Splunk Enterprise を活用した Box の利用状況分析

坂東 佑一

慶應義塾インフォメーションテクノロジーセンター本部

bando@keio.jp

Usage Analysis of Box using Splunk Enterprise

Yuichi Bando

Information Technology Center, Keio University

概要

慶應義塾では，教育研究活動及び事務業務に関わるデータの保存・共有場所として，クラウドストレージである Box を導入し，2016 年より全学向けに提供している．こうしたストレージの利用状況を管理者が把握することはセキュリティや監査の面で非常に重要であるが，Box が標準で備えるログ検索機能だけでは，高度な分析を行うことは少々困難である．そこで，ログデータ分析基盤である Splunk Enterprise（以下 Splunk）を Box と連携させることで，手間を最小限に留めつつ Box の利用状況を詳細に分析・可視化できるようにした．

1 はじめに

本学では，学内外とのデータ共有・コラボレーションの活発化を目的に，2016 年にクラウドストレージである Box[1]を導入した．利用対象者は全教職員・全学生(一貫教育校の生徒・児童は除く)約 50,000 名であり，本学における主要なアプリケーションの一つとして日々活発に利用されている．

こうした Box の利用状況を管理者が適切に把握することは，インシデントの未然防止や早期発見，ユーザへのサービス向上，といった観点で非常に重要である．例えば，短時間の間に人手とは考えられない程の大量のファイルをダウンロードした，といったアクティビティを検知することができれば，これはアカウントが乗っ取られ機械的にデータを盗まれているのではないか，という推測を立て早期に調査に乗り出すことができる．また例えば，利用が少ないユーザを見つけることができれば，積極的に利用促進を行うといった普及活動に役立てることも可能である．こうしたアクティビティを検知するために最も有益な情報の一つがログである．Box では，あらゆる操作ログを取得できる機能を標準で有しているが(図 1)，現状，この標準機能だけで上記のような検知や分析を行うことは少々手間が掛かる．管理者が管理画面にログインし能動的にログを検索しなければならない点，全ユーザのログをエクスポートするためには相当の時間を要する点，高度な分析を行うためにはログ取得後に自身でスクリプト等を書く必要がある点，等が理由である．

そこで，今回我々はログデータ分析基盤である Splunk[2]を活用し，Box の利用状況を最小限の手間で迅速かつ高度に分析・可視化できるよう環境を整備したので，本稿にて報告する．

図 1. Box 標準のログ検索画面

2 Box と Splunk の連携方法

Box と Splunk の連携は非常に容易である．Splunk を標準的なスタンドアロン構成で構築した後，わずか数分程度の追加作業で Box との連携が可能となる．具体的には，Splunk 社から Splunk

Add-on for Box[3]と呼ぶアドオンが提供されているため,このアドオンをインストールしOAuth2.0認証を行えば，Splunk 上で Box のログを扱えるようになる．さらに今回，可視化やインシデントの検知といった高度な分析を行うため，マクニカネットワークス株式会社が提供する Splunk Enterprise 向け Box アドオン[4]を追加で使用した．また，本アドオンでは可視化のために Punchcard[5]，Wordcloud[6]と呼ぶアドオンを使うため，これらも併せてインストールした．これらのソフトウェアは Splunk を除き，すべて無料で利用可能である．ただし，本アドオンの利用には，マクニカネットワークスあるいはそのパートナー会社から Box を購入していることが条件となるので注意されたい．

参考まで，Splunk に取り込まれたデータ量は，本学環境の場合，連携初日のみ約 2.5 GB，それ以降は 1 日当たり約 100 MB 前後であった．今回，表 1 に示す一般的な物理サーバ 1 台のみで Splunk を構築したが，これまでのところ動作に問題は生じておらず，検索も高速に実行できている．使用した各種ソフトウェアのバージョンは表 2 の通りである．

表 1 Splunk 構築に用いたマシンスペック

OS	CentOS 6.7
CPU	Intel Xeon E3-1231 v3 (3.40GHz , 4 cores)
MEMORY	16 GB
HDD	SATA (2TB 7,200 rpm)

表 2 使用したソフトウェアのバージョン

Splunk Enterprise	6.4.0
Splunk Add-on for Box	1.2.0
Splunk Enterprise 向け Box アドオン	1.0
Punchcard	1.0.0
Wordcloud	1.11

