脆弱性対応意思決定支援システム VRDA (Vulnerability Response Decision Assistance) の有効性検証

調査責任者：

Art Manion, CERT® Coordination Center

富樫一哉, JPCERT/CC

調査協力者：

高坂史彦, JPCERT/CC

山口将則, IIJ Technology Inc.

Shawn McCaffrey, Carnegie Mellon University

Jay Kadane, Carnegie Mellon University

Chris King, Carnegie Mellon University

Robert Weiland, Carnegie Mellon University

2009 年 10 月

制作：

JPCERT CC

Software Engineering Institute | Carnegie Mellon

目次

概要

VRDA（Vulnerability Response Decision Assistance：脆弱性対応意思決定支援システム）は、各組織が行う脆弱性への対応に関する意思決定をモデル化し、脆弱性対応の意思決定を支援するエキスパートシステムである。

各組織の脆弱性対応ルールを VRDA が提唱する手法で体系的に整理し、VRDA の動作パラメータとしてエンコード化することにより、組織の脆弱性対応意思決定がより一貫したものになり、また脆弱性対応の優先度をより適切に決められるようになる。VRDA は、観察された事実を基に脆弱性への対応をモデル化し、このモデルを基に組織が実際にどのように脆弱性に対応するかを再現する。

本文書では、VRDA により提示された対応予想がどの程度正確であったかという観点から VRDA の有効性を調査する。各組織の分析担当者による対応判断と比較して、VRDA が提示した対応の正確性を評価するために、3 つの参加組織から脆弱性への対応実績データを入手し分析した。脆弱性情報の収集、意思決定モデルの生成、意思決定モデルに基づく対応予想の提示、および実際に取られた対応判断の記録には、VRDA の実装の 1 つである KENGINE が使用された。

KENGINE により提示された対応予想と実際の対応判断との間の誤差は、必要あるいは十分な脆弱性データの欠落、分析担当者のバイアス、意思決定論理上の欠陥、その他、予期しない理由によって生じることが考えられる。KENGINE により提示された対応予想の正確性は個々の脆弱性対応作業によりばらつきはあるものの、異なる組織、データセット（脆弱性情報、脅威分析項目、脆弱性対応作業）、意思決定モデルの間で比較を行った結果、VRDA の正確性が脆弱性対応を支援するのに十分に実用的であることが明らかになった。

はじめに

ソフトウェアの脆弱性は、暗黙的または明示的なセキュリティポリシーに違反する一連の状態（通常は設計または実装の欠陥）であると定義できる[1]。例えば、バッファオーバーフローの欠陥は、攻撃者による任意のコードの実行を許す可能性がある。このような振る舞いは、ユーザ（攻撃者）が許可なく任意のアクションを実行できてはならないという暗黙のセキュリティポリシーに対する違反となりえる。毎年、何千もの脆弱性が報告されている[2]。システム管理者、ソフトウェア開発者、およびその他の人々は、これら脆弱性に対応するために限られたリソースをいかに最適に使用するかという課題に直面している。

VRDA（Vulnerability Response Decision Assistance：脆弱性対応意思決定支援システム）は、各組織が行う脆弱性への対応に関する意思決定をモデル化し、脆弱性対応の意思決定を支援するエキスパートシステムである。

VRDA は、脆弱性の特徴含む脅威分析情報と、実際に行った対応に関する情報に基づいて、脆弱性への対応予想を提示するディシジョンツリーを生成する。CERT Coordination

[1] http://www.cert.org/encyc_article/#IntVul
[2] http://www.cert.org/stats/#vuls

Center（CERT/CC）[3]および JPCERT/CC[4]の脆弱性分析チームは、脆弱性の分析、優先順位付け、および対応に関する自らの経験に基づいて VRDA を開発した。なお、本文書は、読者がソフトウェアの脆弱性についての知識を持ち、また、VRDA に精通していることを前提としている。より詳細な VRDA の説明については、ハル・バーチ（Hal Burch）、アート・マニオン（Art Manion）、および伊藤友里恵による『VRDA（Vulnerability Response Decision Assistance）』で述べている[5]。

記述的なシステムである VRDA は、組織の脆弱性対応実績を基にその組織の脆弱性対応意思決定ルールをモデル化する。VRDA は、このモデルを基に対応予想を提示するが、提示内容の正否判断はそれぞれの組織に委ねられている。また付加的な VRDA の利用効果として、各組織の脆弱性対応ルールを VRDA が提唱する手法で体系的に整理する過程で、組織の脆弱性対応プロセスの見直しが促進され、改善すべき点が明らかになることも多い。

VRDA は、脆弱性対応業務を行う担当者向けに設計されたものであるが、これまで広くは使用されておらず、運用現場における検証も十分行われていない。VRDA で使用されているコンセプトの有効性を検証し、その実装である KENGINE をテストするために、我々は、脆弱性対応業務へ定常的に取り組んでいる次の 3 つの組織とともに調査を行うこととした。

参加組織 A：（名称非公開）

参加組織 B：IIJ Technology[6]

参加組織 C：CERT/CC の脆弱性分析チーム[7]

VRDA が有効であるためには、現実の世界における組織が行う脆弱性への対応を VRDA が正確に提示する必要がある。本調査では、VRDA により提示された対応予想と実際の対応判断との誤差を分析することにより、VRDA の正確性に焦点を当てた。

VRDA による作業効率の向上は、脆弱性対応の意思決定に必要な労力を、対応業務の品質を下げずに軽減することで示される。しかし、VRDA は意思決定に必要な労力を削減するものと予想されるが、本調査において労力の変化は測定していない。CERT/CC における関連コスト見積り（付録 E）は、VRDA がリソースの割り当てに関してどのように利用可能かを示す一例ではあるが、VRDA を利用していない場合の労力と VRDA を利用した場合の労力の比較は行われていない。

本調査の参加組織についての情報と、組織別の VRDA が提示した対応予想の正確性の結果は、調査結果を総括する結論の章に要約されている。

調査方法論

JPCERT/CC は、KENGINE と呼ばれる VRDA の実装を Ruby on Rails と PostgreSQL を用いて開発した[8]。本調査への参加組織は KENGINE を使用して、脆弱性の脅威分析情報と

[3] http://www.cert.org/
[4] http://www.jpcert.or.jp/
[5] http://www.jpcert.or.jp/research/2008/2007VRDA_paper_JP.pdf
[6] http://www.iij-tech.co.jp
[7] http://www.cert.org/vuls/
[8] http://www.jpcert.or.jp/research/2008/20071212KENGINE.pdf

その対応判断を入力し、脆弱性への対応意思決定ルールをモデル化したディシジョンツリーを生成した。ディシジョンツリーを生成することにより、対応予想を提示可能となった KENGINE からそのデータをエクスポートし、これらデータを Microsoft Excel および Weka を使って分析した[9]。また、参加組織は、さまざまなデータ管理用のスクリプト、および KENGINE の XML データインポート/エクスポート機能を使用してデータの投入、抽出、加工、分析を行った。なお、JPCERT/CC は脆弱性の分析情報を定型フォーマットで提供する VRDA フィード（Atom フィード）を公開している[10]。

調査期間中に KENGINE ソフトウェアに加えられた変更により、ブール値（値は 0 か 1）で表現された二択の分析項目を含むディシジョンツリーの作成方法が改善され、生成されたディシジョンツリーに不正なノードが含まれてしまう問題が解決した。本調査では、これらの変更を含むバージョンの KENGINE が用いられた。

トレーニングデータ

調査に用いられたデータは、１）脆弱性の特徴など脅威を分析した情報、２）実際の対応判断、および、３）VRDA により提示された対応予想、という 3 つに分類することができる。このうち、１）脆弱性の特徴など脅威を分析した情報と、２）実際の対応判断についての情報が、VRDA により組織の意思決定ルールをモデル化するためのトレーニングデータとなる。VRDA はこれらのトレーニングデータを入力として生成されたディシジョンツリーに基づき対応予想を提示する。参加組織 A と IIJ Technology は、実際に行っている脆弱性対応業務から得られたトレーニングデータ、つまり実際の脆弱性への対応実績データを使用した。一方、CERT/CC が利用したトレーニングデータのほとんどは過去に報告された脆弱性に対して CERT/CC が実施したであろう対応を想定して準備された模擬データを使用した。

KENGINE で管理された脆弱性情報には、FACT と呼ばれる脆弱性の特徴など脅威分析情報と、情報の識別番号（CVE、VU#、または JVNVU#）や、脆弱性の概要などの基本的な情報が含まれる。図 1 と図 2 に、脆弱性の基本情報と脆弱性の特徴を表現した FACT を示す。

**** General Information **** Edit

Report ID	: CVE-2008-4822		
Title	: Adobe Flash Player 9.0.124.0 and earlier does not properly interpret policy files		
Memo	: Adobe Flash Player 9.0.124.0 and earlier does not properly interpret policy files, which allows remote attackers to bypass a non-root domain policy.		
URL	: http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2008-4822		
Status	: Closed	**VRDA Feed**	: Published
Created	: 2009/03/17 16:21	**Last Updated**	: 2009/03/18 08:51
Created By	: rweiland		
Tri. Handler	: rweiland	**Vul. Handler**	: rweiland
Surface Completed	: 2009/03/17 16:22		
Detailed Completed	: 2009/03/17 16:25		
Decision Finalized	: 2009/03/18 08:51		
Report Closed	:		

図 1：KENGINE 上の脆弱性基本情報

[9] Weka3：Java で書かれたデータマイニングソフトウェア（http://cs.waikato.ac.nz/ml/weka/）
[10] http://www.jpcert.or.jp/vrdafeed/

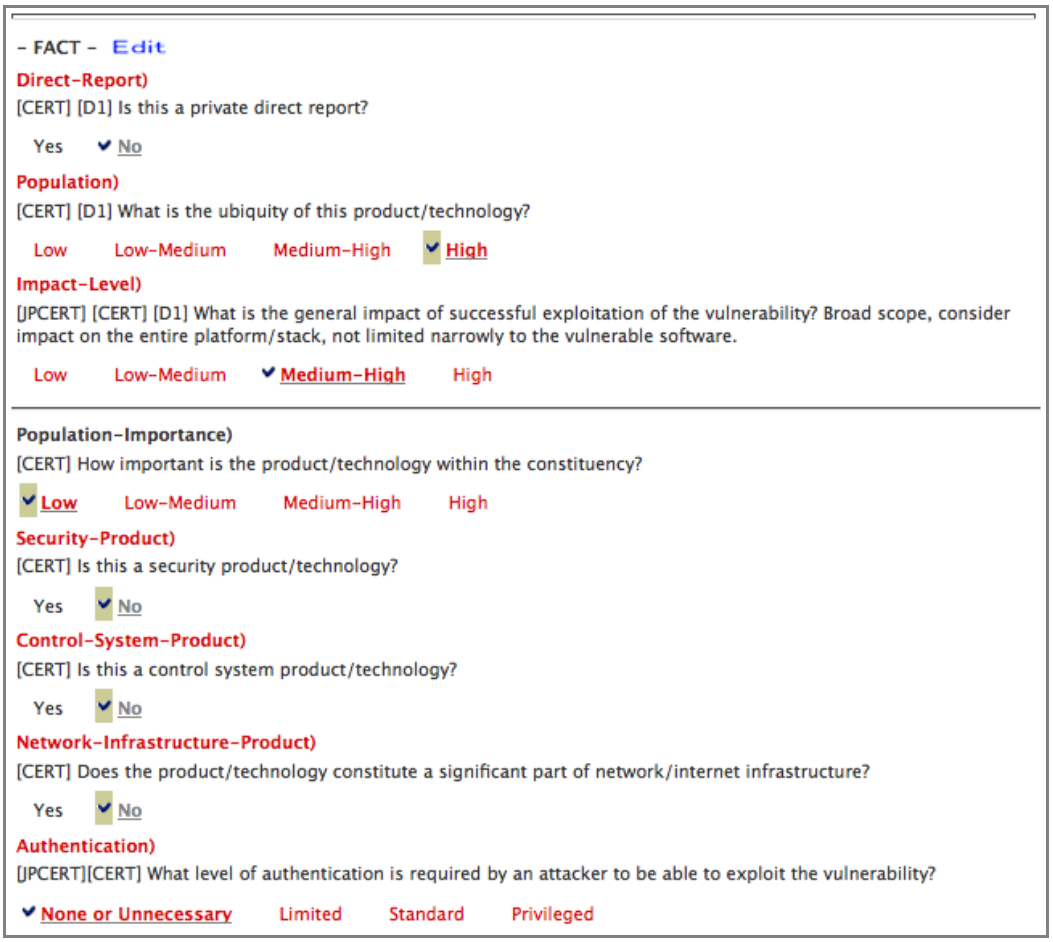


図 2：KENGINE 上で脆弱性の特徴を表現する FACT（抜粋）

参加組織は、トレーニングデータを構成する対応判断実績データを収集するために、各自の脆弱性対応業務において定常的に実施している幾つかの対応タスクを定義し、それぞれのタスクに対しての各脆弱性への対応判断実績（CERT/CC の場合はシミュレートされた対応判断）を記録した（タスクの例としては、「脆弱性の詳細分析」や「注意喚起の発行」など）。これらの組織別の対応タスクについての情報は、付録 A、B、C に記載されている。CERT/CC では、分析担当者はトレーニングデータを入力する作業中、調査に先立って KENGINE に定義済みのディシジョンツリーが存在したため、KENGINE により提示された対応予想が確認可能なケースもあった。これら過去に定義されたディシジョンツリーは調査上意味を持たないため、CERT/CC の分析担当者はトレーニングデータ入力中に KENGINE により提示された対応予想を無視するように指示されていた。しかし、実際の対応判断を記録する際に、分析担当者の意思決定に影響を与えうる情報を完全に遮断する具体的な方法をとらなかったため、担当者のバイアスが意思決定に影響した可能性は否定できない。参加組織 A と IIJ Technology の分析担当者は、対応判断の意思決定を実施する際に KENGINE により提示された対応予想を目にすることはなかった。

