特集論文

おとりを用いた内部犯検知技術

Insider Threat Detection Using Decoy

山本　匠*
Takumi Yamamoto
野澤康平†
Kohei Nozawa
宇谷亮太†
Ryota Udani

*情報技術総合研究所（博士（情報学））
†同研究所

要　旨

サイバー攻撃による情報漏えいが課題であると言われているが，実際には，内部犯による情報漏えいも数多く発生している(1)。内部犯は正規の権限を持つ悪意を持ったユーザーであり，既存のセキュリティー対策ではその兆候を明確に捉えることが難しい。内部犯の悪意を顕在化させるために，囮（おとり）に着目した。普段とは異なる活動をするユーザーに対して，動的に囮のファイルを配置し，囮のファイルへのアクセスの傾向を基に内部犯を絞り込む手法を開発した。

1．ま え が き

“内部不正による情報漏えい等の被害”はIPA（独立行政法人 情報処理推進機構）が公開する“情報セキュリティ10大脅威2025”で第４位であり，10年連続で10大脅威として取り扱われている(2)。また重要インフラや工場などのOT（Operational Technology）環境でも，内部犯による不正は，深刻な脅威として認識されている(3)。このことからも内部犯への対策の検討が必要と言える。本稿では，三菱電機の内部犯に対する技術的な対策の取組みについて述べる。

2．内部犯の脅威

本稿で扱う内部犯の定義は，正規のアカウントを持つ悪意を持ったユーザーである。内部犯に対しては，一度認証されたユーザーを原則信用する，従来の監視・防御モデルでの対策は難しい。そのため，あらかじめ分析しておいたユーザーの正常な振る舞いから，振る舞いの異常をリスク値として評価し高リスクのユーザーをあぶり出す，アノマリー検知やUEBA（User and Entity Behavior Analytics）などの振る舞い分析技術の適用が有効であると考えられている(4)。しかし，ユーザーの業務内容は多岐にわたり，また，同一ユーザーが同じ業務を行うときでもユーザーの振る舞いは変動するため，正確にリスク値を算出することは難しい。また内部犯は見つからないように振る舞い分析技術が低リスクと評価する範囲で不正を働くことが考えられ（例えば，少しずつ機密ファイルをアップロードする），振る舞い分析技術で厳しく監視すれば誤報が頻出してしまう。

これらのことから，内部犯を正確に見つけだすことは非常に難しい課題であると言える。正規の権限の範囲で悪意のある行為を実行するという性質上，内部犯と悪意のないユーザー（“正規ユーザー”という。）との差は，悪意の有無だけであり，振る舞いを受動的に観測しているだけでは識別が困難である。

3．提案する内部犯対策技術

そこで筆者らは，受動的に観測できる情報だけから内部犯と正規ユーザーとを切り分けるのではなく，内部犯なら積極的に行うであろう行為を防御側が能動的に誘発させることで，内部犯の悪意を間接的に顕在化させ，悪意に関連する行為を観測するというアプローチを採用した。内部犯の悪意を間接的に顕在化させるために，普段とは異なる活動をするユーザーに対して，もしそのユーザーが内部犯であれば興味を引きそうな囮ファイルを動的に配置し，囮ファイルへのアクセスの傾向を基に内部犯を絞り込むというのが，提案方式のコンセプトである。

提案方式の概要は次のとおりである（**図１**）。

①アノマリー検知などの振る舞い分析システムでファイルアクセスのログなどを分析し，振る舞いの異常をリスク値として算出する。さらに，対象ユーザー名，リスク値，ファイルアクセスの情報を含むリスク情報を作成する。

