ S 2〕 制御部 1 a は、通信データ D 0 をセッション毎に分割して、分割データ D 1、D 2、D 3 が含まれる分割データ群 D g を生成する。

〔ステップ S 3〕 制御部 1 a は、分割データ群 D g のうちから、同一の識別情報を有し、かつセッション間隔が閾値以下である分割データを抽出する。分割データ D 1、D 2 は識別情報が A で同一である。また、分割データ D 1、D 2 のセッション間隔は閾値以下であるとする（セッション間隔については後述）。この場合、分割データ群 D g のうちから、分割データ D 1、D 2 が抽出される。そして、制御部 1 a は、抽出した分割データ D 1、D 2 を

連結して連結データDLを生成する。

[0017] 〔ステップS4〕制御部1aは、連結データDLに含まれる特定情報にもとづいて通信の危険度を決定する。特定情報は、連結データDLに含まれるコマンド、および連結データDLに含まれるファイルのサイズである。

[0018] よって、制御部1aは、コマンドおよびファイルのサイズにもとづいて危険度を決定する。例えば、コマンドcd1、・・・、cd4のうちでコマンドcd3による操作の危険度が最も高く、ファイルサイズの閾値を3MBとすると、連結データDLには、コマンドcd3が含まれ、また連結データDLに含まれるファイルのサイズ1MBは閾値以下である。したがって、この場合、連結データDLは、危険度が最も高い値に決定される。

[0019] 〔ステップS5〕制御部1aは、ステップS4で決定された危険度にもとづいて、連結データに含まれるファイルからファイル情報を収集する。ファイル情報としては、ファイルのメタ情報、ファイルハッシュ値およびファイル本体がある。

[0020] 制御部1aは、危険度のレベルにもとづいて、連結データに含まれるファイルから、メタ情報、ファイルハッシュ値およびファイル本体の少なくとも1つを適応的に収集する。例えば、危険度が最も高い値に決定された連結データDLからは、ファイル情報として、メタ情報、ファイルハッシュ値およびファイル本体が収集される。

[0021] このように、情報処理装置1では、通信データをセッション毎に分割し、同一識別情報を有しセッション間隔が閾値以下の分割セッションを連結し、連結セッションから通信の危険度にもとづいてファイル情報を収集する。これにより、危険度に応じたデータを効率よく抽出することができる。

[0022] 〔第2の実施の形態〕

次に第2の実施の形態について説明する。図2は第2の実施の形態の情報処理システムの一例を示す図である。情報処理システム1-1は、サーバ装置10、スイッチsw1、sw2およびユーザ端末3a、3b、3c、4a、4b、4cを備える。サーバ装置10は、図1の情報処理装置1の機能を

実現する。

[0023] A 拠点には、サーバ装置 10、スイッチ s w 1 およびユーザ端末 3 a、3 b、3 c が含まれ、B 拠点には、スイッチ s w 2 およびユーザ端末 4 a、4 b、4 c が含まれる。なお、A 拠点および B 拠点は図示しないネットワークで接続される。

[0024] サーバ装置 10 は、ユーザ端末 3 a、3 b、3 c からユーザ端末 4 a、4 b、4 c へ向かう通信データに対して通信の危険度レベルを決定して、通信データに含まれるファイルから所定のファイル情報を収集する。またはサーバ装置 10 は、ユーザ端末 4 a、4 b、4 c からユーザ端末 3 a、3 b、3 c へ向かう通信データに対して通信の危険度レベルを決定して、通信データに含まれるファイルから所定のファイル情報を収集する。

[0025] <機能ブロック>

図 3 はサーバ装置の機能ブロックの一例を示す図である。サーバ装置 10 は、制御部 11 および記憶部 12 を備える。制御部 11 は、図 1 の制御部 1 a の機能を実現し、記憶部 12 は、図 1 の記憶部 1 b の機能を実現する。

[0026] 制御部 11 は、通信インタフェース部 11 a、通信ログ生成部 11 b、セッション分割部 11 c、疑似セッション生成部 11 d、危険度レベル決定部 11 e およびファイル情報収集部 11 f を含む。

[0027] 通信インタフェース部 11 a は、サーバ装置 10 に接続されるネットワークを介した通信インタフェース処理を行って、ネットワークを通じて通信データ（パケット）を受信する。通信ログ生成部 11 b は、受信した通信データを解析し、通信データから実行操作のコマンドやファイルアクセスにもとづき通信ログを生成（再構成）する。

[0028] セッション分割部 11 c は、リモート管理操作プロトコル（リモートコンピューティングデバイスに保存されたファイルをユーザ端末に表示してアクセスできるように設計されたプロトコル）のセッションで、通信ログを分割して、分割セッションデータを生成する。なお、リモート管理操作プロトコルとしては、例えば、SMB（Server Message Block）がある。

[0029] 疑似セッション生成部11dは、分割セッションデータに対して、コマンドの実行アカウントが同一アカウントであり、かつセッション間隔があらかじめ定めた閾値以下である複数の分割セッションデータを抽出する。そして、疑似セッション生成部11dは、抽出した分割セッションデータを連結して疑似セッションデータ（図1の連結データに相当）を生成する。