3 Box の利用状況分析

Splunk を活用することで，様々な Box 上のアクティビティを可視化することができるが，本節では代表的だと考えられる分析例を掲載する．

3.1 地理別ログイン分布

どのような場所からログインされているのか，ログイン分布を地図上にプロットしたものを図 2 に示す．日本以外からの国からもある程度のアクセスがあるが，出張中の教員や，海外勤務中の職員，留学や旅行中の学生の利用ではないかと推測される．なお，図 2 中の黄色の丸をクリックすれば，ログインユーザの詳細を閲覧することも可能である(図 3)．これらを活用すれば，もし出張等で行くとは考え辛い地域からアクセスがあれば，アカウントが乗っ取られているのではないかと疑うことができ，不正ログイン検出の一手段として役立てられると考える．

3.2 ユーザ別イベント回数

Box 上で何らかの操作をした回数が多いユーザ上位 10 名を図 4 に示す．こうした図からパワーユーザを見つけることができれば，ユースケースをヒアリングしたり，周囲の人に利用を勧めてもらったりといった普及活動に繋げられる可能性がある．

3.3 イベント回数分布

曜日別にどの時間帯のイベント回数が多いかを示したものが図 5 である．本学職員の多くは 8:30-17:00 が勤務時間であり，その時間帯の利用が最も多くなっていた．一方で，土日や夜間帯でもある程度の利用があることが新たに判明した．

3.4 ファイル拡張子割合

Box にどのようなファイルがアップロードされているのか，その拡張子を表示したものが図 6 である．PDF や Office，画像関連の拡張子が最も多くなっているが，この他にも実に様々なファイルが保存されていることがわかる．容量無制限で利用できるため，中間ファイル等も含めたバックアップにも利用されていることが推測できる．このように可視化し定期的に確認することで，例えば，mp3 や mp4 のようなファイルが急激に増えた場合，違法なファイルをアップロードしているのではないか，といった推測を早期に立てることができる．

3.5 接続元アプリケーション

Box にどのようなアプリケーション・デバイスからアクセスしているのか, その上位 10 個を図 7 に示す．iPhone や iPad 等のモバイル端末からの利用が多くあることが判明したため，紛失や盗難に備え，Box のモバイル用アプリケーションにパスコードを設定するようユーザに呼びかける，といった広報案立案に役立てられると考える．

3.6 コラボレーション先ドメイン

どのようなドメインに対してコラボレーションの招待メールを送っているのか，その宛先上位 10 ドメインを図 8 に示す．gmail.com や yahoo.co.jp 等のフリーアドレスを除くと，keio.ac.jp や keio.jp 等，学内への共有が多いことが分かった．しかし，本学では Box のアカウントはすべて keio.jp というドメインに関連付いているため, 招待メールは必ず keio.jp のアドレスに送る必要があり, keio.ac.jp のアドレスに送っても上手く共有できない．本分析により，誤って keio.jp 以外のアドレスに招待を送っているユーザがかなり多くいることが判明したため，学内マニュアル等では，必ず keio.jp を宛先にするよう強調記載するよう改訂したい．こうした分析はマニュアルの改善にも役立てられることが分かった．

4 不正利用の兆候の検知

本節では，Box の不正利用の検知に役立つ手法について報告する．まず，どのようなアクティビティを不正利用と見なす，あるいは疑わしいと判断するかだが，例えば，一定期間の間に大量のファイルをダウンロードした，というアクティビティは有益な情報の一つだと考えられる．アカウントが乗っ取られ，大量に情報を持ち出されようとしているのではないか？と疑えるからである．例として，1 日に 50 回以上ファイルをダウンロードしたユーザを探す Splunk のサーチ文を図 9 に示す. Splunk のサーチ文を習得するには多少の慣れが必要ではあるが，図 9 のサーチ文をテンプレートとすれば，対象期間を縮める，対象操作をファイルの「削除」にする，閾値を 100 回にする，等の変更は容易に行える．サーチ文は管理者が任意のタイミングで実行することはもちろん，Splunk のアラート機能を使ってメールでリアルタイムに通知させることも可能である．

次に，不正利用の兆候として考えられる例として，不審な外部ドメインをコラボレータとして招待していないか，ということが考えられる．本学では Box 導入の最大の理由が外部との情報共有を活発化させることであるため，現時点では特定ドメインとのコラボレーションを禁止していないが，一般的な企業であれば有益な機能だと推測する．ここでは例として，gmail.com をコラボレーション先として招待したユーザを探すサーチ文を図 10 に示す．また，該当ユーザが発見された場合に Splunk から通知された実際のアラートメールを図 11 に示す.

