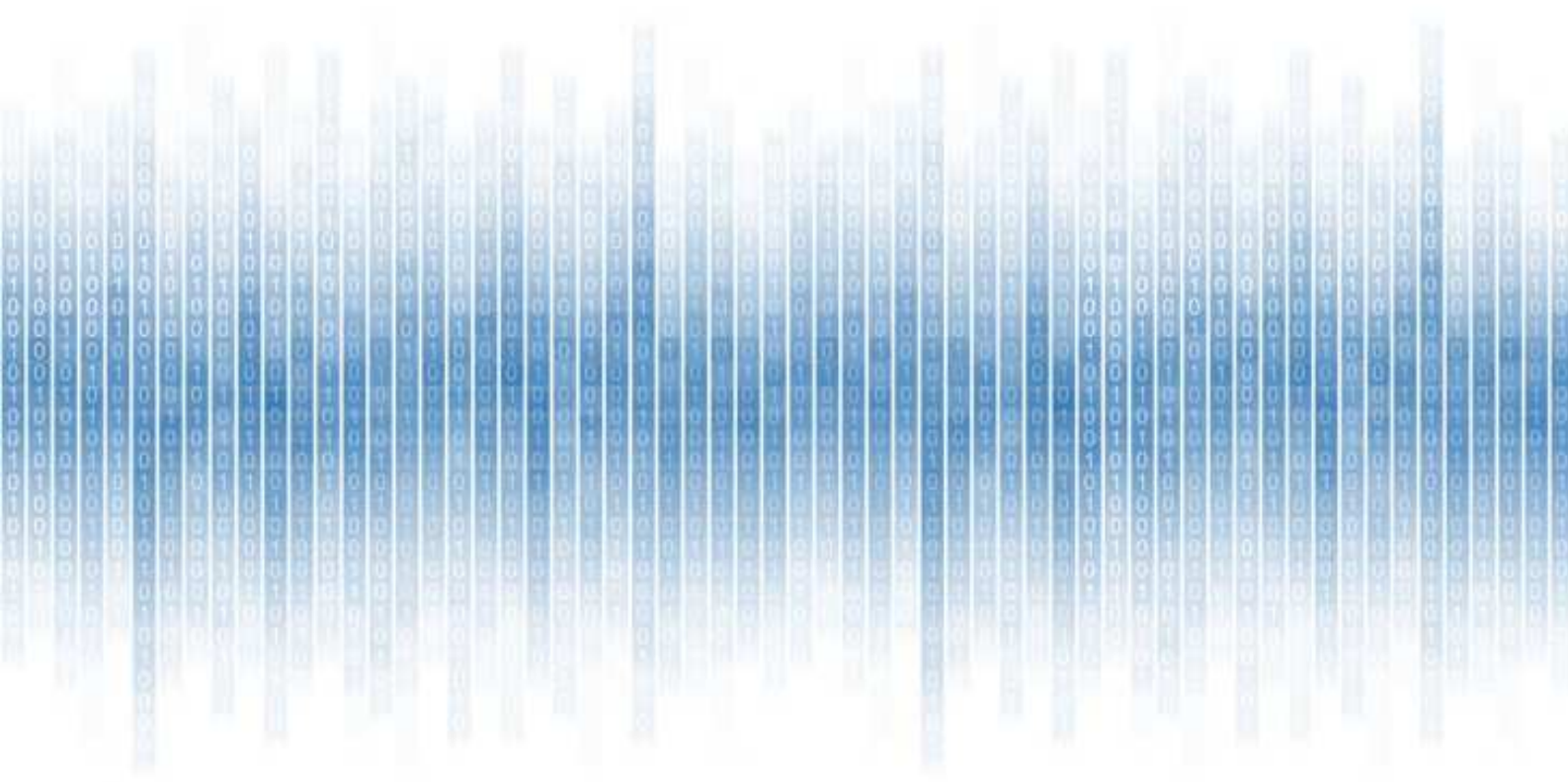

クラウドセキュリティアライアンス

ビッグデータワーキンググループ

ビッグデータのセキュリティ／プライバシーにおける十大脅威

（日本語訳）

2014 年 1 月

© 2012 Cloud Security Alliance – All Rights Reserved.