②囮システムは，振る舞い分析システムからリスク情報を受け取って，高リスクのユーザーに対しては，そのユーザーが興味を引きそうなトピックをファイルアクセスの情報から推定し，そのトピックに関連するコンテンツを含んだ囮ファイルを，様々なトピックの囮ファイルが保存された囮データベースの中から選択する。その後，選択された囮ファイルが，あたかもファイルサーバー上に配置されているかのように当該ユーザーのファイルエクスプローラーの表示を制御する。内部犯にとって魅力的な囮ファイルを配置することで，囮ファイルを選択する確率が増えて，内部犯をより特定しやすくする。配置した囮ファイルに関する情報(ファイル名，対象ユーザー名など)は囮情報として振る舞い分析システムへと渡される。

③振る舞い分析システムは，通常のファイルと同様に，囮ファイルへのアクセスも監視する。振る舞い分析システムは，囮システムからの囮情報を基に，囮ファイルへのアクセスの有無を判断する。囮ファイルへのアクセスがあれば，振る舞い分析システムのリスク値(又はしきい値)にフィードバックすることで，囮ファイルに積極的にアクセスするより高リスクなユーザーに対する監視を強化し，内部犯の候補として絞り込んでいく。

高リスクと判定されたユーザーに絞って囮ファイルを配置することで，普段どおりの業務をしている低リスクの正規ユーザーへの業務阻害を軽減できる。万が一正規ユーザーが高リスクと判定され，囮ファイルが配置されたとしても，普段使っているファイル以外へは積極的にアクセスすることはないと考えられ，業務阻害の可能性は低いと期待する。