不正利用の検知に関しては，現時点では上記のようなサーチを試験的に実行するのみに留まっているが，今後は，より複合的にユーザのアクティビティを観察し，実用に耐える仕組みを模索していきたい.

5 まとめと今後の展望

Box のようなクラウドストレージを利用するに当たり，有事の際に迅速にログを追跡できる環境を整えておくことは重要である．本学ではこうした有事に備え，Splunk を利用し，迅速かつ高度に Box の利用状況を把握する環境を整備した．これまで幸いにも，約 2 年間に及ぶ Box の運用においてデータ流失等の重大なインシデントは観察されておらず，今後も何も起こらないことを祈るばかりであるが，万一の有事の際には，Splunk を活用し迅速に対処していきたいと考える．また，インシデント発生時に限らず，Splunk を利用することで，Box の利用状況を詳細に把握することができるようになったため，利用者へのよりきめ細やかなサポートに繋げられるよう今後活用していきたい.

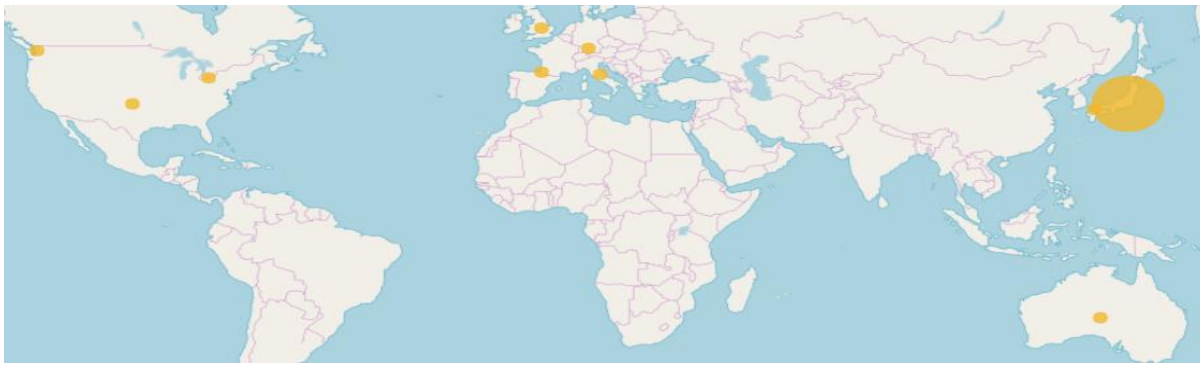

図 2. 地理別ログイン分布

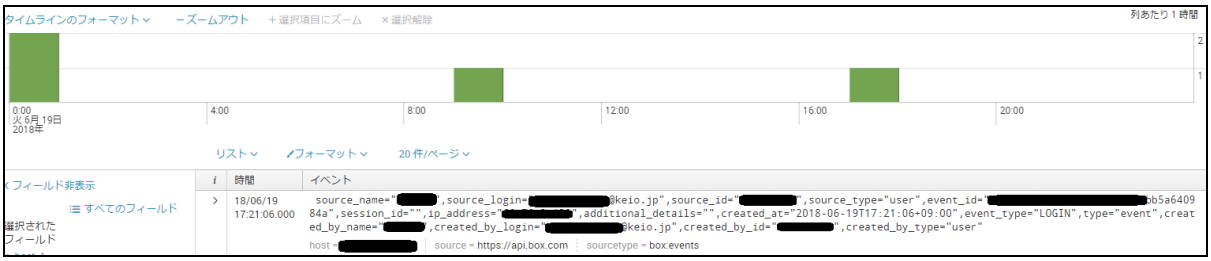

図 3. ログインユーザの詳細

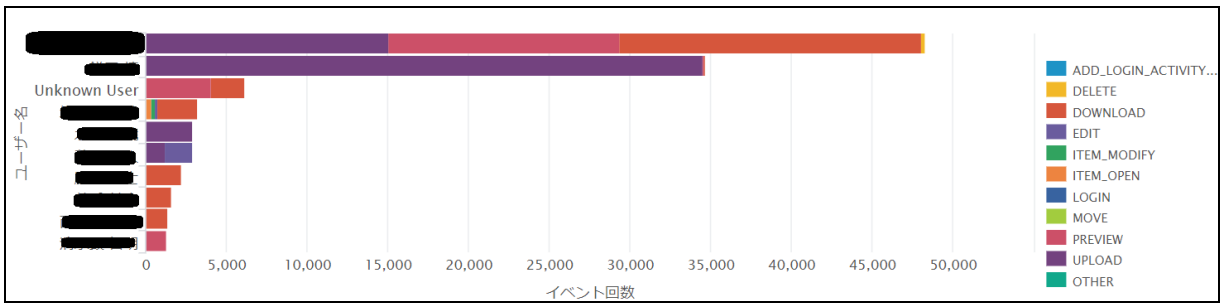

図 4. イベント回数が多いユーザ上位 10 名

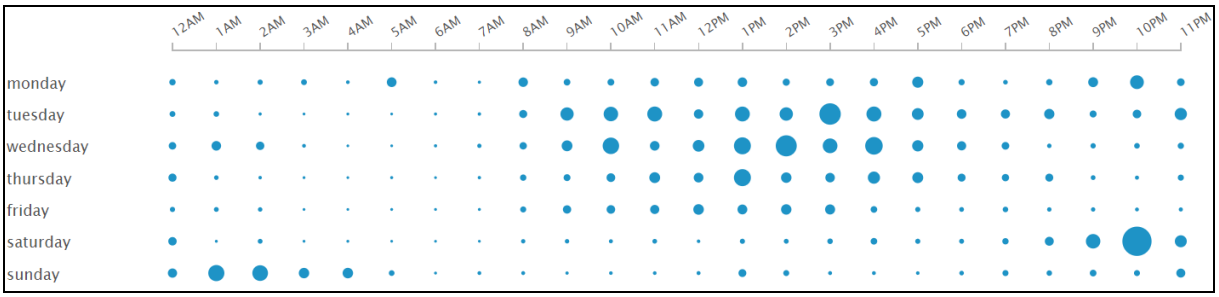

図 5. 曜日・時間帯別イベント回数分布

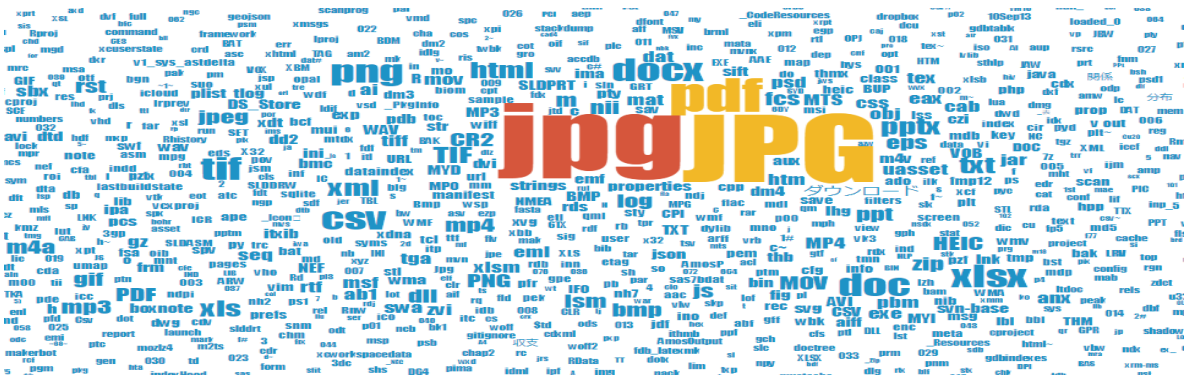

図 6. ファイル拡張子割合

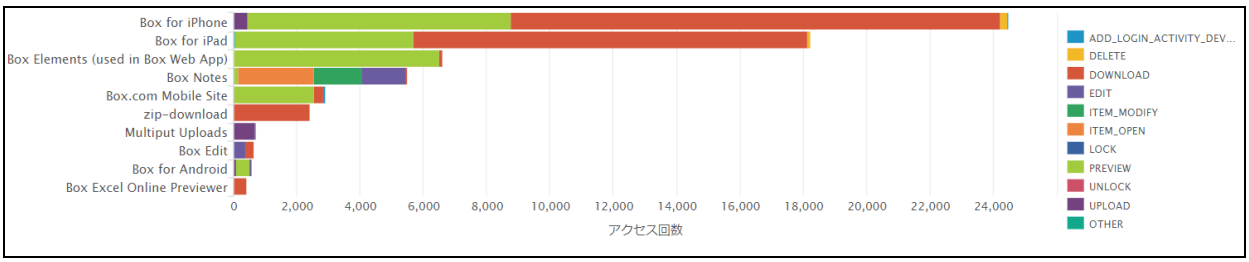

図 7. 接続元アプリケーション上位 10 個

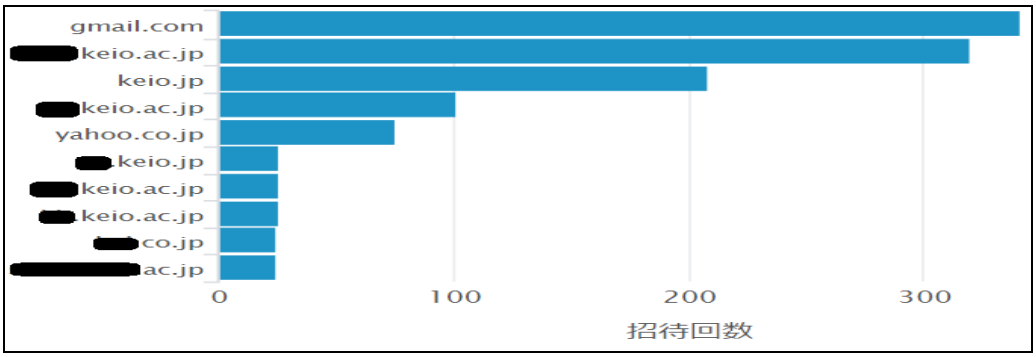