[0030] 危険度レベル決定部11eは、疑似セッションデータに含まれるコマンドおよび疑似セッションデータに含まれるファイルのサイズにもとづいて、危険度レベルを決定する。ファイル情報収集部11fは、疑似セッションデータに含まれるファイルから、危険度レベルにもとづくファイル情報を収集する。

[0031] 記憶部12は、通信データ（通信ログ）、テーブル情報、および収集されたファイル情報等を格納する。また、記憶部12は、サーバ装置10の運用に関する制御情報等を格納する。なお、テーブル情報としては、例えば、コマンドと危険度レベルとの対応関係を示すコマンドー危険度レベル対応テーブルT1、および危険度レベルと収集すべきファイル情報との対応関係を示す危険度レベルーファイル情報対応テーブルT2がある（図13、図14で後述）。

[0032] <ハードウェア>

図4はサーバ装置のハードウェア構成の一例を示す図である。サーバ装置10は、プロセッサ（コンピュータ）100によって全体制御されている。プロセッサ100は、制御部11の機能を実現する。

[0033] プロセッサ100には、バス103を介して、メモリ101、入出力インタフェース102およびネットワークインタフェース104が接続されている。

プロセッサ100は、マルチプロセッサであってもよい。プロセッサ100は、例えば、CPU（Central Processing Unit）、MPU（Micro Processing Unit）、DSP（Digital Signal Processor）、ASIC（Application Specific Integrated Circuit）、またはPLD（Programmable Logic Dev

ice) である。またプロセッサ100は、CPU、MPU、DSP、ASIC、PLDのうちの2以上の要素の組み合わせであってもよい。

[0034] メモリ101は、記憶部12の機能を実現し、サーバ装置10の主記憶装置として使用される。メモリ101には、プロセッサ100に実行させるOS (Operating System) のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。また、メモリ101には、プロセッサ100による処理に要する各種データが格納される。

[0035] また、メモリ101は、サーバ装置10の補助記憶装置としても使用され、OSのプログラム、アプリケーションプログラム、および各種データが格納される。メモリ101は、補助記憶装置として、フラッシュメモリやSSD (Solid State Drive) 等の半導体記憶装置やHDD (Hard Disk Drive) 等の磁気記録媒体を含んでもよい。

[0036] バス103に接続されている周辺機器としては、入出力インタフェース102およびネットワークインタフェース104がある。入出力インタフェース102は、キーボードやマウス等の情報入力装置を接続可能であって、情報入力装置から送られてくる信号をプロセッサ100に送信する。

[0037] さらにまた、入出力インタフェース102は、周辺機器を接続するための通信インタフェースとしても機能する。例えば、入出力インタフェース102は、レーザ光等を利用して、光ディスクに記録されたデータの読み取りを行う光学ドライブ装置を接続することができる。光ディスクには、Blu-ray Disc (登録商標)、CD-ROM (Compact Disc Read Only Memory)、CD-R (Recordable) / RW (Rewritable) 等がある。

[0038] また、入出力インタフェース102は、メモリ装置やメモリリーダライタを接続することができる。メモリ装置は、入出力インタフェース102との通信機能を搭載した記録媒体である。メモリリーダライタは、メモリカードへのデータの書き込み、またはメモリカードからのデータの読み出しを行う装置である。メモリカードは、カード型の記録媒体である。

[0039] ネットワークインタフェース104は、ネットワークに接続してネットワ

ークインタフェース制御を行う。ネットワークインタフェース104には、例えば、NIC (Network Interface Card)、無線LAN (Local Area Network) カード等を使用することもできる。ネットワークインタフェース104で受信されたデータは、メモリ101やプロセッサ100に出力される。

[0040] 以上のようなハードウェア構成によって、サーバ装置10の処理機能を実現することができる。例えば、サーバ装置10は、プロセッサ100がそれぞれ所定のプログラムを実行することで本発明の処理を行うことができる。

[0041] サーバ装置10は、例えば、コンピュータで読み取り可能な記録媒体に記録されたプログラムを実行することにより、本発明の処理機能を実現する。サーバ装置10に実行させる処理内容を記述したプログラムは、様々な記録媒体に記録しておくことができる。

[0042] 例えば、サーバ装置10に実行させるプログラムを補助記憶装置に格納しておくことができる。プロセッサ100は、補助記憶装置内のプログラムの少なくとも一部を主記憶装置にロードし、プログラムを実行する。

[0043] また、光ディスク、メモリ装置、メモリカード等の可搬型記録媒体に記録しておくこともできる。可搬型記録媒体に格納されたプログラムは、例えば、プロセッサ100からの制御により、補助記憶装置にインストールされた後、実行可能となる。またプロセッサ100が、可搬型記録媒体から直接プログラムを読み出して実行することもできる。