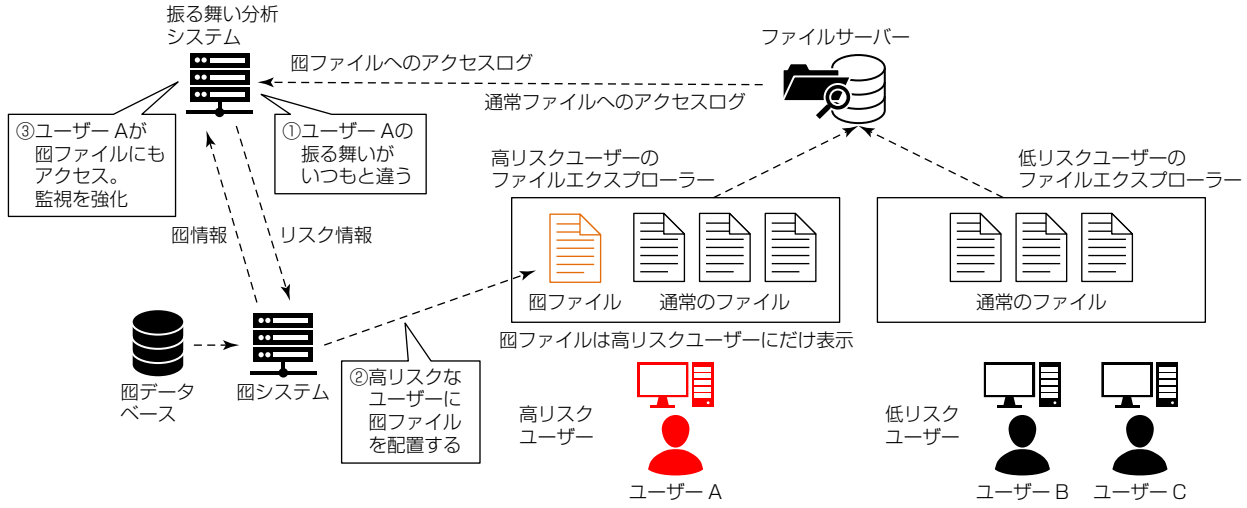


図１−提案方式の概要

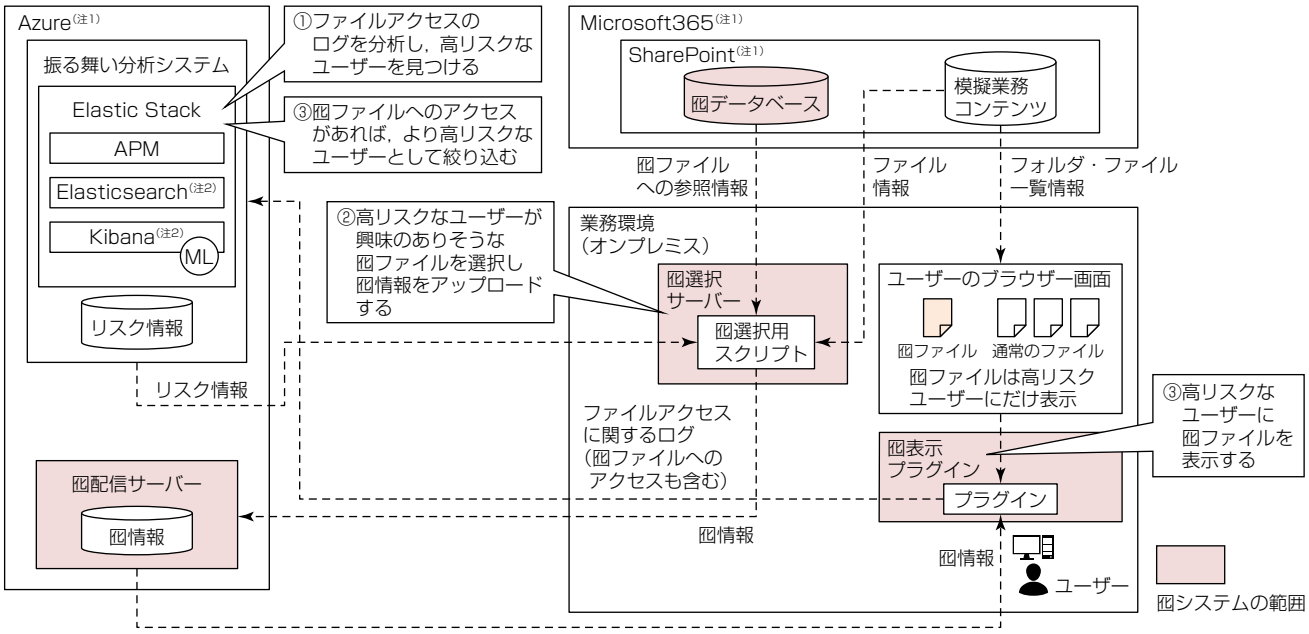
4. 提案方式の試作と動作イメージ

提案方式の実現性の確認のため，試作を行った。試作したシステムの構成を**図２**に示す。

詳細については筆者らの既存研究[5]を参照されたいが，**図２**の囮選択サーバー，囮配信サーバー，囮表示プラグイン，及び囮データベースが連携し，**図１**の囮システムに該当する機能が実現される。昨今のクラウドベースのファイルのやり取りの増加から，今回の試作では**図１**に示すファイルサーバーは利用せず，代わりにSharePoint[6]上に業務で使うファイルを模擬したコンテンツを配置した。振る舞い分析システムはAzure[7]上にElastic Stack[8]を用いて構築した。囮配信サーバーもAzure上に構築した。囮データベースから囮ファイルの参照情報を取得する囮選択サーバーは，業務環境上に構築した。囮データベースは，模擬業務コンテンツと同じSharePoint上に配置した。SharePoint上のファイルにブラウザーでアクセスすることを想定し，対象ユーザーにだけ囮ファイルを閲覧可能にするための制御は，ブラウザーに囮表示プラグインを導入することで実現した。

4.1　試作システムの動作の流れ

次にシステム構成上の動作の流れを述べる(**図２**)。



APM：Application Performance Monitoring，ML：Machine Learning
（注１） Azure，Microsoft365，SharePointは，Microsoft Corp.の登録商標である。
（注２） Elasticsearch，Kibanaは，Elasticsearch B.V.の登録商標である。

図２−試作したシステムの構成

①振る舞い分析システムは，SharePointのファイルアクセスに関するログを分析し，リスク値を算出する。さらに，高リスクのユーザーの名前，リスク値，高リスクと判断されたファイルアクセスの情報を含むリスク情報を作成する。