図 8. コラボレーション(本学からの招待)の相手ドメイン上位 10 個

```
sourcetype="box:events" event_type=DOWNLOAD
| bin span=1d _time
| stats count by _time created_by_name created_by_login
| search count > 50
| sort - count
| rex field=created_by_login ".*¥@(?<user_domain>.*)"
| fillnull value=NULL
| rename count AS 回数, created_by_name AS ユーザー名, user_domain AS 所属
| table _time,ユーザー名,所属,回数
```

図 9. 1 日に 50 回以上ファイルをダウンロードしたユーザを探す Splunk サーチ文

```
sourcetype="box:events" event_type=COLLABORATION_INVITE
| rex field=created_by_login ".*¥@(?<from_maildomain>.*)"
| rex field=accessible_by_login ".*¥@(?<to_maildomain>.*)"
| search from_maildomain = "gmail.com" OR to_maildomain = "gmail.com"
| stats count values(accessible_by_name) AS accessible_by_name values(created_by_name) AS
created_by_name by _time to_maildomain from_maildomain accessible_by_login
| sort - count
| eval date=strftime(_time, "%Y/%m/%d")
| rename date AS 日時
| rename count AS コラボレーション回数
| rename created_by_name AS 招待した人
| rename accessible_by_name AS 招待された人
| rename to_maildomain AS 招待された人のドメイン
| rename from_maildomain AS 招待した人のドメイン
| table 日時,招待した人,招待した人のドメイン,招待された人,招待された人のドメイン,コラボレーション回数
```

図 10.特定ドメイン(例 gmail.com)と外部コラボレーションしたユーザを探す Splunk サーチ文

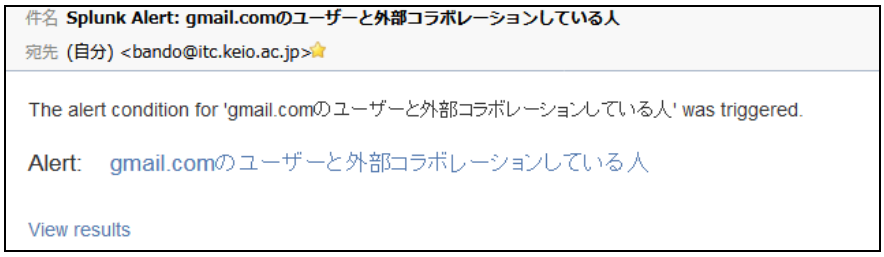

図 11. 実際に Splunk から通知されたメールの例

謝辞

本稿の執筆にあたり，マクニカネットワークス株式会社の床田紘美様から多大なるご支援とご助言を頂きました．この場を借りて厚く御礼申し上げます.

参考文献

[1] “Box 公式ウェブサイト”, https://www.box.com/ja-jp/home
[2] “Splunk 公式ウェブサイト”, https://www.splunk.com/ja_jp
[3] “Splunk Add-on for Box”, https://splunkbase.splunk.com/app/2679/
[4] “Splunk Enterprise 向け Box アドオンを無償提供開始”, https://www.macnica.net/box/
[5] ”Punchcard”, https://splunkbase.splunk.com/app/3129/
[6] “Wordcloud”, https://splunkbase.splunk.com/app/3212/