意思決定ルールのモデル化

意思決定モデルは、分析担当者が各自の対応判断を記録した後、KENGINE を用いてこれらトレーニングデータに基づいて作成される。これらの意思決定モデルは、意思決定の対象となる対応タスク別にディシジョンツリーで表現される。また、一つの対応タスクは、ほかの異なる対応タスクに依存する場合がある。例えば、CERT/CC の対応タスク Publish Technical Alert（注意喚起発行タスク）は、対応タスク Perform Surface Analysis（一次分析タスク）に依存している。つまり、CERT/CC は報告された脆弱性に対して対応タスク Perform Surface Analysis を実施してからでないと、対応タスク Publish Technical Alert を実施しない。このディシジョンツリーは、脆弱性の特徴などあらゆる対応意思決定に影響する判断要素（FACT）、または依存するタスクを各ノードとして構成される。ディシジョンツリーの生成に使用されたアルゴリズムについては、『VRDA（Vulnerability Response Decision Assistance）』の第 2.4 項で説明されている。要約すると、このアルゴリズムは属性（FACT と依存タスク）をツリー内の潜在的ノードとして再帰的に処理する。トレーニングデータに含まれたタスクへの実際の対応判断に関して最も多くの情報を提供し、且つ、カイ二乗検定を経て統計的有意性を示す属性がルートノードとして選択される。その後の属性選択において、ほかの有意な属性が見つかった場合は、ブランチノード（判断の分岐点）が作成される。有意な属性がそれ以上見つからなかった場合は、トレーニングデータ中の実際の対応判断の重み付け値の平均に基づいてリーフノード（対応予想）が作成される。FACT や依存タスクによっては、ディシジョンツリーに含めるのに十分有意な情報を提供しないものもある。短いディシジョンツリーの例を図 4 に示す。

本調査において、いくつかのケースでは、ディシジョンツリーが手作業で作成されたり、KENGINE が生成したディシジョンツリーに対して修正が加えられたりした。手作業での作成や修正は、修正者が、意思決定プロセスを熟知しており、且つ、少ない数の FACT を対応判断の要素としている対応タスクや、KENGINE が生成したディシジョンツリーが十分に対応予想を提示できなかった対応タスクに対して行われた。このため、KENGINE が機械的に生成したディシジョンツリーによって提示される対応予想の正確性と、手作業で作成または修正したディシジョンツリーにより提示される対応予想の正確性の比較も行った。

VRDA において、あるタスクへの対応判断（意思決定）は、基本となる 4 つの優先度レベルで表現することを推奨している。これらは、優先度が高い順に「*Must（必須）*」、「*Should（推奨）*」、「*Might（検討）*」、および「*Won't（不要）*」となる。例えば、CERT/CC の CVE-2008-4822 に関する Publish Vulnerability Note タスクは次のように記述できる。

CVE-2008-4822 について、CERT/CC による Publish Vulnerability Note は{必須|推奨|検討|不要}となる。

本調査において、あるタスクへの対応優先度がどのようにエンコードされたかを表 1 に示す。

表 1：対応優先度のエンコード

対応優先度	値
Must（必須）	3
Should（推奨）	2
Might（検討）	1
Won't（不要）	0

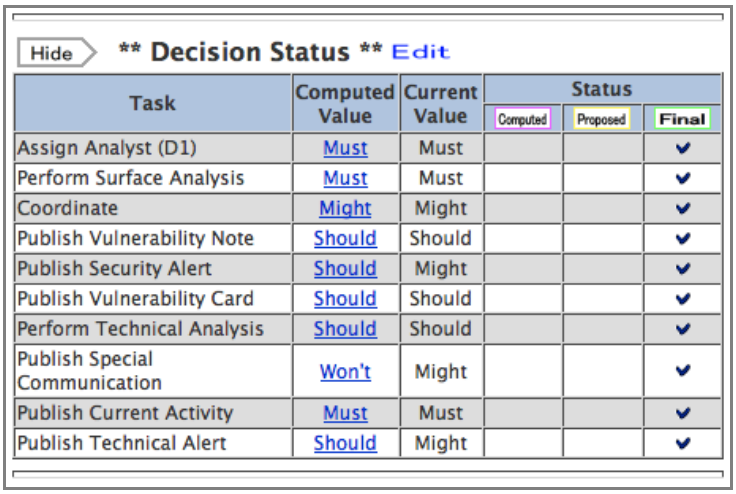
VRDA には、D1 というラベルが付いた一次対応タスク（D1 タスク）というコンセプトが含まれている。組織は D1 タスクにより、特定の処理を必要とする脆弱性を容易に識別できる。D1 タスクは、他のタスクより実施判断のために必要な情報（FACT）が少なく、通常はその後一連の作業として発生しうる追加の分析作業および他のタスクの実施検討を制御するものである（つまり後続して発生しうる作業のフィルタとして機能する）。CERT/CC を例にとると、脆弱性情報に分析担当者を割り当てるという対応タスク（Assign Analyst）は D1 タスクである。このタスクの対応判断を行うために必要な FACT はわずかであり、また、対応優先度は「*Must（必須）*」か「*Won't（不要）*」だけに限られている。対応優先度が「*Won't（不要）*」の場合、それ以上の対応は行われない。

タスクへの対応が 2 つしかない場合のエンコードを表 2 に示す。

表 2：ブール値による対応優先度のエンコード

対応優先度	値
Must（必須）	1
Won't（不要）	0

KENGINE により提示された対応予想と、分析担当者が行う実際の対応判断の例が図 3 に示されている。表中の列”Computed Value” には KENGINE により提示された対応予想が表示され、列”Current Value” には、担当者によって与えられた対応判断が表示されている。



Task	Computed Value	Current Value	Status		
			Computed	Proposed	Final
Assign Analyst (D1)	Must	Must			✔
Perform Surface Analysis	Must	Must			✔
Coordinate	Might	Might			✔
Publish Vulnerability Note	Should	Should			✔
Publish Security Alert	Should	Might			✔
Publish Vulnerability Card	Should	Should			✔
Perform Technical Analysis	Should	Should			✔
Publish Special Communication	Won't	Might			✔
Publish Current Activity	Must	Must			✔
Publish Technical Alert	Should	Might			✔

図 3：KENGINE の決定ステータス

以下、図 4 にタスク Perform Technical Analysis （詳細分析タスク）のディシジョンツリーの例を示す。

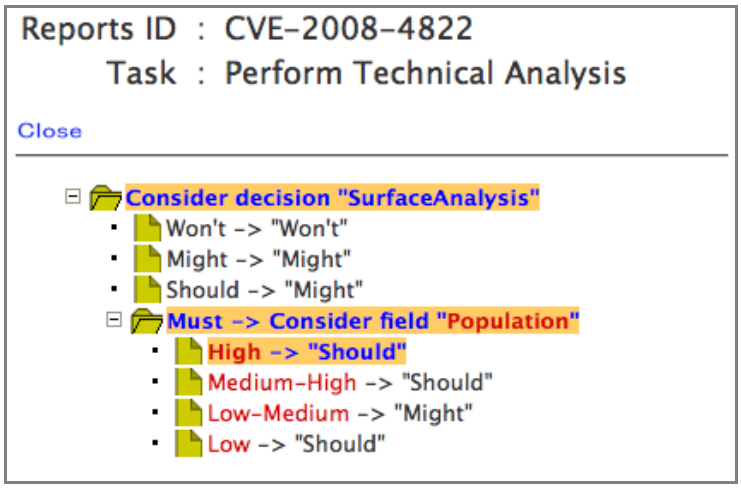


図 4：KENGINE ディシジョンツリー

このディシジョンツリーには明らかな矛盾がある。ディシジョンツリー上に分岐点として含まれている FACT “Population”の取る値が ”Low-Medium（低ー中）” の場合、結果としての対応予想は「*Might（検討）*」になる。しかし、”Population” の取る値が、”Low（低）” の場合、提示される対応予想は、より高い優先度を示す「*Should（推奨）*」になる。この種の矛盾は CERT/CC のディシジョンツリー内に幾つか見られたが、分析担当者による矛盾した対応判断の結果とも考えられる。このような矛盾点を修正し提示される対応予想の一貫性を向上させるために、KENGINE が生成したディシジョンツリーに対して必要最小限の変更を加えた（例えば、Population が取る値が Low（低）の場合の決定を「*Might（検討）*」に変更するなど）。VRDA のディシジョンツリー生成アルゴリズムはこの種の「優先度の一貫性」を強制しておらず、調査において実際にこの矛

盾を修正した場合でも、提示される対応判断の正確性が大きく向上することはなかったため、優先度を一貫させる必要はないとも考えられる。

分析技法

VRDA の実装である KENGINE により提示された対応予想と、分析担当者によって行われた実際の対応判断の誤差は、次のような技法を使って分析された。

- 単純なブール値（値は 0 または 1 のどちらか）で表現される「ヒット率」（誤差がある場合は 0、誤差が無い場合は 1）。
- 1 つ違い、つまり「ニアミス」のヒット率（誤差を示す値が 1 より大きい場合は 0、誤差を示す値が 1 以内の場合は 1）は、NMR（Near Miss Rate）と略される。このレベルの正確性を達成することが、VRDA の設計上の目標である。これについては、『VRDA（Vulnerability Response Decision Assistance）』で次のように述べられている。

  筆者は、結果として得られる対応判断は不完全であると想定している。しかし、それを補うために、ブール値の代わりに、対応予想を優先度を示す表現で提示する。具体的には、「must（必須）」、「should（推奨）」、「might（検討）」、「won't(不要)」の 4 レベルを使用する。目標は、結果として得られる対応予想が、実際に「正しい」値と、1 レベルを超えて異ならないようにすることである[11]。

考えられる対応優先度のレベルが 2 つしかないタスクについての NMR は有用ではない。なぜなら、最大誤差である、1 でさえ NMR の範囲内なので、常に NMR は 100%となるためである。対応優先度が「***Must（必須）***」と「***Won't（不要）***」である典型的な二択のタスク（ブール値タスク）の場合、1 つ違いということは完全に誤っていることと同じである。

- 平均 2 乗誤差（MSE：Mean of Squared Error）。この技法は、提示された対応予想と実際の対応判断の誤差を強調する。この測定値が取り得る範囲は、タスクに対して選択可能な対応優先度レベルの数によって異なる。例えば、前述した 4 つのレベルの対応優先度を持つタスクであれば、MSE の範囲は 0（誤差無し）から 9（誤差が 3 つ）までである。
- 平均誤差（ME：Mean of the Error）。「*実際の対応判断のエンコード値 – 提示された対応予想のエンコード値*」として、代数的に計算される（符号付き数値として）。この技法は、一定のずれの「方向」を明らかにできることがある。負の数は、分析担当者が実際に望ましいと考えた対応判断よりも優先度の高い対応を VRDA が予想していることを示し、正の数はその逆を示す。MSE と同様に、ME の範囲はタスクに対して選択可能な対応優先度レベルの数によって異なる。
- 増分分析のために、異なるサンプルサイズに対するヒット率および MSE の直線近似を測定するには、相関係数（*r*）が使用される。