All rights reserved. You may download, store, display on your computer, view, print, and link to the Cloud Security Alliance Big Data Top Ten at http://www.cloudsecurityalliance.org/, subject to the following: (a) the Document may be used solely for your personal, informational, non-commercial use; (b) the Document may not be modified or altered in any way; (c) the Document may not be redistributed; and (d) the trademark, copyright or other notices may not be removed. You may quote portions of the Guidance as permitted by the Fair Use provisions of the United States Copyright Act, provided that you attribute the portions to the Cloud Security Alliance Big Data Top Ten (2012).

Acknowledgements

CSA Big Data Working Group Co-Chairs

Lead: Sreeranga Rajan, Fujitsu
Co-Chair: Wilco van Ginkel, Verizon
Co-Chair: Neel Sundaresan, eBay

Contributors

Alvaro A. Cárdenas Mora, Fujitsu
Yu Chen, SUNY Binghamton
Adam Fuchs, Sqrrl
Adrian Lane, Securosis
Rongxing Lu, University of Waterloo
Pratyusa Manadhata, HP Labs
Jesus Molina, Fujitsu
Praveen Murthy, Fujitsu
Arnab Roy, Fujitsu
Shiju Sathyadevan, Amrita University

CSA Global Staff

Aaron Alva, Graduate Research Intern
Luciano JR Santos, Research Director
Evan Scoboria, Webmaster
Kendall Scoboria, Graphic Designer
John Yeoh, Research Analyst

© Copyright 2012 Cloud Security Alliance. All rights reserved.

日本語訳の提供について

「ビッグデータのセキュリティ／プライバシーにおける十大脅威」は、Cloud Security Alliance Big Data Working Group よりリリースされている「Top Ten Big Data Security and Privacy Challenges」（2012 年 11 月）の日本語訳です。このドキュメントは、ビッグデータセキュリティに関心のあるクラウドユーザーの教育・啓発を目的として、原文をそのまま翻訳したものであり、日本独自の法令や基準に関する記述は含まれておりません。

なお、日本クラウドセキュリティアライアンスに関する情報は、以下の URL より参照可能ですので、ご覧下さい。

http://www.cloudsecurityalliance.jp/

このドキュメントは、以下の日本クラウドセキュリティアライアンスの有志により作成されています。

日本クラウドセキュリティアライアンス・ビッグデータユーザーワーキンググループ

リーダー：笹原 英司 （特定非営利活動法人ヘルスケアクラウド研究会　医薬学博士）

阿倍 克英 （特定非営利活動法人ヘルスケアクラウド研究会）

里中 慧 （特定非営利活動法人ヘルスケアクラウド研究会）

協力者

諸角 昌宏（マカフィー株式会社）

© Copyright 2012 Cloud Security Alliance. All rights reserved.

目次

© Copyright 2012 Cloud Security Alliance. All rights reserved.

1.0 要約

セキュリティ／プライバシーの脅威は、大規模なクラウドインフラストラクチャ、データソースやフォーマットの多様性、データ収集の流動的な性質、大容量のクラウド間の移動など、ビッグデータの速度、容量、多様性によって増幅される。従って、小規模の静的（流動的に相対する）データをセキュアにするために構築された伝統的なセキュリティのメカニズムでは不十分である。本稿では、ビッグデータに特有なセキュリティ／プライバシーの十大脅威を取り上げる。これらの脅威を取り上げることによって、ビッグデータインフラストラクチャの強化が改めて注目されることを期待する。

2.0 はじめに

ビッグデータという用語は、企業や政府機関が、我々や我々の周辺から収集している大容量のデジタル情報を指している。毎日、我々は 250 京バイトのデータを生成しており、今日世界にあるデータの 90%は最近 2 年間で生成されたものである。セキュリティ／プライバシーの脅威は、大規模なクラウドインフラストラクチャ、データソースやフォーマットの多様性、データ収集の流動的な性質、大容量のクラウド間の移動など、ビッグデータの速度、容量、多様性によって増幅される。大規模なクラウドインフラストラクチャの利用が、大規模なコンピューターネットワーク上で、多様なソフトウェアプラットフォームで広がると同時に、システム全体の攻撃対象領域が増大する。

小規模の静的（流動的に相対する）データをセキュアにするために構築された伝統的なセキュリティのメカニズムでは不十分である。例えば、アノマリ検知の分析は、大量の異常値を生成するかもしれない。同様に、既存のクラウドインフラストラクチャ上で来歴を追加する方法は明らかでない。ストリーミングデータは、セキュリティ／プライバシーのメカニズムからの超高速のレスポンス時間を必要とする。