[0044] <ファイル情報の収集(疑似セッションデータを生成しない場合)>

次に第2の実施の形態の詳細を説明する前に、疑似セッションデータを生成せずにファイル情報を収集する場合のサーバ装置(サーバ装置20と呼ぶ)の動作およびその課題について図5から図8を用いて説明する。

[0045] 図5はコマンドと危険度レベルとの対応関係を示すテーブル情報の一例を示す図である。テーブルT11は、項目として、コマンドおよび危険度レベルを有する。テーブルT11には、通信ログに含まれるコマンドの危険度レベルが登録されている。

[0046] 図5の例では、(コマンド, 危険度レベル) = (net group, 1)、(net

use, 2)、(schtasks, 3)と登録されている(危険度レベルの数値が高いほど危険度が高い)。

ここで、net groupは、リモート管理操作される側の装置(リモート管理操作元装置)が所属しているアカウント情報等をリモート管理操作する側の装置(リモート管理操作先装置)に一覧表示する参照系のコマンドである。

[0047] net useは、リモート管理操作先装置がリモート管理操作元装置の共有リソース(共有フォルダ)に接続するための接続系のコマンドである。schtasksは、リモート管理操作先装置がリモート管理操作元装置におけるタスクスケジューラのタスク処理を制御する更新系のコマンドである。

[0048] リモート管理操作先装置を通じて攻撃を受けた場合、上記のコマンドに対して、危険度が最も高いコマンド(感染拡大が行われる可能性が最も高いコマンド)がschtasksであり、危険度レベルを3としている。次に危険度が高いコマンドがnet useであり、危険度レベルを2としている。そして、3つのコマンドのうちで危険度が最も低いコマンドがnet groupであり、危険度レベルを1としている。

[0049] 図6は危険度レベルとファイル情報との対応関係を示すテーブル情報の一例を示す図である。テーブルT12は、項目として、危険度レベルおよびファイル情報を有する。テーブルT12には、危険度レベルに応じて収集すべきファイル情報が登録されている。

[0050] 図6の例では、(危険度レベル, ファイル情報) = (1, メタ情報)、(2, メタ情報・ファイルハッシュ値)、(3, メタ情報・ファイルハッシュ値・ファイル本体)と登録されている。

[0051] すなわち、危険度レベルが1の場合、ファイルから収集すべきファイル情報は、該ファイルのメタ情報のみと設定されている。危険度レベルが2の場合、ファイルから収集すべきファイル情報は、該ファイルのメタ情報および該ファイルのファイルハッシュ値と設定されている。危険度レベルが3の場合、ファイルから収集すべきファイル情報は、該ファイルのメタ情報、該ファイルのファイルハッシュ値および該ファイルの本体と設定されている。

[0052] 図7、図8は通信ログをセッション分割してからファイル情報を収集するまでの動作の一例を説明するための図である。

図7において、サーバ装置20は、リモート管理操作された通信データを受信し、受信した通信データから通信ログL2を生成する。通信ログL2は、項目として、日付、時刻、セッションID、コマンドおよびオプションを有する。

[0053] 通信ログL2において、ログL2aの(2019-12-26, 10:40, 1, net use, ipc\$)では、2019-12-26の10:40に、net useコマンドによって共有リソースipc\$に接続要求があったことが記録されている。

[0054] ログL2bの(2019-12-26, 10:42, 1, net use, admin\$)では、2019-12-26の10:42に、net useコマンドによって共有フォルダadmin\$に接続要求があったことが記録されている。

[0055] ログL2cの(2019-12-26, 10:45, 1, WRITE, a.exe)では、2019-12-26の10:45に、WRITEコマンドによって共有フォルダadmin\$にa.exeのファイルの書き込みが行われたことが記録されている。

[0056] ログL2dの(2019-12-26, 10:47, 2, schtasks, create)では、2019-12-26の10:47に、schtasksコマンドによってタスクスケジューラにタスク登録(create)が行われたことが記録されている。

[0057] ログL2eの(2019-12-26, 10:47, 2, schtasks, start)では、2019-12-26の10:47に、schtasksコマンドによって登録されたタスクの実行の開始(start)が行われたことが記録されている。

[0058] ログL2fの(2019-12-26, 10:57, 3, net group, admin)では、2019-12-26の10:57に、net groupコマンドによってadminグループの一覧表示が行われたことが記録されている。

[0059] なお、ログL2aは共有リソースへの接続、L2b、L2cは、共有フォルダへのファイル書き込み関連のリモート管理操作であり、これらのリモート管理操作のセッションIDは1になっている。また、ログL2d、L2eは、タスク処理関連のリモート管理操作であり、このリモート管理操作のセ

セッションIDは2になっている。さらに、ログL2fは、アカウント情報の表示関連のリモート管理操作であり、このリモート管理操作のセッションIDは3になっている。

[0060] このような通信ログL2に対して、サーバ装置20では、通信ログL2をセッション毎に分割する。すなわち、通信ログL2は、セッションIDが1の分割セッションデータd11、セッションIDが2の分割セッションデータd12、およびセッションIDが3の分割セッションデータd13に分割される。