②囧選択サーバーは，リスク情報を振る舞い分析システムから定期的に取得し，高リスクのユーザーが直近にアクセスしたファイルの情報(ファイル名やファイルのコンテンツなど)からファイルのトピックを推定する。囧データベースからトピックに関連する囧ファイルへの参照情報を取得後，当該ユーザー向けの囧情報を作成し，囧配信サーバーにアップロードする。

③囧表示プラグインは，定期的に囧配信サーバーから囧情報を取得し，高リスクなユーザーであれば囧ファイルを表示する。今回の実装では，囧システムから振る舞い分析システムに送られる囧情報(**図１**)の代わりに，囧表示プラグインから振る舞い分析システムに対して，SharePoint上のファイルへのアクセスのログと一緒に囧ファイルへのアクセスかどうかの情報(ラベル)も送信している。振る舞い分析システムが囧ファイルへのアクセスを確認した場合，当該ユーザーをより高リスクなユーザーとして絞り込む。

4.2　試作システムの動作イメージ

振る舞い分析のリスク値のダッシュボードや囧ファイルが表示される画面を例に，試作物の動作イメージを述べる(**図３**)。

①振る舞い分析システムが，ユーザーのSharePointのファイルアクセスに関するログを分析し，リスク値を算出する。リスク値をダッシュボードに可視化した例を**図３**に示す。グラフ横軸は時刻，縦軸がリスク値であり，折れ線が各時刻での，あるユーザーの振る舞いに対するリスク値の算出結果を示している。

②振る舞い分析システムは，リスク値が所定のしきい値を超えた場合にアラートを発報し，リスク情報を作成する。囧選択サーバーは，リスク情報を取得することでアラートの有無を確認する。通常時，しきい値はθ_1に設定されている(**図３**のピンクの線)。**図３**の例では，②の時点で，リスク値がしきい値θ_1を超えたため，この直後にリスク情報が作成され囧選択サーバーがアラートを確認する。アラートにはリスク値のほかに，しきい値やアクセスしたファイルの情報などが付与される(**図３**中のダッシュボード下部)。

③アラートが発報されると，当該ユーザーには囧ファイルが表示される。囧データベースから得た情報を基に囧ファイルのメタ情報(タイムスタンプ，ファイルサイズ，更新者など)も再現可能である。ブラウザーで囧ファイルを表示した例を**図４**に示す。**図４**(a)の例では，説明の都合上，囧ファイルかどうかが分かるように，“【囧】”という文字列をファイル名に付与している。**図４**(b)の例では，囧ファイルをブラウザーで開いたときの画面である。なお本稿の図の

中に示されているファイル名やファイルのコンテンツは，実際の業務とは一切関係のない架空のものである。また囮ファイルの用意については今回手作業で行ったが，昨今急速に進化している生成AI技術を活用すれば，品質の高いコンテンツの囮ファイルを大量に自動生成することも可能であると考えるが，生成されたコンテンツの著作権については十分注意が必要である。

④ユーザーが囮ファイルにアクセスしたら，アラート発報のしきい値を，当該ユーザーだけ厳しい値に変更($\theta_1 \rightarrow \theta_2$)し，監視を強める。**図３**の例では，利用するしきい値がピンク色の線(θ_1)から緑色の線(θ_2)に変わって，低いリスク値でもアラートが発報されるようになる(**図３**の④)。囮ファイルへのアクセスが一定期間なければ，しきい値は元の値(θ_1)に戻す。囮ファイルへのアクセスの有無が分かるように，囮ファイルへのアクセスがあった場合，青色の折れ線グラフを，リスク値の折れ線グラフに加えた形で描画している(**図３**の③)。囮ファイルか否かについては，ファイルアクセスログに付与されるラベルを基に判断する。

なお，説明の都合上，**図３**のリスク値は，あらかじめ決めたルールに基づいて算出されている。実運用ではあらかじめ手作業でルールを定義する必要のない機械学習によるアノマリー検知技術を使うことも可能である。