本稿では、ビッグデータに特有のセキュリティ／プライバシーの十大脅威を取り上げる。我々は、重大なセキュリティ/プライバシーの問題に関するドラフトの最初のリストを作成するためにクラウドセキュリティアライアンスの会員にインタビューしセキュリティ専門家向けの業界誌にサーベイを行った。また、公開されている研究資料の調査を行った。その結果、以下のような十大脅威に至った：

1. 分散プログラミングフレームワークにおけるセキュアな計算処理
2. ノンリレーショナルデータストアに対するセキュリティのベストプラクティス
3. セキュアなデータ保存とトランザクションのログ
4. エンドポイントの入力の検証／フィルタリング
5. リアルタイムのセキュリティ／コンプライアンスモニタリング
6. 拡張性があり構成可能なプライバシー保護データマイニング／分析

© Copyright 2012 Cloud Security Alliance. All rights reserved.

7. 暗号化により強化されたアクセス制御とセキュアな通信

8. 粒度の高いアクセス制御

9. 粒度の高い監査

10. データ来歴

本稿では、以下で簡単に説明し、ユースケースを紹介する。

3.0 分散プログラミングフレームワークにおけるセキュアな計算処理

分散プログラミングフレームワークでは、大容量データを計算して保存するために並列処理を利用する。典型的な例は MapReduce フレームワークであり、入力ファイルを複数のチャンク（かたまり）に分割する。MapReduce の最初のフェーズでは、個々のチャンクの Mapper がデータを読み込み、一定の計算処理を行って、鍵と値のペアのリストを出力する。次のフェーズでは、Reducer が個々の鍵に附属する値を結びつけて、結果を出力する。主な攻撃防止手段としては、Mapper のセキュア化と、信頼できない Mapper に存在するデータのセキュア化の 2 種類がある。

3.1 ユースケース

信頼できない Mapper が誤った結果を返す可能性があり、それによって不正確な集約結果を生成する。大規模なデータセットの場合、判別することは不可能に近く、特に科学／金融計算においては重大な損害を生む結果となる。

小売業者の消費者データは、ターゲット広告や顧客セグメンテーションのためにマーケティング代理店が分析することがよくある。これらの作業には、大規模のデータセット上での高度な並列処理が含まれており、特に Hadoop のような MapReduce フレームワークに適している。しかしながら、データの Mapper に、意図的若しくは意図的でない漏えいが含まれる可能性がある。例えば、Mapper が、プライベートな記録を分析し、特別な値を外に出して、ユーザーのプライバシーを侵害する可能性がある。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

4.0 ノンリレーショナルデータストアに対するセキュリティのベストプラクティス

NoSQL によって普及したノンリレーショナルデータストアは、セキュリティインフラストラクチャに関しては、まだ進化の途上にある。例えば、NoSQL インジェクション向けの堅牢なソリューションは未成熟である。個々の NoSQL DB は、分析の世界から提示された異なる課題に取り組むよう構築されており、それゆえ設計段階のいかなる時点においても、モデルの一部となることはなかった。NoSQL データベースを利用する開発者は、通常、ミドルウェアにセキュリティを組み込んできた。NoSQL データベースは、データベースの中で明確にそれを強化するためのサポートを提供していない。しかしながら、NoSQL データベースにおけるクラスタの観点は、このようなセキュリティプラクティスの堅牢性に対する追加的な課題を示している。

4.1 ユースケース

大規模な非構造化データセットを取り扱う企業は、巨大な容量のデータを収納／処理する点に関して、伝統的なリレーショナルデータベースを NoSQL データベースに移植することの恩恵を得る可能性がある。一般的に、NoSQL データベースのセキュリティの思想は、外部の強化された機能に依存している。セキュリティインシデントを減らすために、企業は、ミドルウェアのセキュリティポリシーを見直して、エンジンに項目を追加すると同時に、運用の機能で妥協することなしに RDB に対抗できるように NoSQL データベース自身を強化する必要がある。

5.0 セキュアなデータ保存とトランザクションのログ

データとトランザクションのログは、多層のストレージメディアに保存される。手動で各層間をデータ移動させることは、IT マネージャーにどのデータがいつ移動されたかを直接コントロールさせることになる。しかしながら、データセットの容量は指数関数的に増加し続けており、拡張性と可用性のためにビッグデータストレージ管理の自動階層化が求められる。自動階層化ソリューションは、どこにデータが保存されるか、どれがセキュアなデータ保存の新たな脅威となるかを追跡することはない。新たな機能として、権限のないアクセスを遮断し、常時、可用性を維持することが必須となる。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