NMR、MSE、および ME は、対応優先度レベルの数による影響を受けるため、選択可能な対応優先度レベルの数が少ないタスクは、その他の条件がすべて同じであれば、対応優先度レベルの数が多いタスクよりも平均誤差の値が小さくなる傾向がある。

[11] 出典：“Vulnerability Response Decision Assistance” (http://www.jpcert.or.jp/research/2008/2007VRDA_paper_JP.pdf)

上記のすべての測定値において、誤差は次のように計算される。

誤差 = 実際の対応判断のエンコード値 - 提示された対応予想のエンコード値

このほかに、正確性の評価には直接繋がらないが、次に示す 3 つの技法も使用された。

- 提示された対応予想と実際の対応判断の平均偏差と標準偏差。標準偏差が小さい場合、それはそのタスクに対する回答がより一定していることを示しており、そのタスクが意思決定の対象というよりは、どちらかというと固定的に定義可能な手続きにすぎないことを意味している可能性がある。この分析は、付録 D に記載されている。
- 脆弱性対応プロセスの複雑度評価。ディシジョンツリー内の平均ノード数、および、ディシジョンツリー内に含まれる選択肢の合計で測定される。
- コスト見積り（CERT/CC の場合のみ）。タスクごとの想定コストおよび提示された対応判断に基づいて計算される（付録 E を参照）。

分析

VRDA は、非常に多様となりうる組織の脆弱性対応ルールを体系的に整理した情報を入力として機能するため、ある特定の組織での利用において VRDA がどの程度の正確性を示すことができるのかを事前に算出することは困難である。NMR の説明で述べたように、VRDA は、実際の優先度レベルに対して 1 つの優先度レベルの誤差は許容の範囲として、タスクの対応予想を提示することを目標に設計された。これは、VRDA が脆弱性対応業務の支援に役立つためには、この程度の正確性で十分だと考えられているからである。ただし、この目標は実用的見地から選択されたものであり、厳密な数学的根拠があるわけではない。各参加組織における VRDA の正確性は、NMR、ヒット率、および MSE によって測定された。高い NMR とヒット率は、提示される対応予想の正確性が高いことを示している。高い MSE は、誤差の程度が大きい（予想と実際の判断のずれが大きい）ことを示している。

VRDA が実用可能なディシジョンツリーを生成するためには、十分な量のトレーニングデータが必要である。ディシジョンツリー生成アルゴリズムが、ツリー上に最終的な対応判断（リーフノード）や、重要な判断要素（ブランチノード）を含んだツリーを作成できないようなケースは、トレーニングデータの不足によって起きる。これに関連した 2 つの疑問は、第 1 として実用に耐える正確性を持つディシジョンツリーを生成するためにどれだけの量のトレーニングデータが必要かということであり、第 2 は、トレーニングデータがさまざまな種類の脆弱性情報をどの程度適切にカバーする必要があるかということである。第 1 の疑問を調査するために、CERT/CC のデータを使ってトレーニングデータの増分分析を実行した。しかし、第 2 の疑問については、CERT/CC のトレーニングデータとして十分に多様な種類の脆弱性情報が含まれていたと表明できることに留まる。

次の章では各組織の脆弱性対応プロセスの複雑度と、提示される対応予想の正確性の関係について考察する。

脆弱性対応プロセスの複雑度

参加組織の脆弱性対応プロセスの複雑度は、VRDA が提示できる対応予想の正確性と関連している可能性がある。例えば、複雑な脆弱性対応プロセスを正確にモデル化することは、シンプルな脆弱性対応プロセスのモデル化よりも困難であると考えられる。VRDA では、脆弱性対応プロセスの複雑度は、各組織の FACT（脅威分析項目）、FACT 値（各脅威分析項目が取る指標）、タスク、および、タスクへの対応の定義された数によって表現される。各参加組織の脆弱性対応プロセスの相対的な複雑度は、いくつかの手段によって測定された。その手段とは、組織毎に定義された FACT、FACT 値、タスク、および、タスクへの対応優先度レベルの数、タスクのディシジョンツリーに含まれたノードの平均数、タスクのディシジョンツリー上に含めることができる選択肢の数の合計である。これらの複雑度の測定値を表 3 と表 4 に示す。表 4 では、複雑度と正確性を比較している。

表 3：脆弱性対応プロセスの複雑度

参加組織	プロセスの複雑度			
	FACT	FACT 値	タスク	対応優先度レベル数
参加組織 A	21	79	5	20
IIJ Technology	8	22	5	10
CERT/CC	16	53	10	38

表 4：脆弱性対応プロセスの複雑度と正確性

参加組織	プロセスの複雑度		正確性	
	平均ノード数	ツリー分岐点数	ヒット率	MSE
参加組織 A	15.0	403	67%	0.81
IIJ Technology	2.2	118	100%	0.00
CERT/CC	25.1	580	70%	0.56

なお、タスクのディシジョンツリー上に含めることができる選択肢の数の合計である「ツリー分岐点数」の値は、次のように計算される。

1. 各タスクについて、定義されたすべての FACT 値の数と、すべての依存するタスクに定義された対応優先度レベルの数の合計を足し合わせる。これにより、各タスクで取りうることができる選択肢の数が算出される。

2. その後、参加組織毎に各タスクの選択肢の数を合計して、参加組織の選択肢の総数を算出する。

本文書では、参加組織の FACT、FACT 値、タスク間の依存関係、および、依存するタスクについて、すべての情報は提供していないが、一部のタスクおよび FACT の情報は、付録 A、B、および C に記載されている。

最も明確な関係を示しているのは、IIJ Technology から得られた調査結果であり、脆弱性対応プロセスの複雑度が最も低く、VRDA が提示する対応予想の正確性が最も高くなっていることがわかる。

参加組織 A

参加組織 A は、日本の大手多国籍企業内の事業部門に対して、脆弱性情報の分析と配信を行う CSIRT（Computer Security Incident Response Team：コンピュータセキュリティインシデント対応チーム）である。参加組織 A は、71 件の脆弱性に関する対応実績データを準備したが、その中で KENGINE により機械的に対応予想を提示できなかったデータは調査対象外とした。参加組織 A は、トレーニングデータとして実際の脆弱性対応業務のデータを使用した。参加組織 A の脆弱性対応プロセスの概要とタスクのリストは、付録 A に記載されている。

参加組織 A における VRDA の正確性を調査した結果を表 5 に示す。

表 5：参加組織 A における VRDA の正確性

タスク	サンプルサイズ	ヒット率	NMR	MSE	ME
全体的パフォーマンス	341	67%	88%	0.81	-0.11
Detail Analysis (D1)（詳細分析）	66	44%	74%	1.56	-0.32
Warning Level Alert（警戒レベル注意喚起）	67	63%	93%	0.60	-0.12
Inform All Contact Points（すべての連絡先へ通知）	69	74%	91%	0.59	-0.16
Inform Specific Contact Points（特定の連絡先へ通知）	69	75%	94%	0.49	-0.14
Critical Level Alert（重大レベル注意喚起）	70	80%	89%	0.83	0.20

タスク Detail Analysis の統計データが示している若干高い MSE と低いヒット率は、提示された対応予想と実際の対応判断が多くのケースで不一致であったことを意味する。また、提示された対応予想が示す優先度は、一般に実際に分析担当者により与えられた優先度より高くなっている（ME = -0.32）。これら不一致の原因は、トレーニングデータが不十分であったことが考えられる（入力された脆弱性の数が少なすぎたか、タスク Detail Analysis について適切な対応予想を行うために必要な情報（FACT）が欠けていたなど）。参加組織 A の脆弱性対応プロセスを調べた結果、タスク Detail Analysis が D1 タスクであることがわかった。これは、タスク Detail Analysis に対する対応判断が、限られた数の FACT の分析のみが行われており、その他の FACT の分析結果が与えられる前に行われていたことを意味する。このことが、より正確な対応予想を行うために必要な情報（FACT）の欠落の原因であった可能性がある。

各タスクの誤差カウントの分布を表 6 に示す。ヒストグラムのような視覚効果を与えるために、値がゼロのセルは空白のままにしてある。誤差は「*実際の対応判断のエンコード値 - 提示された対応予想のエンコード値*」として計算されており、負の数値が大きいほど実際に望ましい対応優先度より高い対応優先度が予想されたことを意味し、正の数値が大きい場合はその逆を意味する。表内の影付きの領域は、1 未満の誤差（NMR）を示しており、これが正確性に関する VRDA の設計上の目標である。

表 6：参加組織 A における VRDA の予想と実際の誤差分布

タスク	誤差						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
全体的パフォーマンス	4	23	40	230	31	8	5
Detail Analysis (D1) （詳細分析）	3	13	4	29	16	1	
Warning Level Alert （警戒レベル注意喚起）		4	11	42	9	1	
Inform All Contact Points （すべての連絡先へ通知）		4	10	51	2	1	1
Inform Specific Contact Points （特定の連絡先へ通知）	1	2	9	52	4	1	
Critical Level Alert （重大レベル注意喚起）			6	56		4	4
			設計目標（NMR）				

NMR の測定結果によれば、参加組織 A について VRDA の対応予想の正確性は妥当であったと言えるが、タスク Detail Analysis は例外である。タスク Detail Analysis の対応判断に使用された FACT（同タスクのディシジョンツリー上に現れるもの）を見直し、71 件のうち 66 件しか予想されなかった原因を調べることにより、タスク Detail Analysis についての対応予想の正確性が向上する可能性がある。

なお、参考情報として、参加組織 A の対応予想と実際の判断の誤差を強調する MSE 値と脆弱性情報の件数との対応を示す MSE 分布を付録 F の図 8 に示す。

参加組織 B：IIJ Technology

本調査の 2 番目の参加組織は、幅広い規模のネットワークやシステムの構築・運用サービスを提供している IIJ Technology である。同社は、運用サービスの一環として、顧客のシステム構成に即した脆弱性情報を収集、影響度を判定した上で情報提供を行い、顧客のセキュリティ対応活動を支援するサービス（*脆弱性情報判定サービス*）を提供している。このため、IIJ Technology は同社の顧客が使用している IT 資産に関する正確なインベントリおよび配備情報を維持管理しており、この点が他の参加組織とは異なる。IIJ Technology のタスクに定義された対応優先度は、「*Must（必須）*」と「*Won't（不要）*」の 2 つだけである。トレーニングデータは、63 件の脆弱性について提供されており、5 つのタスクが定義されている。IIJ Technology は本調査のために、特定の 1 顧客に対する実際のサービス運用における実績データをトレーニングデータとして使用した。IIJ Technology の脆弱性対応プロセスの概要とタスクに関する情報は、付録 B に記載されている。

表 7 および表 8 に示すように、VRDA は IIJ Technology のタスクに対し、すべて実際の対応判断と完全に一致した対応予想を提示した。これは、脆弱性対応プロセスが比較的単純であることに起因する可能性が高い。すなわち、すべてのタスクの対応優先度が「*Must （必須）*」と「*Won't（不要）*」の 2 つに限られ、定義された FACT と FACT 値の数が少なく、また、少ない数の分析担当者が対応判断に関与したためと考えられる（より数の少ない分析担当者の関与は、担当者別の対応判断のバラつきが抑止され一貫性向上に繋がる）。また、タスク Ignore、Emergency Level Alert、Critical Level Alert においては、それぞれのディシジョンツリーにノードが 1 つしかない（考えられる対応が 1 つしかない）。正確性が 100%で選択可能な対応がディシジョンツリー上に 1 つしかないこれらのタスクは、対応判断の是非を問う必要がない静的なプロセスであると考えることもできる。しかし、IIJ Technology のトレーニングデータを見ると、タスク Emergency Level Alert または Critical Level Alert の基準を満たす脆弱性が実際 1 つもないことがわかる。したがって、VRDA はトレーニングデータに基づいて正しい対応（「*Won't（不要）*」）を予想している。IIJ Technology は、一部の脆弱性が実際にタスク Emergency Level Alert または Critical Level Alert を必要とする場合があることを前提としているので、これらのタスクは静的なプロセスに変更するべきではない。トレーニングデータのセットがより大きいか包括的であれば、より現実的なモデルが生成されるはずである。また、IIJ Technology のデータセットには、タスク Ignore が実行されなかった（無視されずに対応の検討が必要と判断された）脆弱性だけが含まれていたためタスク Ignore に対する対応予想の提示結果は 100%の正確性を示すことが予想できた。実際に VRDA により提示されたタスク Ignore の対応予想はすべて、”Ignore = 不要” であり、脆弱性情報を無視しない（つまり、何らかの対応が必要）という予想を提示した。