[0061] 図8において、サーバ装置20は、図5に示したテーブルT11にもとづいて、分割セッションデータd11、d12、d13それぞれの危険度レベルを決定する。分割セッションデータd11は、net useコマンドが含まれるので危険度レベルが2と決定され、分割セッションデータd12は、schtasksコマンドが含まれるので危険度レベルが3と決定される。さらに、分割セッションデータd13は、net groupコマンドが含まれるので危険度レベルが1と決定される。

[0062] サーバ装置20は、危険度レベルを決定した後、分割セッションデータd11、d12、d13のうちからファイル情報を収集する。この場合、分割セッションデータd12、d13にはファイルは含まれず、分割セッションデータd11にはファイルが含まれているから、分割セッションデータd11に含まれるファイルからファイル情報が収集される。

[0063] 分割セッションデータd11には、a.exeファイルが含まれており、また、a.exeファイルは、危険度レベルが2のnet useコマンドに紐づけられている。

このため、サーバ装置20は、図6に示したテーブルT12にもとづいて、a.exeファイルからメタ情報およびファイルハッシュ値をファイル情報として収集する。したがって、通信ログL2からは、ファイル情報としてメタ情報およびファイルハッシュ値が収集されて、ストレージに格納されることになる。

- [0064] しかし、上記のようなサーバ装置20の制御では、ファイル情報の収集精度が低く、収集すべきファイル情報が不十分である。なぜなら、通信ログL2には、schtasksという危険度レベルが3のコマンドが含まれているため、本来、通信ログL2から収集したいファイル情報は、メタ情報、ファイルハッシュ値およびファイル本体であるはずなのに、メタ情報およびファイルハッシュ値しか収集できていないからである。
- [0065] このような問題が生じるのは、a. exeファイルが、危険度レベルが3のschtasksコマンドに紐づけられていないことに起因する。a. exeファイルが、危険度レベルが3のschtasksコマンドに紐づけられるような通信ログL2の分割制御が行われるならば、ファイル情報としてファイル本体まで収集することができるようになる。
- [0066] ここで、危険度レベルに応じて、収集すべきファイル情報が異なる理由について説明する。上記のようなnet useやschtasks等のコマンドは、サーバの運用管理者によって正常業務で通常使用されるコマンドである。
- [0067] このため、どのコマンドに対してもリモート管理操作で書き込まれたファイルからファイル本体をすべて収集してストレージに格納してしまうと、ストレージ容量が逼迫してしまうという問題が生じてしまう。
- [0068] また、そのようなファイル情報の収集制御をしてしまうと、攻撃が実際に行われた場合、大量のファイルのうちに不正ファイルが埋没してしまい、正常ファイルの中から不正ファイルを検出することが困難になる。
- [0069] このような理由によって、危険度レベルに応じて異なるファイル情報を収集するようにしている。攻撃者が使用した可能性の高いファイルからファイル本体を収集するようにすれば、よりファイル情報の収集精度を向上させることができる。本発明はこのような点に鑑みてなされたものであり、ファイル情報の収集精度を向上させ、危険度の高い通信に関するデータを効率よく抽出するものである。
- [0070] <本発明のファイル情報の収集（疑似セッションデータを生成する場合）>

次に第2の実施の形態の動作について以降詳しく説明する。第2の実施の形態のサーバ装置10では、例えば、危険度レベルが3のschtasksコマンドに紐づけられるような通信ログの分割制御（疑似セッションデータの生成）を行う。さらにファイルサイズの判定処理も行って、危険度レベルの高いファイルから精度よくファイル本体まで収集することができるようにするものである。

[0071] （通信ログの生成）

図9は通信ログの一例を示す図である。サーバ装置10内の制御部11は、リモート管理操作された通信データを受信し、受信した通信データから通信ログL1を生成する。通信ログL1は、項目として、日付、時刻、セッションID、コマンドおよびオプションを有する。

[0072] 通信ログL1において、ログL1aの（2019-12-26, 10:40, 1, net use, ipc\$, account:A）では、2019-12-26の10:40に、アカウントAにより、net useコマンドによって共有リソースipc\$に接続要求があったことが記録されている。

[0073] ログL1bの（2019-12-26, 10:42, 1, net use, admin\$, account:A）では、2019-12-26の10:42に、アカウントAにより、net useコマンドによって共有フォルダadmin\$に接続要求があったことが記録されている。

[0074] ログL1cの（2019-12-26, 10:45, 1, WRITE, a.exe, account:A, 1MB）では、2019-12-26の10:45に、アカウントAにより、WRITEコマンドによって共有フォルダadmin\$にa.exeのファイルの書き込みが行われたことが記録されている。また、ファイルサイズは1MBと記録されている。

[0075] ログL1dの（2019-12-26, 10:47, 2, schtasks, create, account:A）では、2019-12-26の10:47に、アカウントAにより、schtasksコマンドによってタスクスケジューラにタスク登録（create）が行われたことが記録されている。

[0076] ログL1eの（2019-12-26, 10:47, 2, schtasks, start, account:A）では、2019-12-26の10:47に、アカウントAにより、schtasksコマンドによって