5.1 ユースケース

ある製造企業は、様々な部門からのデータを統合したいと考えている。このデータの中にはほとんど引き出されないものがある一方、同じデータプールを継続的に利用する部門もある。自動階層化ストレージシステムは、ほとんど利用されないデータをより下位の（安い）層に格納することによって、製造企業の経費を節約するであろう。しかしながら、このデータには、一般的ではないが重要な情報を含む研究開発結果が含まれている可能性もある。下位層ではしばしば低いセキュリティが提供されることがあるので、企業は、慎重に階層化戦略を研究すべきである。

6.0 エンドポイントの入力の検証／フィルタリング

企業環境におけるビッグデータのユースケースの多くで、エンドポイントデバイスなど様々なソースからのデータ収集が要求される。例えば、セキュリティ情報イベント管理システム（SIEM）は、企業ネットワーク上にある数百万のハードウェアデバイスやソフトウェアアプリケーションからイベントログを収集する可能性がある。データ収集プロセスにおける重要な課題として、入力の検証がある。どのようにしてデータを信頼することができるのか？ どのようにして入力データのソースが悪意のないことを検証でき、どのようにして収集物から悪意のある入力をフィルタリングすることができるのか？ 入力の検証とフィルタリングは、特に Bring Your Own Device （BYOD）モデルなど、信頼できない入力ソースにより引き起こされる手強い課題である。

6.1 ユースケース

気象センサーから収集されたデータや、iPhone アプリケーションから送信されたフィードバックの投票には、同じような検証上の問題が存在する。動機付けられた相手が、「不正を働く」仮想的なセンサーを生成したり、結果を偽装するために iPhone の ID になりすましたりすることが可能かもしれない。これが、収集したデータの容量によって一層複雑化して、読み込みデータ数／投票が数百万を超える可能性がある。これらの業務を効率的に実行するためには、大規模なデータセットの入力を検証するアルゴリズムを生成する必要がある。

7.0 リアルタイムのセキュリティ／コンプライアンスモニタリング

（セキュリティ）デバイスによって数多くの警告が生成されると、リアルタイムのセキュリティモニタリングが常に問題となってくる。これらの警告は（相関関係の有無に関わらず）大量の誤検知につ

© Copyright 2012 Cloud Security Alliance. All rights reserved.

ながり、取り出した量を人間が処理できなくなると、大抵無視されるか単にクリックされるだけになる。この問題は、データの流れの容量や速度によって、ビッグデータと共に増大する可能性がある。しかしながら、ビッグデータ技術は、これらの技術が異なるタイプのデータの高速な処理・分析を可能にするという意味で、機会をもたらす可能性もある。その出番になった時、例えば拡張性のあるセキュリティ分析に基づいてリアルタイムのアノマリ検知を提供するために利用することが可能である。

7.1 ユースケース

ユースケースによって異なる可能性があるが、産業や政府（機関）の大半はリアルタイムセキュリティ分析から恩恵を受ける。「誰が、どのデータに、どのソースから、いつアクセスしているのか」「攻撃を受けているのか」「A という行動のせいで、コンプライアンス基準 C に違反していないか」など、共通のユースケースが存在する。実際、これらは新しいものではないが、その分野に関して、より速く、より良い意思決定（例．誤検知の減少）を実行するために、自由に扱うことができるより多くのデータを有することが異なる点である。しかしながら、新しいユースケースが定義されたり、ビッグデータの代わりに既存のユースケースを我々が再定義したりすることが可能である。例えば、健康医療産業の場合、潜在的に納税者向けに高額のお金を節約したり、請求の支払がより正確になったり、請求に関連する不正を削減したりするなどして、ビッグデータの恩恵を受ける。しかしながら同時に、保存された記録は非常に機微であり、同一データの慎重な保護を要求する HIPAA あるいはその他の地域／地方の規制を遵守しなければならない可能性がある。意図した若しくは意図しない個人情報の異常な取得をリアルタイムに検知することによって、医療機関は、発生した損害を迅速に修復し、さらなる誤使用を防止することが可能となる。

8.0 拡張性があり構成可能なプライバシー保護データマイニング／分析

ビッグデータは、潜在的にプライバシーの侵害、侵略的なマーケティング、市民の自由の制限、国家や企業によるコントロールの増大を可能にする独裁者のトラブルの兆候と見なされる可能性がある。