表 7：IIJ Technology における VRDA の正確性

タスク	サンプル サイズ	ヒット率	NMR	MSE	ME
全体的パフォーマンス	315	100%	100%	0.00	0.00
Ignore（無視）	63	100%	100%	0.00	0.00
Emergency Level Alert （緊急レベル注意喚起）	63	100%	100%	0.00	0.00
Critical Level Alert （重大レベル注意喚起）	63	100%	100%	0.00	0.00
Warning Level Alert （警戒レベル注意喚起）	63	100%	100%	0.00	0.00
FYI（参考）	63	100%	100%	0.00	0.00

表 8：IIJ Technology における VRDA の予想と実際の誤差分布

タスク	**誤差**						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
全体的パフォーマンス				315			
Ignore（無視）				63			
Emergency Level Alert （緊急レベル注意喚起）				63			
Critical Level Alert （重大レベル注意喚起）				63			
Warning Level Alert （警戒レベル注意喚起）				63			
FYI（参考）				63			
			設計目標（NMR）				

他に 2 つの要因が、IIJ Technology における VRDA の予想の正確性の高さに影響したと考えられる。本調査を開始する以前から、IIJ Technology では、同社の顧客向けに脆弱性情報を提供するサービスを運用していることもあり、明確に定義された既存の脆弱性対応プロセスが存在しており、小人数で構成された分析チームが一貫してそのプロセスに従っていた。また、IIJ Technology は顧客の IT 資産に関するインベントリと配置情報を維持管理しており、LAPT（Lightweight Affected Product Tag）[12]FACT に基づいて正確なディシジョンツリーを生成するために、これら IT 資産情報が役立った可能性がある。Ubiquity（普及率）や Population Importance（影響があるシステムの重要性）などの LAPT FACT は、脆弱性そのものではなく脆弱性の影響を受ける可能性がある製品や技術についての脅威分析上意味のある情報を記述したものである。IIJ Technology のディシジョンツリーを調べた結果、多くのノードが製品の利用状況や配置に関連する FACT であることがわかった。つまり、IIJ Technology が維持管理する顧客の IT 資産情報を利用すれば、ツリー上の各ノードを LAPT FACT として定義することも可能であり、これら IT 資産情報に基づいた対応予想が可能であることを意味している。これは、正確なインベントリおよび配備情報が、正確な対応予想に貢献したという考えを裏付ける事実である。

参加組織 C：CERT/CC

3 番目の参加組織は CERT/CC の脆弱性分析チームである。CERT/CC は、アメリカ国内および全世界を対象として脆弱性と対応タスクに優先順位を付ける。CERT/CC が使用する FACT とタスクの情報は、付録 C を参照のこと。

CERT/CC は、トレーニングデータとして大きく 2 種類の脆弱性情報データセットを使用した。中核となるセットには、各種の脆弱性（例えば、コードの実行を許すスタックおよびヒープオーバーフロー、リモートおよびローカルの脆弱性、ディレクトリトラバーサル、SQL インジェクション、クロスサイトスクリプティング、競合状態など）が適度に網羅された脆弱性情報のサンプルがおよそ 50 件含まれた。もう一方のデータセットは、2007 年から 2009 年までに開示されている NVD（National Vulnerability Database）[13] と Vulnerability Notes Database[14]に記載された脆弱性の中から、おおまかな年代順に選択された。

D1 分析（1 番目の意思決定ポイントについての分析）

前述したように、VRDA には複数の意思決定ポイントを含むというコンセプトが盛り込まれている。CERT/CC は、2 つの異なる意思決定ポイント、D1 および D2 を使用した（D1,D2 はそれぞれ 1 番目、2 番目の意思決定ポイントであることを意味する）。D1 タスクである Assign Analyst は、その後に行われる D2 タスクの対応判断のために必要な情報の収集や分析などの労力を費やす前に、D2 タスクの実施検討の対象となる脆弱性情報を絞り込むために使用される。D1 タスクである Assign Analyst は、D2 タスクに先行して、より少ない情報量で対応の意思決定が行われる。つまり、最小の対応コストを掛けて、その後に発生するより対応コストが高いタスクの実施対象を絞り込む仕組みである。

[12] 出典：“Vulnerability Response Decision Assistance” (http://www.jpcert.or.jp/research/2008/2007VRDA_paper_JP.pdf)
[13] http://web.nvd.nist.gov/view/vuln/search
[14] http://www.kb.cert.org/vuls

CERT/CC は、タスク Assign Analyst（分析担当者割り当て）を実施するかを判断するために、次の FACT のサブセットを使用している。

1. 報告が非公開であるかどうか（Direct Report）
2. 脆弱性が与える影響（Impact）
3. 影響を受けるシステムの普及率（Ubiquity）

このようにタスク Assign Analyst（分析担当者割り当て）は、単純な意思決定プロセスで対応判断できるため、手作業で KENGINE 上にディシジョンツリーを作成することとした。CERT/CC は、FACT Direct Report（直接報告）と、Control System Product（制御システム）または Network Infrastructure Product（ネットワークインフラストラクチャ製品）に関する分析結果が真（true）である場合には、常にタスク Assign Analyst を行う（「*Must*（*必須*）」）。それ以外の場合は、FACT Impact（影響）と Ubiquity（普及率）の分析結果によってタスク Assign Analyst 対応を決めている。図 5 は、16,025 件の脆弱性に対しての FACT Impact と Ubiquity の分析結果の分布を示している。

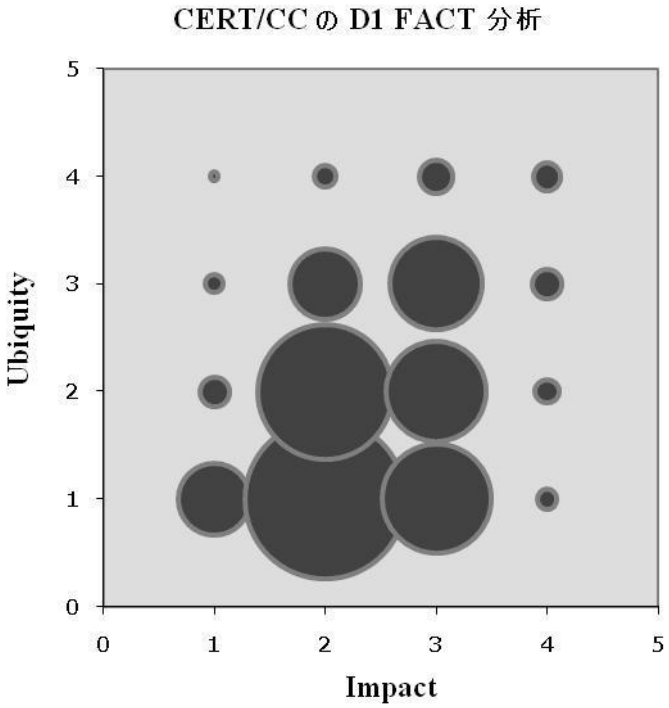


図 5：CERT/CC の D1 FACT 分布

CERT/CC の分析担当者は、Impact と Ubiquity の分析結果の組み合わせ分布を見て、どの組み合わせ条件が成り立った場合に、Assign Analyst タスクを実施することが最も適切かを検討した。例えば、CERT/CC は、Ubiquity の分析結果が 2 より小さく（値 2 は *Low-Medium"低－中"を示す*）Impact の分析結果が 2 より小さい（値 2 は *Low-Medium "低－中" を示す*）脆弱性情報に対しては、タスク Assign Analyst を行わないこととした。VRDA を使用することで、このような脆弱性の分析結果の分布情報を得ることができ、このケースにおいては、分布情報を基に手作業でタスク Assign Analyst のディシジョン

ツリーを作成した。なお、この分布情報そのものは、脆弱性の分析結果を別の形で表現したものであり、VRDA による対応予想の正確性を測るものではなく、あくまで VRDA の他の活用方法の一例を紹介したものである。

このように手作業で作成したタスク Assign Analyst ディシジョンツリーを、トレーニングデータを基に KENGINE によって生成された同タスクのディシジョンツリーと、それぞれが提示する対応予想の正確性について比較した。2 つのディシジョンツリーは似ており、両方とも FACT Ubiquity と Impact を判断要素として考慮している。異なる点として、KENGINE により生成されたディシジョンツリーでは、FACT Authentication（認証）も判断要素として考慮されていたが、Direct Report（直接報告）は考慮されなかった。双方のディシジョンツリーの正確性の比較結果を示す表 9 では、手作業で作成されたディシジョンツリーと機械的に生成されたディシジョンツリーとの間で正確性に大きな違いは見られない。これはおそらく、CERT/CC がタスク Assign Analyst の対応判断について長い経験を持っており熟知していることと、同タスクの対応判断自体が比較的単純である（限られた数の FACT を考慮して、以降の作業が必要かどうかを判断する）ことが原因として考えられる。このことから、十分に理解された意思決定プロセスに基づく単純な対応判断の場合、専門の分析担当者は、KENGINE によって生成されたディシジョンツリーと同程度の正確性を示すディシジョンツリーを作成できると考えられる。

表 9：CERT/CC のタスク Assign Analyst のツリー作成方法別の誤差比較

ツリー作成方法	サンプルサイズ	ヒット率	NMR	MSE	ME
手作業 Assign Analyst (D1)	215	90%	100%	0.10	0.05
KENGINE 生成 Assign Analyst (D1)	215	90%	100%	0.10	0.05

なお、タスク Assign Analyst が取り得る対応は 2 つだけなので、NMR は常に 100%となる。

D2 分析（2 番目の意思決定ポイントについての分析）

D2 タスクの調査データは、215 件の脆弱性について提供された。これらの脆弱性それぞれについて、10 個のタスクへの対応判断があり、サンプルサイズは最大 2150 となる。このデータ中、KENGINE が対応予想を提示することができなかったタスク Publish Current Activity の対応予想における 3 つのケースは、正確性の評価対象から除かれた。CERT/CC における VRDA の正確性を示す統計データを表 10 に示す。前述した D1 タスク Assign Analyst 以外のすべてのディシジョンツリーがトレーニングデータを基に KENGINE によって生成された。

表 10：CERT/CC における VRDA の正確性

タスク	サンプル サイズ	ヒット率	NMR	MSE	ME
全体的パフォーマンス	2,147	70%	94%	0.56	0.05
Assign Analyst (D1)（分析担当者割り当て）	215	90%	100%	0.10	0.05
Perform Surface Analysis（一次分析）	215	88%	90%	0.83	0.20
Perform Technical Analysis（詳細分析）	215	68%	93%	0.56	-0.02
Coordinate（連絡調整）	215	58%	93%	0.68	0.01
Publish Vulnerability Card（脆弱性カード）	215	68%	95%	0.50	0.08
Publish Vulnerability Note（脆弱性ノート）	215	66%	94%	0.57	0.12
Publish Technical Alert（テクニカルアラート）	215	60%	89%	0.81	0.08
Publish Security Alert（セキュリティアラート）	215	63%	92%	0.67	0.01
Publish Special Communication（会員への情報提供）	215	80%	99%	0.27	0.21
Publish Current Activity（関連活動状況の共有）	212	57%	94%	0.60	-0.21

NMR、MSE、および ME は全般的に期待した結果であったが、ヒット率はやや期待はずれであった。提示された対応予想の正確性は設計上の目標に近いが、より高い正確性を示すことができると予想していた。予想と実際との誤差の分布を表 11 に示す。