登録されたタスクの実行の開始 (start) が行われたことが記録されている。

[0077] ログ L 1 f の (2019-12-26, 10:57, 3, net group, admin, account:B) では、2019-12-26の10:57に、アカウント B により、net group コマンドによって admin グループの一覧表示が行われたことが記録されている。

[0078] なお、ログ L 1 a は共有リソースへの接続であり、L 1 b、L 1 c は、共有フォルダへのファイル書き込み関連のリモート管理操作であり、このリモート管理操作のセッション ID は 1 になっている。また、ログ L 1 d、L 1 e は、タスク処理関連のリモート管理操作であり、このリモート管理操作のセッション ID は 2 になっている。さらに、ログ L 1 f は、アカウント情報の表示関連のリモート管理操作であり、このリモート管理操作のセッション ID は 3 になっている。

[0079] (分割セッションデータの生成)

図 10 は分割セッションデータの一例を示す図である。制御部 11 は、通信ログ L 1 に対して、通信ログ L 1 をセッション毎に分割する。すなわち、通信ログ L 1 は、セッション ID が 1 の分割セッションデータ d 1、セッション ID が 2 の分割セッションデータ d 2、およびセッション ID が 3 の分割セッションデータ d 3 に分割される。

[0080] (セッション間隔)

次に疑似セッションデータを生成するために用いられるセッション間隔について説明する。図 11 はセッション間隔の一例を説明するための図である。横軸は時間である。セッション s e 1、s e 2、s e 3 が図 11 に示すように連続して行われたとする。

[0081] セッション s e 1 の終了時刻が t 1、セッション s e 2 の開始時刻が t 2 ($> t 1$) のとき、時刻 t 1、t 2 の時間間隔が、セッション s e 1、s e 2 のセッション間隔 t a となる (この場合は正の値)。

[0082] また、セッション s e 2 の終了時刻が t 4、セッション s e 3 の開始時刻が t 3 ($< t 4$) のとき、時刻 t 3、t 4 の時間間隔が、セッション s e 2、s e 3 のセッション間隔 t b となる (この場合は負の値)。

[0083] (セッション間隔の閾値)

セッション間隔の閾値は、例えば、サーバ装置 10 の制御を適用する環境内で観測されるセッション間隔の平均値とする。ここで、同一の送信元、送信先で観測されるセッションの開始時刻を t_s 、セッションの終了時刻を t_e とし、セッション間隔 t_{int} を $t_{int} = t_e - t_s$ とする。

[0084] このとき、環境内で観測されるセッション間隔を $T = \{t_{int,1}, t_{int,2}, t_{int,3}, \dots, t_{int,n}\}$ とすると、セッション間隔 S_{th} は、以下の式 (1) で算出される。なお、 n はデータの総数である。

[0085] [数1]

$$S_{th} = \frac{1}{n} \sum_{i=1}^n |t_{int,i}| \quad \dots (1)$$

[0086] (疑似セッションデータの生成)

図 12 は疑似セッションデータの生成の一例を示す図である。図 10 に示したような分割セッションデータの生成後、制御部 11 は、疑似セッションデータを生成する。制御部 11 は、複数の分割セッションデータのうち、同一のアカウント（または同一のクレデンシャル）の分割セッションデータであり、かつその分割セッションデータのセッション間隔が閾値以下である分割セッションデータを連結して疑似セッションデータを生成する。

[0087] 制御部 11 は、まず、分割セッションデータ d1、d2、d3 のうち、同一のアカウントを持つ分割セッションデータを抽出する。ここでは、分割セッションデータ d1、d2 はアカウント A であり、分割セッションデータ d3 はアカウント B であるから、同じアカウント A を持つ分割セッションデータ d1、d2 が抽出される。

[0088] 次に、制御部 11 は、分割セッションデータ d1、d2 のセッション間隔が閾値以下であるか否かを判定する。分割セッションデータ d1 の終了時刻、すなわちログ L1c の時刻は 10:45 であり、分割セッションデータ d2 の開始時刻、すなわちログ L1d の時刻は 10:47 である。

[0089] よって、分割セッションデータ d 1、d 2 のセッション間隔は 2 分であり、セッション間隔の閾値を 5 分とすると、分割セッションデータ d 1、d 2 はセッション間隔の閾値以下であり条件を満たすことになる。

[0090] したがって、制御部 11 は、同一のアカウントを有し、かつセッション間隔が閾値以下である分割セッションデータ d 1、d 2 を連結して疑似セッションデータ d p を生成する。

[0091] (テーブル構成)

図 13 はコマンドー危険度レベル対応テーブルの一例を示す図である。コマンドー危険度レベル対応テーブル T 1 は、項目として、コマンドおよび危険度レベルを有し、通信ログに含まれるコマンドの危険度レベルがあらかじめ登録されている。

[0092] 図 13 の例では、(コマンド, 危険度レベル) = (net group, 1)、(net use, 2)、(schtasks (ファイルサイズが閾値以上), 2)、(schtasks (ファイルサイズが閾値未満), 3) と登録されている。