最近、企業のマーケティングを目的としたデータ分析の活用方法に関する分析により、どのようにして、当人の父親が知る前に十代の若者が妊娠したことを小売事業者が確認することができるかが事例として示された。同様に、ユーザーのプライバシーを保持するために、分析用データの匿名化だけでは十分でない。例えば AOL は、学術目的で匿名化された検索ログを公表したが、その検索者によって簡単にユーザーが特定された。Netflix は、同社の映像スコアを IMDB のスコアで修正することで匿名化したデータセットのユーザーが特定されてしまった時、同様の問題に直面した。このようなことから、意図しないプライバシーの公開を防止するためのガイドラインや推奨を策定することが重要である。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

8.1 ユースケース

企業や政府機関によって収集されたユーザーデータは、内部の分析者、場合によっては外部委託先やビジネスパートナーによって継続的にマイニング／分析される。悪意のある内部関係者や信頼できないパートナーが、これらのデータセットを悪用して、顧客からプライベートな情報を抜き出すことは可能である。同様に、諜報機関は膨大な量のデータの収集を必要とする。データソースは多岐に渡り、チャットルーム、個人のブログやネットワークルーターが含まれる可能性がある。しかしながら、収集されたデータの大半は元来悪意のないものであり、保存したり　匿名化を維持したりする必要はない。

堅牢で拡張性のあるプライバシー保護マイニングアルゴリズムによって、適切な情報を収集し、ユーザーの安全性を高める機会が増えるであろう。

9.0 暗号化により強化されたアクセス制御とセキュアな通信

最も機微な個人データが、エンドツーエンドでセキュアであり、権限を有する主体だけがアクセスできることを保証するためには、データがアクセス制御ポリシーに基づいて暗号化されている必要がある。属性ベース暗号（ABE）など、この分野に特化した研究を一層充実し、効率的で拡張性のあるものにする必要がある。分散した主体間で認証や同意、公平性を保証するためには、暗号化によるセキュアな通信フレームワークが導入される必要がある。

9.1 ユースケース

機微なデータは、日常的にクラウド上に暗号化されていない状態で保存されている。データ、特に大規模のデータセットを暗号化する際の大きな問題は、暗号化されたデータに対するオール・オア・ナッシングの検索ポリシーであり、ユーザーは、記録や検索の共有のようなきめの細かいアクションを簡単に遂行できなくなる可能性がある。ABE は、暗号化されたデータに関連する属性が鍵の解除に利用されるところで、公開鍵の暗号システムを活用することによってこの問題を軽減する。他方、我々は、分析に有用なデータなど、暗号化されていない、比較的機微でないデータも保有している。このようなデータは、暗号化によりセキュアな通信フレームワークを利用して、セキュアかつ取り決めされた方法により伝達される必要がある。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

10.0 粒度の高いアクセス制御

アクセス制御の観点から問題となるセキュリティの特性は機密性であり、アクセスすべきでない人によるデータへのアクセスを抑制することである。過程の細かいアクセスメカニズムの問題は、そうでなければ共有されたであろうデータが、目に見えるセキュリティを保証するために、より厳格な分類へと排除されることがよくある点である。粒度の高いアクセス制御によって、機密性に妥協することなく可能な限りデータを共有する剣の代わりとなるメスがデータ管理者に付与される。

10.1 ユースケース

ビッグデータ分析やクラウドコンピューティングでは、次第に、スキーマの量およびセキュリティ要件の量で膨大なデータセットを処理する点に注目が集まっている。データに関する法律およびポリシーの制限は、様々なソースに起因している。サーベンス・オクスリー法（SOX 法）は企業の財務情報を保護するための要件を設けており、医療保険の相互運用性と説明責任に関する法律(HIPAA）は、個人健康記録の共有に関する様々な制限を含んでいる。米国機密保護法の大統領令13526（Executive Order 13526）は、国家安全情報保護のための精巧なシステムを概説している。また、プライバシーポリシーや共有同意書、企業ポリシーも、データの取り扱いに関する要件を示している。この過度な制限を管理することによって、アプリケーション開発費用の増大や、誰もほとんど分析に参加できない壁に囲まれた庭のようなアプローチを招く結果となった。粒度の高いアクセス制御は、この急激に複雑化するセキュリティ環境に、分析システムを適合させるために必要である。