表 11：CERT/CC における VRDA の予想と実際の誤差分布

タスク	誤差						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
全体的パフォーマンス	4	40	256	1,500	264	52	31
Assign Analyst (D1) （分析担当者割り当て）			6	193	16		
Perform Surface Analysis （一次分析）	2	1	3	190	1	2	16
Perform Technical Analysis （詳細分析）		3	39	146	16	9	2
Coordinate （連絡調整）		4	44	125	32	7	3
Publish Vulnerability Card （脆弱性カード）		2	27	147	31	6	2
Publish Vulnerability Note （脆弱性ノート）		2	28	141	33	9	2
Publish Technical Alert （テクニカルアラート）	1	9	27	129	36	10	3
Publish Security Alert （セキュリティアラート）	1	10	26	135	36	6	1
Publish Special Communication （会員への情報提供）				173	40		2
Publish Current Activity （関連活動状況の共有）		9	56	121	23	3	
			設計目標（NMR）				

注目すべき異常値の 1 つは、タスク Perform Surface Analysis における 16 件の完全なミス（誤差 = 3）である。これら 16 件の脆弱性に対しては、他のタスクについても対応予想の誤差が大きいものが多かった。これら 16 件の脆弱性に対しての対応予想の平均 MSE

は 2.45 で、標準偏差は 1.09 である。これに比べて、全体の平均 MSE は 0.56 で、標準偏差は 0.81 である。16 件の脆弱性のハミング距離の計算と目視による確認では、これら脆弱性の明白な類似性や、ほかの脆弱性情報と著しく異なる点は見つからなかった。CERT/CC が調査に用いた脆弱性の MSE 分布は、付録 F の図 9 に示されている。

さらに分析担当者が手を加えた（手作業で修正あるいは作成した）ディシジョンツリーと KENGINE により機械的に生成したツリーとの対応予想の正確性を比較する目的で、CERT/CC の分析担当者が各タスクのディシジョンツリーの誤差率を再び計算した。分析担当者はいくつかのタスクについて、KENGINE により生成されたディシジョンツリーに小さな修正を加えた。これら修正は、KENGINE が対応予想を提示できなかったタスクに対して対応予想に到達するためにツリー上にパスを追加したり、前述した矛盾点を含むディシジョンツリーの例で説明したような小さな矛盾を修正したりするといった目的で行われた。また、タスク Publish Technical Alert、Publish Security Alert、Publish Special Communication、および Publish Current Activity については、KENGINE が生成したディシジョンツリーを基に修正したのではなく、分析担当者は手作業でディシジョンツリー全体を作成した。表 12 と表 13 は、CERT/CC の分析担当者が修正あるいは作成したディシジョンツリーの正確性と誤差分布を示している。手作業で作成した 4 つのタスクのディシジョンツリーには、アスタリスク（*）が付けられている。

表 12：手作業で修正あるいは作成されたツリーの正確性

タスク	サンプルサイズ	ヒット率	NMR	MSE	ME
全体的パフォーマンス	2,141	66%	91%	0.72	0.04
Assign Analyst (D1)（分析担当者割り当て）	215	90%	100%	0.10	0.05
Perform Surface Analysis（一次分析）	215	88%	91%	0.79	0.20
Perform Technical Analysis（詳細分析）	215	66%	93%	0.64	-0.04
Coordinate（連絡調整）	215	56%	93%	0.73	-0.24
Publish Vulnerability Card（脆弱性カード）	215	68%	95%	0.51	-0.11
Publish Vulnerability Note（脆弱性ノート）	215	68%	95%	0.50	-0.41
Publish Technical Alert *（テクニカルアラート）	215	61%	87%	0.94	0.25
Publish Security Alert *（セキュリティアラート）	212	50%	87%	1.05	0.37
Publish Special Communication *（会員への情報提供）	210	73%	98%	0.37	0.09
Publish Current Activity *（関連活動状況の共有）	214	36%	72%	1.53	-0.15

表 13：手作業で修正あるいは作成されたツリーによる予想と実際の誤差分布

タスク	誤差						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
全体的パフォーマンス	4	57	255	1,280	236	54	35
Assign Analyst (D1) （分析担当者割り当て）			5	194	16		
Perform Surface Analysis （一次分析）	1	1	4	190	1	2	16
Perform Technical Analysis （詳細分析）	2	2	40	142	18	9	2
Coordinate （連絡調整）		26	53	97	28	8	3
Publish Vulnerability Card （脆弱性カード）		4	31	74	22	1	
Publish Vulnerability Note （脆弱性ノート）		4	27	41	4		
Publish Technical Alert * （テクニカルアラート）	1	6	18	131	38	15	6
Publish Security Alert * （セキュリティアラート）		2	11	135	45	16	6
Publish Special Communication * （会員への状況提供）		2	18	154	34		2
Publish Current Activity * （関連活動状況の共有）		10	48	122	30	3	
			設計目標（NMR）				

表 14 と表 15 は、KENGINE が生成したディシジョンツリーの正確性を示す統計値と、分析担当者が修正あるいは作成したディシジョンツリーの正確性を示す統計値との誤差

を示している。誤差は、手作業で修正あるいは作成したツリーでの統計値から、KENGINE により生成されたツリーでの統計値を差し引いた値で表わされている。空白のセルは、誤差がなかったことを意味する。アスタリスクが付いたタスクは、手作業でディシジョンツリーが作成されたことを示している。

表 14：ツリーの作成方法による正確性の変化（*修正ツリーの統計値 – 自動生成ツリーの統計値*）

タスク	サンプルサイズ	ヒット率	NMR	MSE	ME
全体的パフォーマンス	-6	-4%	-3%	+0.16	-0.02
Assign Analyst (D1) （分析担当者割り当て）					
Perform Surface Analysis （一次分析）				-0.04	+0.01
Perform Technical Analysis （詳細分析）		-2%		+0.08	-0.01
Coordinate （連絡調整）		-2%	-1%	+0.05	-0.26
Publish Vulnerability Card （脆弱性カード）					-0.20
Publish Vulnerability Note （脆弱性ノート）		+3%	+1%	-0.07	-0.52
Publish Technical Alert * （テクニカルアラート）		+1%	-2%	+0.13	+0.17
Publish Security Alert * （セキュリティアラート）	-3	-13%	-4%	+0.38	+0.36
Publish Special Communication * （会員への情報提供）	-5	-7%	-1%	+0.10	-0.13
Publish Current Activity * （関連活動状況の共有）	+2	-21%	-22%	+0.93	+0.06

表 15：ツリー作成方法による誤差分布の変化（*修正ツリーの統計値 - 自動生成ツリーの統計値*）

タスク	誤差						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
全体的パフォーマンス	+1	-13		-92	+24	+63	+11
Assign Analyst (D1) （分析担当者割り当て）			-1	+1			
Perform Surface Analysis （一次分析）	-1		+1				
Perform Technical Analysis （テクニカルアラート）	+2	-1	+1	-4	+2		
Coordinate （連絡調整）		+2	+4	-4	-2		
Publish Vulnerability Card （脆弱性カード）			+1	-1			
Publish Vulnerability Note （脆弱性ノート）				+6	-3	-3	
Publish Technical Alert * （テクニカルアラート）		-3	-9	+2	+2	+5	+3
Publish Security Alert * （セキュリティアラート）		-4	+19	-30	-1	+8	+5
Publish Special Communication * （会員への情報提供）		+2	+18	-19	-6		
Publish Current Activity * （関連活動状況の共有）		-9	-34	-43	+32	+53	+3
			設計目標（NMR）				

興味深いことに、分析担当の専門家が手を加えたディシジョンツリーのいくつかは正確性が大幅に低下している（ヒット率と NMR が低下し、MSE が上がっている）。これは、

より複雑な意思決定において、VRDA の方が専門家より適切なディシジョンツリーを作成していることを示唆している。理論的には、入力（FACT と依存タスク）の数が増えるほど、すべての組み合わせを人的に評価することが難しくなるので、これは意外なことではない。なお、前述したディシジョンツリー上の優先度の明らかな矛盾点を修正しても、正確性に大きな影響はなかった。

また、KENGINE により生成されたディシジョンツリーを調べた結果、FACT Access Required（必要なアクセス経路）、User Interaction Required（必要なユーザの関与）、Information Source Reliability（情報源の信頼性）、および Security Product（セキュリティ製品）がディシジョンツリー上まったく考慮されていないことがわかった。脆弱性への対応を判断するためにこれらの FACT のいずれかが重要であると考えていた分析担当者にとって、これは驚きであった。これらの FACT は対応の意思決定に役立っていないと考えられるため、CERT/CC がこれらの FACT を分析するために労力を費やすのを今後やめることもあり得る。

増分分析

実用に耐える正確性を示すディシジョンツリーを作成することを目的として、どの程度の量のトレーニングデータが必要かを調べるために、CERT/CC のデータを使って増分分析が実行された。脆弱性は、データ数 215 フルセットからランダムに選択され、50、100、150、200、および 215（完全なセット）それぞれの増分セットが作られた。それぞれの増分は累積的であり、前のセットからの脆弱性を含む。前述のように、CERT/CC のフルセットは、一般的な種類の脆弱性を網羅するように選択された約 50 件の脆弱性と、NVD からおおまかな年代順に選択されたその他の脆弱性で構成されていた。この分析では、相関係数（*r*）によって、サンプルサイズに対するヒット率および MSE の直線近似が測定される。*r* の値が 1.00 の場合は、完全な直線関係がある。*r* の値が小さくなるほど直線近似が低下するが、それによってほかの種類の関係が妨げられることはない。

表 16 と表 17 では、異なるサンプルサイズでのヒット率と MSE を比較している。

表 16：増分ごとのヒット率

タスク	サンプルサイズ					r
	50	100	150	200	215	
全体的パフォーマンス	69%	71%	69%	70%	70%	0.19
Assign Analyst (D1) （分析担当者割り当て）	92%	90%	90%	89%	90%	-0.81
Perform Surface Analysis （一次分析）	72%	89%	87%	88%	88%	0.73
Perform Technical Analysis （詳細分析）	66%	61%	66%	68%	68%	0.60
Coordinate（連絡調整）	54%	59%	58%	58%	58%	0.61
Publish Vulnerability Card （脆弱性カード）	72%	72%	69%	66%	68%	-0.91
Publish Vulnerability Note （脆弱性ノート）	76%	72%	68%	66%	66%	-0.98
Publish Technical Alert （テクニカルアラート）	62%	64%	54%	59%	60%	-0.44
Publish Security Alert （セキュリティアラート）	72%	65%	56%	62%	63%	-0.65
Publish Special Communication （会員への情報提供）	68%	79%	76%	80%	80%	0.81
Publish Current Activity （関連活動状況の共有）	54%	58%	59%	60%	57%	0.66
	増分ごとのヒット率					

表 17：増分ごとの MSE

タスク	サンプルサイズ					r
	50	100	150	200	215	
全体的パフォーマンス	0.50	0.50	0.56	0.58	0.56	0.90
Assign Analyst (D1) （分析担当者割り当て）	0.08	0.10	0.10	0.11	0.10	0.81
Perform Surface Analysis （一次分析）	0.46	0.66	0.72	0.85	0.83	0.97
Perform Technical Analysis （テクニカルアラート）	0.58	0.66	0.61	0.58	0.56	-0.43
Coordinate（連絡調整）	0.84	0.70	0.75	0.69	0.68	-0.81
Publish Vulnerability Card （脆弱性カード）	0.40	0.45	0.48	0.55	0.50	0.92
Publish Vulnerability Note （脆弱性ノート）	0.36	0.45	0.49	0.55	0.57	0.99
Publish Technical Alert （テクニカルアラート）	0.56	0.50	0.77	0.85	0.81	0.89
Publish Security Alert （セキュリティアラート）	0.76	0.49	0.70	0.71	0.67	0.07
Publish Special Communication （会員への状況提供）	0.38	0.30	0.35	0.28	0.27	-0.80
Publish Current Activity （関連活動状況の共有）	0.58	0.66	0.60	0.60	0.60	-0.09
	増分ごとの MSE					

全体的に、ヒット率と MSE はサンプルサイズとの間に明確な相関関係は見られなかった。したがって、ランダムに選択された 50 のサンプルサイズは、VRDA が実用に耐える正確なディシジョンツリーを生成するために必要とするトレーニングデータの最小数であると言える。

タスク Publish Vulnerability Note は一貫して、サンプルサイズが大きくなるにつれてヒット率が下がり（r = -0.98）、MSE が上がった（r = 0.99）。つまり、タスク Publish Vulnerability Note はサンプルサイズが大きくなるほど正確性は下がった。同様にタスク Perform Surface Analysis も、サンプルサイズが大きくなると MSE は上がった（r = 0.97）。これらのタスクは、サンプルサイズが大きくなるに連れて正確性がわずかに低下を示す例である。逆に、タスク Coordinate、Publish Special Communication、および Publish Current Activity の正確性は、サンプルサイズが大きくなるに連れてやや向上している。