[0093] net group は、情報の参照を行うための参照系コマンドの 1 つであり、net use は、共有リソースに接続するための接続系コマンドの 1 つである。また、schtasks は、データ処理の更新を行うための更新系コマンドの 1 つである。

[0094] ここで、net group コマンドの危険度レベルが 1 であり、net use コマンドの危険度レベルが 2 であることは図 5 と同じであるが、schtasks の危険度レベルの設定が異なっている。

[0095] サーバ装置 10 では、同じ schtasks コマンドであっても、schtasks コマンドに紐づけられるファイルのサイズが閾値以上の場合は危険度レベルを 2 とし、schtasks コマンドに紐づけられるファイルのサイズが閾値未満の場合は危険度レベルを 3 としている。

[0096] 図 14 は危険度レベルーファイル情報対応テーブルの一例を示す図である。危険度レベルーファイル情報対応テーブル T 2 は、項目として、危険度レベルおよびファイル情報を有し、危険度レベルに応じて収集すべきファイル情報が登録されている。

[0097] 図14の例では、(危険度レベル, ファイル情報) = (1, メタ情報)、
(2, メタ情報・ファイルハッシュ値)、(3, メタ情報・ファイルハッシュ
値・ファイル本体)と登録されている。

[0098] すなわち、危険度レベルが1の場合、ファイルから収集すべきファイル情
報は、該ファイルのメタ情報のみとなる。メタ情報には例えば、ファイル名
、ファイルの拡張子およびタイムスタンプ(ファイルの変更・修正が行われ
た時期を示す情報)がある。

[0099] また、危険度レベルが2の場合、ファイルから収集すべきファイル情報は
、該ファイルのメタ情報および該ファイルのファイルハッシュ値となる。さ
らに、危険度レベルが3の場合、ファイルから収集すべきファイル情報は、
該ファイルのメタ情報、該ファイルのファイルハッシュ値および該ファイル
の本体となる。このように、危険度レベルが高くなるほど、抽出されたファ
イルから収集すべきファイル情報の種類が多くなるように設定されている。

[0100] <ファイルサイズの閾値>

ファイルサイズの閾値(上限値)は、サーバ装置20の制御を適用する環
境内で過去に観測された拡張子と実際のファイルの種類が異なっているファ
イルのサイズ、または実行形式のファイルのサイズから算出される平均に標
準偏差を加算したものとする。

[0101] ここで、環境内で過去に観測された拡張子と実際のファイルの種類が異な
っているファイル、または実行形式のファイルのサイズ S を $S = \{S_1, S_2, S_3, \dots, S_n\}$ とすると、平均 μ は以下の式(2)で表せる(n はデータ
の総数である)。したがって、ファイルサイズの閾値 F_{th} は以下の式(3)
で算出される。なお、 x_i は個々の数値である。

[0102] [数2]

$$\mu = \frac{1}{n} \sum_{i=1}^n S_i \quad \dots(2)$$

[0103]

[数3]

$$F_{th} = \mu + \sqrt{\sum_{i=1}^n \frac{1}{n} (x_i - \mu)^2} \quad \dots (3)$$

[0104] なお、通常のアプリケーションで使用される実行ファイルのサイズは、5 MBから10 MB程度のもが多いが、マルウェアのファイルは例えば、数百KBから5 MB未満のもが多い。このため、上記のようにファイルサイズの閾値（上限値）を定めて、不正ファイルの特定を図っている。

[0105] （危険度レベルの決定）

図15は危険度レベルの決定の一例を示す図である。制御部11は、図13に示したコマンドー危険度レベル対応テーブルT1にもとづいて、疑似セッションデータd pおよび分割セッションデータd 3の危険度レベルを決定する。

[0106] ファイルサイズの閾値を3 MBとする。疑似セッションデータd pには、schtasksコマンドとファイルとが含まれており、該ファイルのサイズは1 MBとなっている。

したがって、疑似セッションデータd pは、ファイルサイズが閾値未満のファイルに紐づくschtasksコマンドを有するから、コマンドー危険度レベル対応テーブルT1にもとづいて、危険度レベルが3と決定される。なお、分割セッションデータd 3には、net groupコマンドが含まれるので危険度レベルが1と決定されている。

[0107] （ファイル情報の収集）

図16はファイル情報の収集の一例を示す図である。制御部11は、危険度レベルを決定した後、疑似セッションデータd pおよび分割セッションデータd 3のうちからファイル情報を収集する。

[0108] この場合、分割セッションデータd 3にはファイルは含まれず、疑似セッションデータd pにはファイルが含まれているから、疑似セッションデータd pに含まれるファイルからファイル情報が収集される。

[0109] 疑似セッションデータ d p には、a. exe ファイルが含まれており、また、a. exe ファイルは、危険度レベルが 3 の schtasks コマンドに紐づけられている。

このため、制御部 11 は、図 14 に示した危険度レベル—ファイル情報対応テーブル T2 にもとづいて、a. exe ファイルからメタ情報、ファイルハッシュ値およびファイル本体をファイル情報として収集する。したがって、通信ログ L1 からは、不正ファイルのファイル情報としてメタ情報、ファイルハッシュ値およびファイル本体が収集されて、記憶部 12（ストレージ）に格納されることになる。

