内部不正防止策へのWebアクセスログ活用に関する考察
A Study on Web access log utilization for internal anti-fraud measures.

上河内栄治・マネジメント分科会・情報セキュリティ大学院大学

Abstract: There has been a remarkable increase in publication which includes research presentation on internal fraud since the outbreak of the incident where a huge number of personal data leaked via internal fraud in 2014. Although a lot of suggestions came out from researchers based on the facts to counter internal frauds, they've failed so far to demonstrate the validation of their proposals with actual data, except for presenting survey results. This paper introduces the results observed through the statistical analysis on the 'actual' web access logs of an anonymous cooperating company, and my future plans for further research especially on utilizing more of actual data.

1.研究の目的

セキュリティオペレーションセンター（SOC）では、インシデント発生にいち早く気づくことを目的として多くのログを監視している。

監視ログに「組織風土の変化を知る情報」が含まれているとすれば、ログのモニタリングが「内部不正を防止する取り組み」にも活用できることになり、SOC機能のさらなる価値向上につなげることができる。

このことから、監視対象ログから「**組織の小さな不正発生を発見する**」という観点で、Webアクセスログの実データを使った分析に取り組んだ。

2. 着目点

研究用のログを提供いただいた企業ではファイアウォール統合型のプロキシサーバーを使用しており、設定したWebアクセスポリシーに違反するアクセスを検知すると、「違反理由をログに記録し通信を遮断する」運用が行われていた。

Webアクセスポリシー違反が故意ではなくミスやスリップで発生している場合、発生要因はヒューマンエラーであり不正ではないが、もし、**「悪いことだと知りながら発生させた」**のであれば、不正（ルール違反）の発生を知る情報となる。さらに通信遮断の発生率が組織により異なる場合、**組織風土の違いを示す指標にも活用できる**と考えた。

3. 使用データの概要

ログを提供いただいた企業のプロフィールとログデータの概要が表3.1、2015年9月のWebアクセス件数とWebアクセスポリシー違反発生率(以降、違反率)を日付別と時間別でグラフにまとめたものが、それぞれ図3.1、図3.2である。

違反率はそれぞれ**土日祝日、および、22時以降の深夜時間**で高くなる傾向がわかった。

なおアクセス数が最も多い昼休みは違反率が低下している。

この状況は、10月、11月でも同様の傾向であった。

表3.1 ログ提供企業とログデータのプロフィール

企業プロフィール	
業態	ICTの運用，保守と関連するコールセンタ業務
従業員数	約2,000人
拠点数	全国約100拠点
ログデータの概要	
ログの種類	Webアクセスログ(Proxyログ)
収集期間	2015年9月～11月(3ヶ月間)
ログ件数	約90,000,000件/月
記録情報	日時，送信元アドレス，アクセス結果，アクション，通信量，接続先情報，ポート番号，カテゴリー，ポート番号，カテゴリー，フィルタリング結果，フィルタリング理由，等29項目

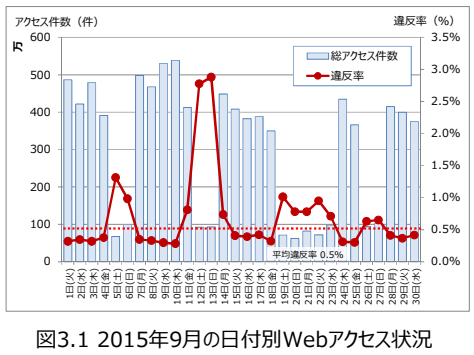

図3.1 2015年9月の日付別Webアクセス状況

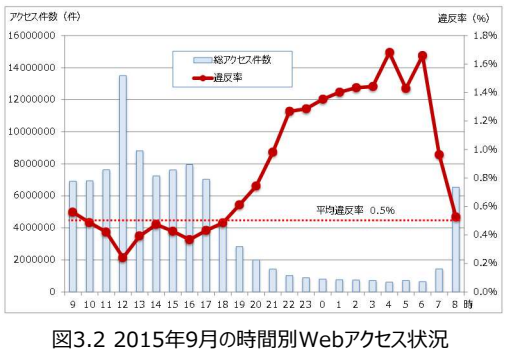

図3.2 2015年9月の時間別Webアクセス状況

4.統計分析結果

つづいて統計分析から相関が確認できた代表的な項目を紹介する。

図4.1 は1時間単位で集計したWebアクセス件数と勤務していた人数をそれぞれ目的変数と説明変数に設定し回帰分析を行った結果で、当然だが、強い正の相関があった。

なお勤務人数については、ユニークなIPアドレス数のカウントで代用した。（以降も同様）

図4.1 Webアクセス件数と勤務人数の相関

図4.2は同様に違反率を目的変数、Webアクセス件数を説明変数に設定して回帰分析を行った結果で負の相関が確認できた。

なお、説明変数に勤務人数を設定した場合も負の相関があった。

Webアクセスポリシ違反がヒューマンエラーで発生してる場合、エラー率は一定のはずだが、負の相関を示したことで、**「違反は意図的な不正（ルール違反）をきっかけに発生している」**可能性が高いことが分かった。

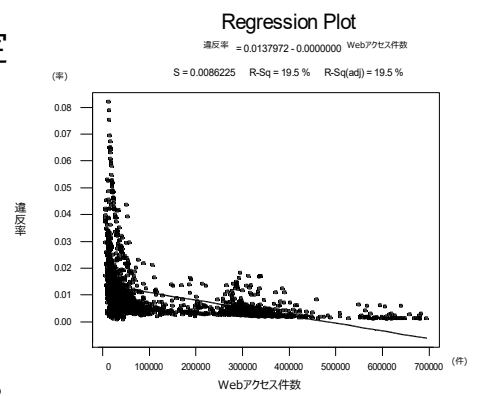

図4.2 違反率と勤務人数の相関

なお表4.1は違反率を従属変数、組織を独立変数に設定して分散分析を行った結果である。

組織毎の違反率に有意な差があることもわかった。

表4.1 違反率と組織の分散分析結果

One-way ANOVA: 違反率 versus 組織

Analysis of Variance for 違反率

Source	DF	SS	MS	F	P
組織	32	0.0008488	0.0000265	7.10	0.000
Error	66	0.0002466	0.0000037		
Total	98	0.0010954			

Individual 95% CIs For Mean

5.考察

今回取り組んだ「Webアクセスログ分析」から、Webアクセスポリシー違反は不正（ルール違反）によって発生している可能性が高いこと、そして組織により発生率の傾向が異なることがわかった。

これまで「発見的コントロール」で使用されていたWebアクセスログが、**「小さなルール違反の発見」にも活用できる可能性が高い**と判断することができた。

残念ながら今回はデータ不足により、退職や懲戒処分など、組織風土との関係が予想される変数とは相関が確認できなかったが、今後も研究を継続し、組織風土の変化把握に活用できる確証を得たい。

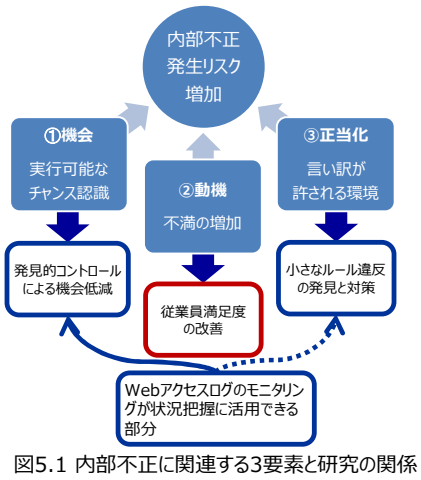

図5.1 内部不正に関連する3要素と研究の関係