結論

本調査における 3 つの参加組織の調査データの概要、対応プロセスの複雑度、および、VRDA の実装である KENGINE による対応予想の正確性を、表 18 に要約する。

表 18：参加組織の調査データの概要と VRDA の正確性

参加組織	調査データ			複雑度	正確性		
	脆弱性	サンプルサイズ	収集期間	ツリー平均ノード	ヒット率	NMR	MSE
参加組織 A	71	341	1 年	15.0	67%	88%	0.81
IIJ Technology	63	315	1 年	2.2	100%	100%	0.00
CERT/CC	215	2,147	4 か月	25.1	70%	94%	0.56

本調査から、次のような暫定的な結論を導き出すことができる。

- VRDA により提示される対応予想は、実用的な設計上の目標（NMR）から見て妥当な正確性を示している。
- VRDA によって機械的に生成されたディシジョンツリーは、専門の分析担当者によって作成されたディシジョンツリーと比較して、同等あるいはより高い正確性を示す。
  - 専門家が意思決定プロセスを熟知している比較的単純なタスクの場合、専門家は正確なツリーを作成できる。
  - より複雑なタスクや、専門家が意思決定プロセスをよく理解していないタスクの場合は、VRDA の方がより正確である。
- 比較的単純な意思決定プロセスにより対応判断されるタスクの場合、VRDA は非常に高い正確性を示すことができる（IIJ Technology のデータで実証されたとおり）。

VRDA の対応予想の正確性と直接関係はないが、タスクと FACT の把握や整理、FACT 分析や対応判断などのデータの入力、VRDA が提示した対応予想が実際の対応と大きく異なった場合の調査といった行為によって、副次的な効果も確認された。例えば、組織

の脆弱性対応プロセス上の誤りが起きやすく問題に繋がりうる点が明らかになり、CERT/CC において、いくつかの FACT の定義に見直しが入り、より明確に定義された。

また、KENGINE が生成したディシジョンツリー上に、これまで分析担当者が意思決定のために重要であると考えていた一連の FACT が考慮されていないことに驚かされた。

本調査を通して、VRDA に対するいくつかの修正と、さらなる調査に関するアイデアが得られた。

VRDA の欠点の 1 つは、タスクを実施するコストを直接考慮しないことである。タスクの対応判断を行う専門の分析担当者が優先度を選択するときにおそらくタスクのコストを考慮するため、タスクのコストはある程度まで意思決定モデルに組み込まれる（そうでなければ、分析担当者は単にすべての脆弱性情報についてすべてのタスクを最も高い優先度で実施する判断を行うかも知れない）。

KENGINE の実装、あるいは、おそらく VRDA そのものは、非論理的に見えるディシジョンツリーを生成することがある。このようなディシジョンツリーは、トレーニングデータが示す内容と正しく一致していれば、実際に正確である可能性がある。この考えは、論理的な矛盾点を含んだ KENGINE により生成されたディシジョンツリーと、これら矛盾点を修正したディシジョンツリーの比較によって裏付けされている。この比較では、矛盾を取り除くように修正されたディシジョンツリーにおいて正確性の著しい向上は見られない。しかし、KENGINE あるいは VRDA のディシジョンツリー作成アルゴリズムに、小さな欠陥や十分に理解されていない振る舞いがある可能性もあるため、同アルゴリズムに、ディシジョンツリーの論理的一貫性を強制するための修正を加えることも考えられる。

KENGINE を改善することによって、提示される対応予想の正確性が向上する可能性がある。ディジョンツリー生成アルゴリズムで使われている静的に定義されたパラメータを、ユーザがある程度調整可能とすることで、ディシジョンツリーの生成処理を、より細かく制御できると考えられる。また、ディシジョンツリーの管理ユーザインタフェースがより直感的になれば、分析担当者がより正確で複雑なディシジョンツリーを作成する手助けになるであろう。

意思決定支援およびエキスパートシステムについてよく触れられる理念として、正確なシステムの作成には十分かつ正確な知識が不可欠であり、意思決定のメカニズム自体はそれほど重要ではないというものである。この考えに従うと、意思決定に使用される入力データ（FACT）を改良する方がより生産的なアプローチであるということになる。これは、脆弱性情報をはじめとする脅威分析情報の定義とエンコードの方法を改良したり、本調査で使用されている FACT セットに含まれていないが意思決定により強く影響する新しい情報（FACT）を特定したりすることを意味する。また、別のアプローチとして、ディシジョンツリーごとに単純さと正確性のバランスを比較検討し最適解を求める考えもある。「節約の法則（parsimony）」では、すべてのことが同じであれば、「より単純なモデルの方が複雑なモデルより好ましい」と述べられている[15]。例えば、単純なツリーと複雑なツリーの 2 つがあり、これらが同じ正確性を示す場合、単純なツリーの方が好ましい。VRDA のさらなるパフォーマンス向上のために、代替的なディシジョンツリー生成メカニズムや、前向き連鎖などの推測ルールの適用検討も考えられる。

まだ試されていない VRDA の別の側面として、既存の脆弱性脅威評価基準および脆弱性データとの統合がある。KENGINE はスタンドアロン実装であるが、VRDA のコンセプ

[15] 出典：“On the Problem of Selecting Categories and Model Subsets in Decision Trees”（http://www.decisionsciences.org/Proceedings/DSI2008/docs/330-5989.pdf）

トはほかの脆弱性管理および優先順位付けツールに組み込むこともできる。例えば、VRDA 意思決定支援システムを、NVD が提供する CVSS（Common Vulnerability Scoring System）[16]データなど、ソースが異なる脆弱性データと組み合わせることもできる。他のツールやデータソースとの統合だけでは VRAD そのものの有効性を測定できないが、KENGINE と比較することで、VRDA の異なる評価を得ることができる。

[16] http://www.first.org/cvss/

付録 A：参加組織 A の情報

参加組織 A は日本の大手多国籍企業の本社の CSIRT である。参加組織 A は、脆弱性情報の分析と展開を自社の各部署に対して行っている。

参加組織 A のタスク

Detail Analysis (D1)（詳細分析）
報告された脆弱性の詳細を分析する。

Warning Level Alert（警戒レベル注意喚起）
脆弱性情報を参考情報として提供するための準備をする。

Inform All Contact Points（すべての連絡先へ通知）
すべての連絡先に脆弱性情報（Warning Level Alert または Critical Level Alert）を通知する。

Inform Specific Contact Points（特定の連絡先へ通知）
特定の連絡先に脆弱性情報（Warning Level Alert または Critical Level Alert）を通知する。

Critical Level Alert（重大レベル注意喚起）
脆弱性の対応勧告を行う。

参加組織 A の脆弱性対応プロセス

参加組織 A における脆弱性対応プロセスの概要を図 6 に示す。

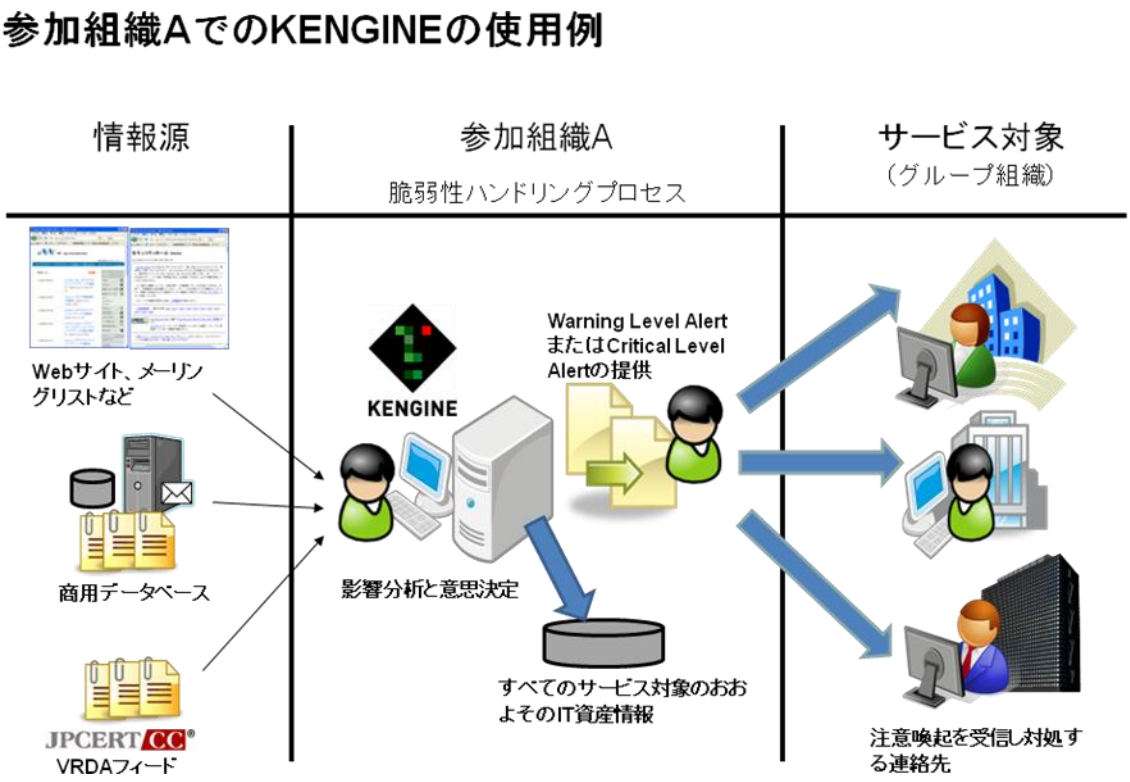


図 6：参加組織 A での使用例

付録 B：IIJ Technology の情報

IIJ Technology は、幅広い規模のネットワークやシステムの構築・運用サービスを提供している。IIJ Technology は、同社の顧客のシステム構成に即した脆弱性情報を収集、影響度を判定した上で、情報提供を行い顧客のセキュリティ対応活動を支援するサービス（*脆弱性情報判定サービス*）を提供している。このため、IIJ Technology は同社の顧客が使用している IT 資産に関する正確なインベントリおよび配備情報を維持管理している。本調査において、VRDA を評価するために、IIJ Technology は、特定の顧客に対して実施した脆弱性対応業務から過去約 1 年間に相当する実績データを使用した。

IIJ Technology のタスク

Ignore（無視）

サービス対象である特定顧客において、影響を受ける製品が使用されていない場合は、脆弱性情報に対して対応を行わない。

Emergency Level Alert（緊急レベル注意喚起）

最も高いレベルの注意喚起をサービス対象に提供する。機密データが保存されているホストに対するインターネットからの攻撃や、特権ユーザとしてのアクセス権の取得が可能となる攻撃、またはその両方の攻撃の危険性がある場合に発行する注意喚起。

Critical Level Alert（重大レベル注意喚起）

中レベルの注意喚起をサービス対象に提供する。機密データが保存されているホストに対するインターネットからの攻撃、非特権ユーザとしてのアクセス権の取得が可能となる攻撃、またはその両方の攻撃の危険性がある場合に発行する注意喚起。