11.0 粒度の高い監査

リアルタイムのセキュリティモニタリング（12 章参照）を利用して、我々は攻撃が起きた瞬間に通知されるよう試みている。実際には、これがいつも当てはまるとは限らない（例．最新の攻撃、本当は正しいのに見落とされた場合）。見落とされた攻撃の真相を究明するためには、我々は監査情報が必要である。これは、何が起きて、何を誤ったのかを理解するためだけでなく、コンプライアンスや法規制、フォレンジックの理由からも重要である。そのような観点から監査は目新しいものではないが、適用範囲や粒度が異なることがある。例えば、我々はより多くのデータオブジェクトを処理しなければならないが、それらは（必ずしも必要ではないが）分散されている可能性がある。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

11.1 ユースケース

コンプライアンス要件（例．HIPAA、PCI、SOX 法）は、金融機関に対し、粒度の高い監査記録を要求する。加えて、プライベートな情報を含む記録の損失は 1 件当たり 200 ドルと推計されている。地理的な場所にもよるが、データ違反の場合、その後に法的手続が取られる可能性がある。金融機関の主要人員は、社会保障番号（SSN）など個人情報（PI）を含む大規模なデータセットへのアクセスを要求する。例えば、マーケティング企業の場合、オンライン広告に関する顧客中心の手法を最適化するために、個人のソーシャルメディア情報にアクセスしたいと考える。

12.0 データ来歴

来歴を可能にするビッグデータアプリケーションのプログラミング環境から生成される大規模な来歴グラフにより、来歴のメタデータは複雑化していく。メタデータのセキュリティ／秘密性アプリケーションへの依存度を検知するために行うこのような大規模の来歴グラフ分析はコンピュータ処理上集中的なものになる。

12.1 ユースケース

いくつかの主要なセキュリティアプリケーションは、生成に関する詳細など、デジタル記録の履歴を要求する。例えば、金融機関のインサイダートレーディングの検知や、研究調査のためにデータの正確性を決定する場合がある。これらのセキュリティ評価は元々時間に厳格であり、この情報を含む来歴のメタデータを処理するために、速いアルゴリズムを要求する。加えて、データ来歴は、PCI、SOX 法など、コンプライアンス要件のための監査ログを補完するものである。

13.0 結論

ビッグデータが定着している。実際それなしで、データを消費し、新たな形態のデータを生成し、データ主導のアルゴリズムを含む次世代のアプリケーションを想像することはできない。コンピュータ環境がより安価になり、アプリケーション環境がネットワーク化され、システム／分析環境がクラウド上で共有されると共に、システマティックな方法で対応しなければならない脅威として、セキュリティやアクセス制御、圧縮、暗号化、コンプライアンスが挙がっている。クラウドセキュリティアライアンス（CSA）のビッグデータワーキンググループ（BDWG）は、これらの脅威を認識し、標準化されたシステマティックな方法で対処することをミッションとしている。

© Copyright 2012 Cloud Security Alliance. All rights reserved.

本稿で我々は、ビッグデータの処理およびコンピューティングのインフラストラクチャをよりセキュアなものにするために対処することが必要なセキュリティ／プライバシーの十大脅威を取り上げてきた。これらビッグデータ特有の上位十項目リストに共通な要素の中には、複数のインフラストラクチャ層（ストレージおよびコンピューティング）の利用、セキュリティ問題の観点から完全に精査されてこなかった NoSQL データベース（ビッグデータの容量により高速のスループットが要求される）など、新たなコンピューターインフラストラクチャの利用、大規模データセット向けの暗号化における拡張性の欠如、小容量データ向けには実用的なリアルタイムモニタリング技術における拡張性の欠如、データを生成するデバイスの多様性、そして、プライバシー／セキュリティを確実なものにするために個別のアプローチにつながる様々な法務／ポリシーの過剰な制限による混乱から生じるものがある。十大脅威リストの項目の多くは、このようなタイプの脅威を分析するためにビッグデータ処理インフラストラクチャ全体の攻撃対象領域に特有な点を明らかにするのに役立つ。我々は、OpenMobius という、オープンソースで大容量、分散化されたデータ処理、分析、ツールのプラットフォームを、eBay 研究所より実証テストベッドとして利用する計画である。

我々は、本稿が、研究開発コミュニティにおいて共同研究的に十大脅威への注目を高める行動を促進し、ビッグデータプラットフォームにおけるセキュリティ／プライバシーの拡大につながることを期待する。

© Copyright 2012 Cloud Security Alliance. All rights reserved.