[0110] このように、制御部 11 では、通信ログ L1 から上記のような疑似セッションデータを生成する分割制御を行い、さらにファイルサイズの判定を行う。これにより、a. exe ファイルが、ファイル本体まで収集するための危険度レベルに対応付けられるので、ファイル情報としてファイル本体まで収集することができる。

[0111] （フローチャート）

図 17 は疑似セッションデータの生成動作の一例を示すフローチャートである。

〔ステップ S11〕 制御部 11 は、通信データを分割し、分割セッションデータを生成する。

[0112] 〔ステップ S12〕 制御部 11 は、同一アカウント（または同一クレデンシャル）で実行される分割セッションデータのセッション間隔を算出する。

〔ステップ S13〕 制御部 11 は、セッション間隔が閾値以下か否かを判定する。閾値以下の場合はステップ S14 に処理が進み、閾値を超える場合はステップ S16 に処理が進む。

[0113] 〔ステップ S14〕 制御部 11 は、セッション連結の処理を行い、分割セッションデータを連結する。

〔ステップ S15〕 制御部 11 は、閾値の時間内に続く分割セッションデータがあるか否かを判定する。分割セッションデータがある場合はステップ

S 1 3 に処理が戻り、ない場合はステップ S 1 6 に処理が進む。

[0114] 〔ステップ S 1 6〕 制御部 1 1 は、セッション連結の処理を終える。

 〔ステップ S 1 7〕 制御部 1 1 は、連結した分割セッションデータを疑似セッションデータとする。

[0115] 図 1 8 は危険度レベルにもとづくファイル情報収集の動作の一例を示すフローチャートである。

 〔ステップ S 2 1〕 制御部 1 1 は、疑似セッションデータ内のコマンドの分析を行う。

[0116] 〔ステップ S 2 2〕 制御部 1 1 は、疑似セッションデータ内に更新系コマンド（例えば、schtasks）が含まれるか否かを判定する。更新系コマンドが含まれる場合はステップ S 2 3 に処理が進み、含まれない場合はステップ S 2 4 に処理が進む。

[0117] 〔ステップ S 2 3〕 制御部 1 1 は、疑似セッションデータ内のファイルのサイズが閾値未満か否かを判定する。閾値未満の場合はステップ S 2 5 a に処理が進み、閾値以上の場合はステップ S 2 5 b に処理が進む。

[0118] 〔ステップ S 2 4〕 制御部 1 1 は、疑似セッションデータ内のコマンドが接続系（例えば、net use）、または参照系（例えば、net group）のいずれであるかを判定する。接続系コマンドの場合はステップ S 2 5 b に処理が進み、参照系コマンドの場合はステップ S 2 5 c に処理が進む。

[0119] 〔ステップ S 2 5 a〕 制御部 1 1 は、疑似セッションデータの危険度レベルを 3 に決定する。

 〔ステップ S 2 5 b〕 制御部 1 1 は、疑似セッションデータの危険度レベルを 2 に決定する。

[0120] 〔ステップ S 2 5 c〕 制御部 1 1 は、疑似セッションデータの危険度レベルを 1 に決定する。

 〔ステップ S 2 6 a〕 制御部 1 1 は、疑似セッションデータ内のファイルからファイル情報としてメタ情報、ファイルハッシュ値およびファイル本体を収集する。

[0121] 〔ステップS 2 6 b〕制御部 1 1 は、疑似セッションデータ内のファイルからファイル情報としてメタ情報およびファイルハッシュ値を収集する。

 〔ステップS 2 6 c〕制御部 1 1 は、疑似セッションデータ内のファイルからファイル情報としてメタ情報を収集する。

[0122] 〔ステップS 2 7〕制御部 1 1 は、収集したファイル情報を記憶部 1 2 に格納する。

 図 1 9 は危険度レベルにもとづくファイル情報収集の動作の一例を示すフローチャートである。図 1 9 の動作では、危険度レベル 3 のときに疑似セッションデータ内にファイルがあればサイズに関係なくファイル本体まで一旦収集しておく。そして、ファイルサイズが閾値以上と判定された場合は収集したファイルを除去し、閾値未満であればそのまま収集しておくといった動作である。

[0123] 〔ステップS 3 1〕制御部 1 1 は、疑似セッションデータ内にリモート管理操作で書き込まれたファイルのサイズが閾値未満か否かを判定する。閾値未満の場合はステップS 3 2 に処理が進み、閾値以上の場合はステップS 3 3 に処理が進む。

[0124] 〔ステップS 3 2〕制御部 1 1 は、ファイル本体を取得するためのファイル本体取得スイッチをオンにする。

 〔ステップS 3 3〕制御部 1 1 は、ファイル本体取得スイッチをオフにする。

[0125] 〔ステップS 3 4〕制御部 1 1 は、疑似セッションデータの危険度レベルを決定する。

 〔ステップS 3 5〕制御部 1 1 は、決定した危険度レベルが 1 であるか否かを判定する。危険度レベルが 1 の場合はステップS 3 6 に処理が進み、1 でない場合はステップS 3 7 に処理が進む。