Warning Level Alert（警戒レベル注意喚起）

最も低いレベルの注意喚起をサービス対象に提供する。ネットワーク内部からの攻撃の可能性がある場合に発行する注意喚起。

FYI（For Your Information）（参考）

参考情報として脆弱性情報を提供する。ローカルユーザからの攻撃が可能な脆弱性などの場合をカバーする。

IIJ Technology の脆弱性対応プロセス

IIJ Technology における脆弱性対応プロセスの概要を図 7 に示す。

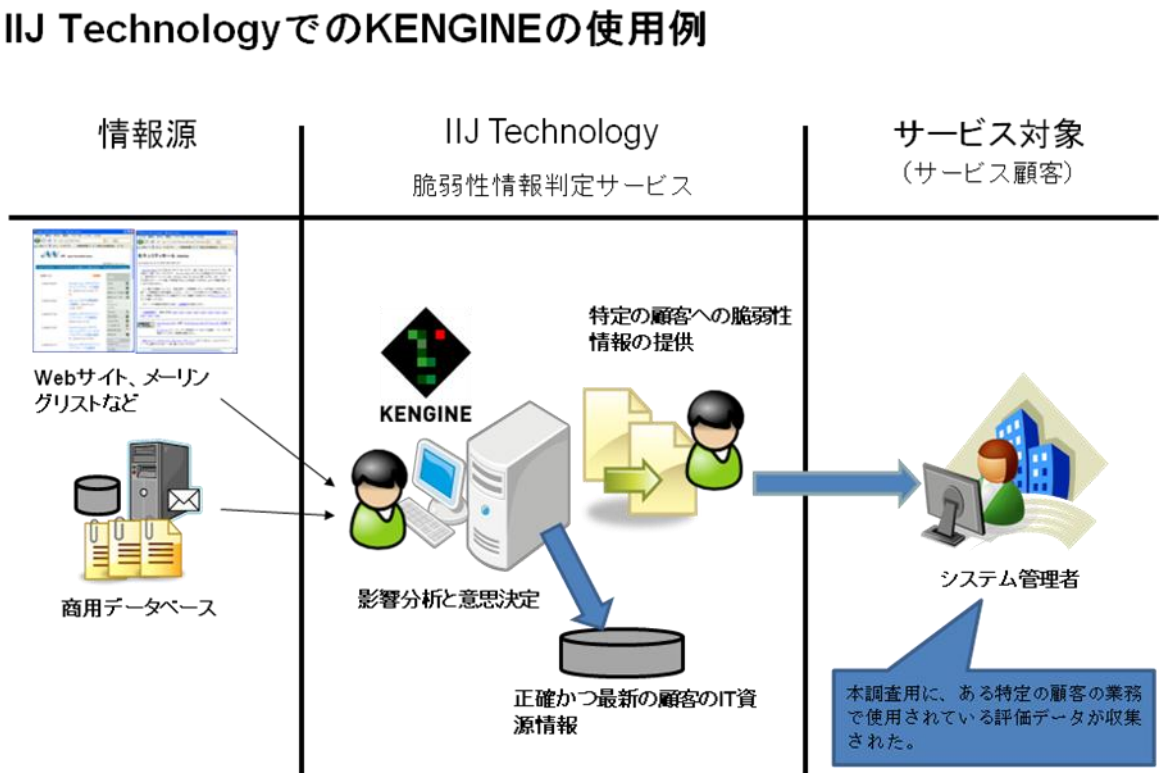


図 7：IIJ Technology での使用例

付録 C：CERT/CC の FACT とタスク

CERT/CC の FACT

製品または技術に関連する FACT は、LAPT（Lightweight Affected Product Tag）としてマーク付けされる。LAPT FACT は、脆弱性そのものの特徴ではなく、脆弱性の影響を受ける可能性がある製品または技術に関して脅威分析影響を与える特徴を記述する。

Direct Report（直接報告）（D1）

脆弱性は非公開で CERT/CC に直接報告されたものか。

{はい|いいえ}

Direct Report は、通常は脆弱性情報を公開する前に対応を調整するため、CERT/CC に提供される機密報告である。

Control System Product（制御システム製品）（LAPT）

影響を受ける製品は制御システム製品か。

{はい|いいえ}

制御システム製品の主な目的は、重要な産業、公益事業、エネルギー、または交通機関を支援することである。これには、SCADA システムと DCS システムが含まれる。たまたま制御システムによって使用される商品や一般の IT 製品および技術は、制御システム製品とは見なされない。

Network Infrastructure Product（ネットワークインフラストラクチャ製品）（LAPT）

その脆弱性はネットワークインフラストラクチャ製品に影響するか。

{はい|いいえ}

ネットワークインフラストラクチャ製品は、中核のインターネット機能（例えば、DNS、ルーティングプロトコル（BGP）、一般的なネットワークプロトコルなど）を支える。当該製品または技術なしではインターネットの機能が大幅に低下する場合、この質問への答えは「はい」になる。

Security Product（セキュリティ製品）（LAPT）

その脆弱性はセキュリティ製品に影響するか。

{はい|いいえ}

セキュリティ製品は、セキュリティを向上させることを主な目的とし、それを期待される製品である。例えば、ファイアウォール、暗号化システムなどがある。

Ubiquity（普及率）（LAPT）

脆弱性のあるシステムのサービス対象（Constituency）内での普及率（システム数）はどの程度か。

{なし|低|低－中|中－高|高}

この FACT は、過去に「Population」と呼ばれていた。

CERT/CC は、ある製品の Ubiquity を評価する際、インターネット全体を考慮する。この FACT は、インストール数とユーザベースの両方を考慮する。例えば、広く使用されている Web アプリケーションは、1 つのインスタンスとして存在しているだけでも、多くのユーザがいるため、十分に Ubiquity の評価値が高くなる可能性がある。各組織においては、より具体的なサービス対象範囲があり、おそらくより厳密な定義が存在するものと考えられる。Ubiquity は、システムの総数に対して脆弱性があるシステムの比率として計算することができる。

- 低：～<1%。ほとんどのソフトウェア。
- 低－中：～2－5%。Apple Safari、Apple Mac OS X などのソフトウェア。
- 中－高：～5-20%。Mozilla Firefox、Apache HTTP Server、Cisco IOS などのソフトウェア。
- 高：インターネットシステム、製品、および技術の合計の 20%以上のもの。これより高比率のカテゴリはない。ユーザ数は数千万から数億にも達する。Microsoft Windows、Adobe Flash、TCP/IP、BIND、GNU/Linux カーネル、sendmail などのソフトウェアが含まれる。

Population Importance（影響を受けるシステムの重要性）（LAPT）

サービス対象（Constituency）内での脆弱性の影響を受けるシステムの重要性はどの程度か。

{低|低－中|中－高|高}

以下の例はガイドラインである。組織またはサービス対象は、ほかの基準を使ってユーザにとっての重要性を決めることもできる。

- 低：家庭のデスクトップ。
- 低－中：業務用のデスクトップ、家庭用（業務用以外）のサーバ。
- 中－高：重要でない業務用サーバ、重要でないネットワークサーバ。
- 高：重要な業務用サーバ、重要なネットワークサーバ。

サービス対象内で異なる役割を担っている各システムに脆弱性が影響する場合、これらシステムの重要性を判断するには、2 つの考え方がある。例えば、数多くの業務用 Windows デスクトップ（低－中）と、数える程度の Windows データベースサーバ（高）に影響する脆弱性の場合、脆弱性の影響を受けるシステムの重要性は次の 2 つの方法で判断できる。

1. 影響を受けるシステムの数による判断

デスクトップシステムの数は非常に多く、データベースサーバの数は少ないため、影響を受けるシステムの重要性を 低－中 とする。

2. 影響を受けるシステムの最も高い重要性による判断

データベースサーバの重要性が高なので、影響を受けるシステムの重要性は高になる。

Impact（影響）

システムに対する脆弱性の悪用が成功した場合の影響はどの程度深刻か。

{低|低－中|中－高|高}

- 低：機能の低下やサービスの中断。
- 低－中：サービス不能状態、クラッシュ、情報漏えい、検出回避。
- 中－高：一般ユーザ権限での任意コード実行、一般ユーザ権限取得。
- 高：管理者権限での任意コード実行、昇格された権限の取得。

Information Source Reliability（情報源の信頼性）

脆弱性情報の最初または主要な情報源はどの程度信頼できるか。

{低|中|高}

- 低：一般的なセキュリティ Web サイトまたはフォーラム、セキュリティ業界紙。
- 中：セキュリティベンダ。
- 高：ベンダ、CSIRT、脆弱性調整担当者。

Access Required（必要なアクセス経路）

脆弱性を悪用するために攻撃者が必要とするアクセス経路。

{ルーティングあり|ルーティングなし|ローカルマシン|物理アクセス}

この FACT は、攻撃者が脆弱性を悪用するために必要とする物理的/論理的近接性を測定する。

- ルーティングあり：インターネットを介して攻撃できる。
- ルーティングなし：Ethernet などのローカルセグメントから攻撃する必要がある。また、Bluetooth や 802.11 などの物理的近接性を含む。
- ローカルマシン：攻撃者は、攻撃対象のシステム上のある種のセッション、シェル、またはリモートデスクトップを必要とする。
- 物理アクセス：攻撃者が攻撃対象のシステムに接触するか、物理コンソールにログインする必要がある。

Authentication（認証）

脆弱性を悪用するために攻撃者が必要とする認証のレベル。

{なし|限定的|標準|特権}

- なし：匿名または認証なし（IP アドレス、rhost、および匿名 FTP アクセスは、「なし」と見なされる）。
- 限定的：自己登録によるおそらく検証済みの電子メールアドレス。
- 標準：管理者によって意図的に与えられたログイン。
- 特権：特定のログインまたはグループメンバシップ（root、bin、Administrators など）が必要。

User Interaction Required（必要なユーザの関与）

攻撃者が脆弱性を悪用するために必要となる、攻撃者以外の者（潜在的な被害者）によるアクションは何か。

{不要|簡単|複雑}

正当なユーザ（攻撃者ではなく悪意もないユーザ ー おそらく犠牲者）によるアクションを必要とする脆弱性の場合、攻撃者がタイミングよくそのアクションを発生させるのがどの程度困難か。

- 不要：正当なユーザがアクションを実行しなくても脆弱性を悪用できる（例えばリスニングサービスに対する攻撃など）。
- 簡単：常識的な人にとって有害には見えない標準的アクション（リンクのクリックやファイルの表示など）を実行するように、正当なユーザを納得させる必要がある。
- 複雑：難しいアクションや疑わしいアクションを実行するように、正当なユーザを納得させる必要がある。正当なユーザが上位の権限を持っている必要がある場合は、そのユーザがより懐疑的になる可能性が高い。

Technical Difficulty（技術的な難易度）

攻撃者が脆弱性を悪用しようとしたときに直面する技術的な困難の程度。

{低|低ー中|中ー高|高}

この FACT はツールを考慮しない。例えば、ポイントアンドクリック攻撃ツールがあるという理由で脆弱性の技術的難易度を低いと見なすことはない。

- 低：専門知識、運、またはその両方が不要であるか、ほとんど必要ない（クロスサイトスクリプティングなど）。
- 低ー中：多少の専門知識、運、またはその両方が必要（ほとんどのバッファオーバーフロー、小規模および/または制御されたスペースでの正しい推測、Windows 関数呼び出しの知識）。
- 中ー高：専門知識、運、またはその両方が必要（中規模なスペースでの正しい推測、カーネルの専門知識）。
- 高：膨大な専門知識、運、またはその両方が必要（BIOS の専門知識、大規模および/または変化し続けるスペースでの正しい推測）。

Availability of Remediation（対策の有無）

提供されている解決策、回避策、それ以外の対策のレベル。

{なし|非公式パッチあり|公式回避策あり|公式パッチあり}

- なし：何も提供されていない。
- 非公式パッチあり：非公式な回避策や非公式パッチが提供されている。
- 公式回避策あり：ベンダから公式な回避策が提供されている。
- 公式パッチあり：ベンダから公式パッチが提供されている。

Incident Activity（インシデントの発生状況）

この脆弱性に関して見られたインシデント活動のレベル。

{活動なし|Exploit あり|活動あり}

- 活動なし：知られている Exploit（悪用）もインシデントもない。
- Exploit/PoC あり：Exploit コード、もしくは脆弱性を検証する PoC code が存在する。
- 活動あり：インシデントが報告されている。

Quality of Public Information（一般公開情報の質）

脆弱性について入手できる一般公開情報の質。

{許容不能|許容可能|高}

- 許容不能：公開データがない、または公開データがひどく欠けているか誤解を招きやすい。
- 許容可能：公開データが提供されているが、いくつかの領域で欠落がある。
- 高：理解しやすく、信頼できる、完全な公開データ。

Public Attention（一般の関心）

脆弱性が一般から受けている注目の程度。

{なし|低|低－中|中－高|高}

- なし：一般が注目しているという情報や公開の情報は認識されていない。
- 低：標準的なバグ/脆弱性メーリングリスト/フィード/サイトで論じられている。
- 低－中：標準リストおよびセキュリティ業界紙で論じられている。
- 中－高：標準リスト、複数の業界紙による発表、およびおそらく地方または地域の報道機関の外で論じられている。
- 高：全国的または国際的な大手メディアによってカバーされている。