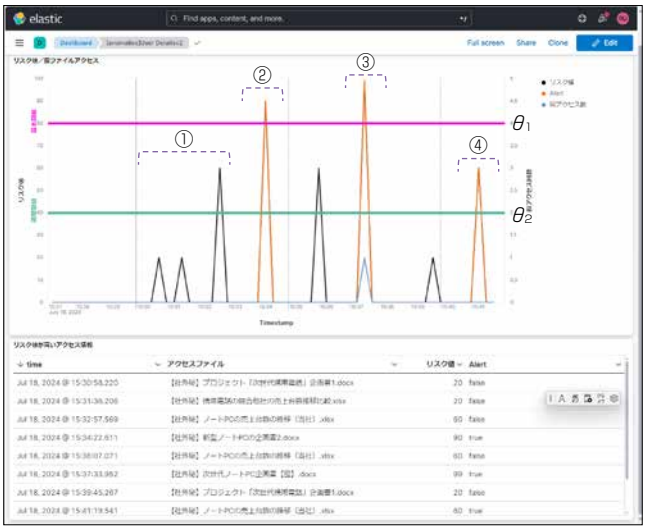


図３−振る舞い分析のリスク値のダッシュボードの例

(a)　SharePoint画面上に表示される囮ファイル

(b)　囮ファイルをブラウザーで開いたときの画面

図４−囮ファイルを表示した例

4.3　対策を無効化する可能性への対応

最後に，この対策の導入を知った内部犯が対策を無効化する可能性が考えられる。プラグインを無効化する攻撃については，プラグインの導入を組織のグループポリシーで強制し，機密ファイルが置かれたSharePointにプラグインなしではアクセスできない環境にすることで対策可能である。ファイルのメタ情報や囮情報などを確認し囮ファイルを見分けようとする内部犯に対しては，当該行動を危険な振る舞いとしてスコアリングし，相補的な監視をすることで対策可能である。

5. む　す　び

従来の方法では検知が困難であった内部犯の脅威に対して，内部犯の悪意を間接的に顕在化させるために，普段とは異なる活動をするユーザーに動的に興味を引きそうな囮ファイルを配置し，囮ファイルへのアクセスの傾向を基に内部犯を絞り込む手法を提案した。また簡単な実装例を示し，実現可能性を示した。

今後は，過去の内部犯の事例などを基にユーザースタディーを行い，提案方式の有効性の考察を行っていく。

参 考 文 献

(1) Proofpoint, Inc.：2022 COST OF INSIDER THREATS GLOBAL REPORT（2022）
https://protectera.com.au/wp-content/uploads/2022/03/The-Cost-of-Insider-Threats-2022-Global-Report.pdf

(2) 独立行政法人 情報処理推進機構：情報セキュリティ10大脅威 2025（2025）
https://www.ipa.go.jp/security/10threats/10threats2025.html

(3) National Counterintelligence and Security Center：INSIDER THREAT MITIGATION FOR U.S. CRITICAL INFRASTRUCTURE ENTITIES, GUIDELINES FROM AN INTELLIGENCE PERSPECTIVE（2024）
https://www.dni.gov/files/NCSC/documents/nittf/20240926_Insider-Threat-Mitigation-for-US-Critical-Infrastructure.pdf

(4) 日本アイ・ビー・エム㈱：ユーザーおよびエンティティーの行動分析（UEBA）とは
https://www.ibm.com/jp-ja/topics/ueba

(5) 山本　匠，ほか：おとりを用いた内部犯検知システムの提案，コンピュータセキュリティシンポジウム2024論文集，494～500（2024）

(6) Microsoft Corp.：What is SharePoint?
https://support.microsoft.com/en-us/office/what-is-SharePoint-97b915e6-651b-43b2-827d-fb25777f446f

(7) Microsoft Corp.：What is Azure?
https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-azure

(8) Elasticsearch B.V.：Elastic Stack
https://www.elastic.co/jp/elastic-stack