[0126] 〔ステップS 3 6〕制御部 1 1 は、危険度レベル 1 のファイル情報（メタ情報）を収集する。ステップS 4 0 に処理が進む。

 〔ステップS 3 7〕制御部 1 1 は、決定した危険度レベルが 2 であるか否

かを判定する。危険度レベルが2の場合はステップS 3 8に処理が進み、2でない場合はステップS 3 9に処理が進む。

[0127] 〔ステップS 3 8〕制御部1 1は、危険度レベル2のファイル情報（メタ情報およびファイルハッシュ値）を収集する。ステップS 4 0に処理が進む。

〔ステップS 3 9〕制御部1 1は、危険度レベルが3であると認識し、危険度レベル3のファイル情報（メタ情報、ファイルハッシュ値およびファイル本体）を収集する。

[0128] 〔ステップS 4 0〕制御部1 1は、ファイル本体取得スイッチの状態（オンかオフか）を判定する。オフの場合はステップS 4 1に処理が進み、オンの場合はステップS 4 2に処理が進む。

[0129] 〔ステップS 4 1〕制御部1 1は、収集されたファイル情報のうちからファイル本体を除去する。

〔ステップS 4 2〕制御部1 1は、収集したファイル情報を記憶部1 2に格納する。

[0130] （効果）

次に本発明を適用した場合の効果について説明する。前提として、マルウェアのファイルサイズは3 MB以下が多く、そのようなファイルを収集したいとする。

[0131] 想定状況としては、ファイルを1件書き込み、5分以内に更新系操作を実行する操作を100回／日とし、ファイル書き込みと別セッションで実行される更新系操作の割合を80％とする。また、3 MB以下のファイルの書き込みは全体に一様に分布し、40％（40件／日）起こっているとする。

[0132] このとき、取得できる3 MB以下のファイル件数は（セッション間隔が5分以内のセッションを連結する場合）、本発明の適用前では8件（＝100回×（1－0.8）×0.4）である。これに対し、本発明を適用した場合は40件（＝100件×0.4）となり、ファイル取得率が5倍になり、マルウェアのファイルを効率よく取得することができる。

- [0133] 以上説明したように、本発明によれば、通信データをセッション毎に分割し、同一識別情報を有しセッション間隔が閾値以下の分割セッションを連結し、連結セッションから通信の危険度にもとづいてファイル情報を収集する。これにより、ファイル情報の収集精度が向上し、また危険度の高い通信に関するデータを効率よく抽出することが可能になる。また、攻撃全容調査をすみやかに行うことができ、素早い対処に繋げることができるため、攻撃による損失を最小限にすることが可能になる。
- [0134] 上記で説明した本発明の情報処理装置１およびサーバ装置１０は、コンピュータによって実現することができる。この場合、情報処理装置１およびサーバ装置１０が有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。
- [0135] 処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、磁気記憶部、光ディスク、光磁気記録媒体、半導体メモリ等がある。磁気記憶部には、ハードディスク装置（ＨＤＤ）、フレキシブルディスク（ＦＤ）、磁気テープ等がある。光ディスクには、ＣＤ－ＲＯＭ／ＲＷ等がある。光磁気記録媒体には、ＭＯ（Magneto Optical disk）等がある。
- [0136] プログラムを流通させる場合、例えば、そのプログラムが記録されたＣＤ－ＲＯＭ等の可搬型記録媒体が販売される。また、プログラムをサーバコンピュータの記憶部に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。
- [0137] プログラムを実行するコンピュータは、例えば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶部に格納する。そして、コンピュータは、自己の記憶部からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。

[0138] また、コンピュータは、ネットワークを介して接続されたサーバコンピュータからプログラムが転送される毎に、逐次、受け取ったプログラムに従った処理を実行することもできる。また、上記の処理機能の少なくとも一部を、DSP、ASIC、PLD等の電子回路で実現することもできる。

[0139] 以上、実施の形態を例示したが、実施の形態で示した各部の構成は同様の機能を有する他のものに置換することができる。また、他の任意の構成物や工程が付加されてもよい。さらに、前述した実施の形態のうちの任意の2以上の構成（特徴）を組み合わせたものであってもよい。

[0140] 上記については単に本発明の原理を示すものである。さらに、多数の変形、変更が当業者にとって可能であり、本発明は上記に示し、説明した正確な構成および応用例に限定されるものではなく、対応するすべての変形例および均等物は、添付の請求項およびその均等物による本発明の範囲とみなされる。

符号の説明

- [0141] 1 情報処理装置
- 1 a 制御部
 - 1 b 記憶部
 - c d 1、c d 2、c d 3、c d 4 コマンド
 - A、B 識別情報
 - D 0 通信データ
 - D g 分割データ群
 - D 1、D 2、D 3 分割データ
 - D L 連結データ