CERT/CC のタスク

特に断りがない限り、すべてのタスクに相対的優先度を示す 4 つのレベル、「*Must（必須）*」、「*Should（推奨）*」、「*Might（検討）*」、および「*Won't（不要）*」がある。VRDA は予想を行うときにコストを直接考慮することはないが、タスク別に実施するための労力の側面でおおよそのコストが示されている。また、前行程としての依存関係があるタスクの有無も示されている。

Assign Analyst（分析担当者割り当て）（D1）

報告を分析担当者に割り当てるべきか。このタスクのオプションは、「*Must（必須）*」と「*Won't（不要）*」の 2 つだけである。このタスクに答えるために必要な FACT は、脆弱性情報を最初にカタログ化する段階に収集される。LAPT FACT に対する分析は、既存の LAPT データから得ることができる。3 つの FACT、*Direct Report（直接報告）*、*Ubiquity（普及率）*、および *Impact（影響）*が必要である。他の関連する LAPT FACT は、*Population Importance（影響があるシステムの重要性）*、*Network Infrastructure Product（ネットワークインフラストラクチャ製品）*、および *Control System Product（制御システム製品）*である。

労力：15 分

依存関係：なし。

Perform Surface Analysis（一次分析）

分析担当者が報告を検討して背景調査に適度な労力を費やすべきか。Assign Analyst（D1）タスクに答える作業の一環として、ある程度の表層分析が実行される。

労力：30 分

依存関係：Assign Analyst（D1）

Perform Technical Analysis（詳細分析）

分析担当者が脆弱性の技術的側面を理解するために、多くの労力を費やすべきか。これには、脆弱性の再現とリバースエンジニアリングが含まれる。

労力：場合によって異なる（2 時間～5 日）。

依存関係：Perform Surface Analysis

Coordinate（連絡調整）

分析担当者がベンダおよびほかの関係者とコミュニケーションをとるべきか。

労力：ベンダの数、期間、および複雑度によって大きく異なる（1 時間～7 時間）。

依存関係：Perform Surface Analysis

Publish Vulnerability Card（脆弱性カード）

分析担当者が脆弱性カード（Vulnerability Card）を公開するべきか。

労力：1～4 時間

依存関係：Perform Surface Analysis

Publish Vulnerability Note（脆弱性ノート）

分析担当者が脆弱性メモ（Vulnerability Note）を公開するべきか。

労力：1 時間～2 日

依存関係：Perform Surface Analysis および Publish Vulnerability Card

Publish Technical Alert（テクニカルアラート）

分析担当者が US-CERT Technical Cyber Security Alert を公開するべきか。

労力：2 時間～1 日

依存関係：Perform Surface Analysis

Publish Security Alert（セキュリティアラート）

分析担当者が US-CERT Cyber Security Alert を公開するべきか。

労力：1～4 時間

依存関係：Perform Surface Analysis

Publish Special Communication（会員への状況提供）

分析担当者が特別なコミュニケーションをとるべきか。

労力：1～4 時間

依存関係：Perform Surface Analysis

Publish Current Activity（関連活動状況の共有）

分析担当者が US-CERT Current Activity エントリを公開するべきか。

労力：1～2 時間

依存関係：Perform Surface Analysis

付録 D：タスクへの対応予想・実績の分散

表 19、表 20、および表 21 は、それぞれ各参加組織において KENGINE により提示された対応予想値と実際の対応値の平均偏差と標準偏差を示している。標準偏差が小さい場合は、そのタスクが意思決定ではなく、ケースバイケースの意思決定が不要な、より静的な通常の手順であるか、あるいは、タスクが一貫して同じように回答されている可能性がある。標準偏差が小さいタスクの対応判断プロセスを調べることで、組織の脆弱性情報ハンドリング手順の見直しに繋がる可能性がある。

表 19：参加組織 A のタスクごとの分散

タスク	サンプルサイズ	予想		実際	
		平均	標準偏差	平均	標準偏差
すべてのタスク	341	0.97	1.13	0.86	1.28
Detail Analysis（詳細分析）（D1）	66	2.21	0.77	1.89	1.40
Warning Level Alert（警戒レベル注意喚起）	67	0.93	1.05	0.81	1.20
Inform All Contact Points（すべての連絡先へ通知）	69	0.81	1.10	0.65	1.16
Inform Specific Contact Points（特定の連絡先へ通知）	69	0.77	1.07	0.62	1.13
Critical Level Alert（重大レベル注意喚起）	70	0.19	0.39	0.39	0.97

表 20：IIJ Technology のタスクごとの分散

タスク	サンプルサイズ	予想		実際	
		平均	標準偏差	平均	標準偏差
すべてのタスク	315	0.20	0.40	0.2	0.40
Ignore（無視）	63	0.00	0.00	0.00	0.00
Emergency Level Alert（緊急レベル注意喚起）	63	0.00	0.00	0.00	0.00
Critical Level Alert（重大レベル注意喚起）	63	0.00	0.00	0.00	0.00
Warning Level Alert（警戒レベル注意喚起）	63	0.16	0.37	0.16	0.37
FYI（参考）	63	0.84	0.37	0.84	0.37

IIJ Technology における予想の正確性は 100 パーセントであった。その主な理由は、比較的単純な脆弱性対応プロセス（対応がブール値のタスクが 5 個）、一貫した対応ポリシー、そして顧客インベントリと配置情報が詳細に把握できたことである。Ignore、Emergency Level Alert、および Critical Level Alert タスクはまったく実行されず（これらのタスクは提示された対応の平均も実際の対応の平均も 0、すなわち「*Won't（不要）*」）、意思決定から静的プロセス（脆弱性情報を決して無視せず、決して Critical Level Alert を発しない）に移行させる候補になり得る。しかし、これらのタスクを意思決定ポイントとして残すのには別の理由がある。本調査に用いたサンプルセット内の脆弱性情報を調べた結果、これらのタスクは実際にはどれも Critical Level Alert または Emergency Level Alert を実施する条件に該当していないことがわかった。また、本調査の対象となった脆弱性はすべてタスク Ignore（無視）の実施が適用されないものであった（実際に無視された脆弱性情報は調査対象データに含まれなかった）。したがって、このセット内のすべての脆弱性情報について、タスク Ignore の対応予想が「Ignore = 不要」（つまり、無視しない）となるものと予想される。

表 21：CERT/CC のタスクごとの分散

タスク	サンプルサイズ	予想		実際	
		平均	標準偏差	平均	標準偏差
すべてのタスク	2,147	0.92	1.05	0.97	1.05
Assign Analyst（分析担当者割り当て）（D1）	215	0.61	0.49	0.66	0.47
Perform Surface Analysis（一次分析）	215	1.84	1.46	2.04	1.38
Perform Technical Analysis（詳細分析）	215	1.10	0.94	1.08	0.90
Coordinate（連絡調整）	215	0.63	0.68	0.65	0.87
Publish Vulnerability Card（脆弱性カード）	215	0.93	0.85	1.02	0.89
Publish Vulnerability Note（脆弱性ノート）	215	0.91	0.84	1.03	0.89
Publish Technical Alert（テクニカルアラート）	215	0.70	1.05	0.78	1.01
Publish Security Alert（セキュリティアラート）	215	0.74	1.06	0.75	1.01
Publish Special Communication（会員への状況提供）	215	0.00	0.00	0.21	0.47
Publish Current Activity（関連活動状況の共有）	212	1.70	1.06	1.49	1.09

付録 E：コスト見積り

VRDA が提示する対応予想に十分な正確性があれば、VRDA を使って必要な対応コスト（労力）を見積ることができる。CERT/CC は一例として、予想された対応と実際の対応の相違が 1 を超えない（NMR の範囲内）タスクを見積り対象とした。VRDA が提示する対応予想とタスクごとの労力の見積り（付録 C を参照）を使用すれば、優先度別にタスクを処理するために必要なリソースを見積ることができる。コストの見積りを表 22 に示す。

表 22：CERT/CC の労力見積り（時間単位に四捨五入されている）

タスク	優先度		
	必須	推奨	検討
すべてのタスク	408	3,643	3,958
Assign Analyst（分析担当者割り当て）（D1）	54	54	54
Perform Surface Analysis（一次分析）	65		
Perform Technical Analysis（詳細分析）	21	2,142	483
Coordinate（連絡調整）	29	485	2,537
Publish Vulnerability Card（脆弱性カード）	3	168	145
Publish Vulnerability Note（脆弱性ノート）	34	459	570
Publish Technical Alert（テクニカルアラート）	75	175	45
Publish Security Alert（セキュリティアラート）	40	88	33
Publish Special Communication（会員への状況提供）			
Publish Current Activity（関連活動の状況共有）	89	74	93

優先度ごとの労力は累積された値ではない。この見積りによると、215 件の脆弱性情報から予想されたすべての「*Must （必須）*」タスクを CERT/CC が完了させるには、408 時間分の労力が必要である。すべての「*Must （必須）*」タスクと「*Should（推奨）*」タスクを完了させるには、408 時間と 3,643 時間を合計した 4,051 時間分の労力が必要である。

D1 タスク Assign Analyst に回答するためには、最終的な優先度の決定に関係なく、215 件の脆弱性すべてに対して労力を費やす必要がある。Perform Technical Analysis や Coordinate などいくつかのタスクでは、必要な労力が大きく異なるため、より正確な見積りのためにはタスクをより注意深く分類する必要がある（例えば、必要な労力が少ない Coordinate タスク、必要な労力が中程度の Perform Technical Analysis タスクなど）。

付録 F：脆弱性情報の MSE の分布

参加組織 A と CERT/CC の脆弱性情報の MSE の分布を、それぞれ図 8 と図 9 に示す。脆弱性情報の MSE 分布とは、脆弱性情報毎の対応予想の平均 MSE の分布である。例えば、図 8 において、すべてのタスクの対応予想の平均 MSE が 0（全タスクについて予想と実績の完全一致）の脆弱性情報が 15 件であることを示す。

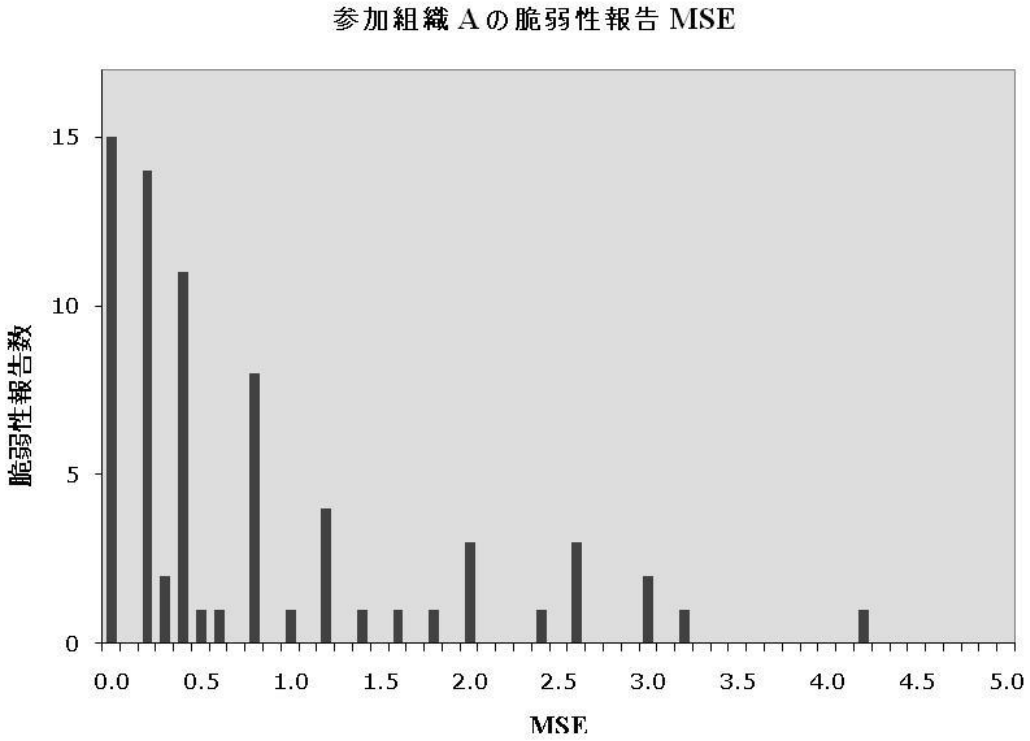


図 8：参加組織 A の脆弱性情報の MSE の分布

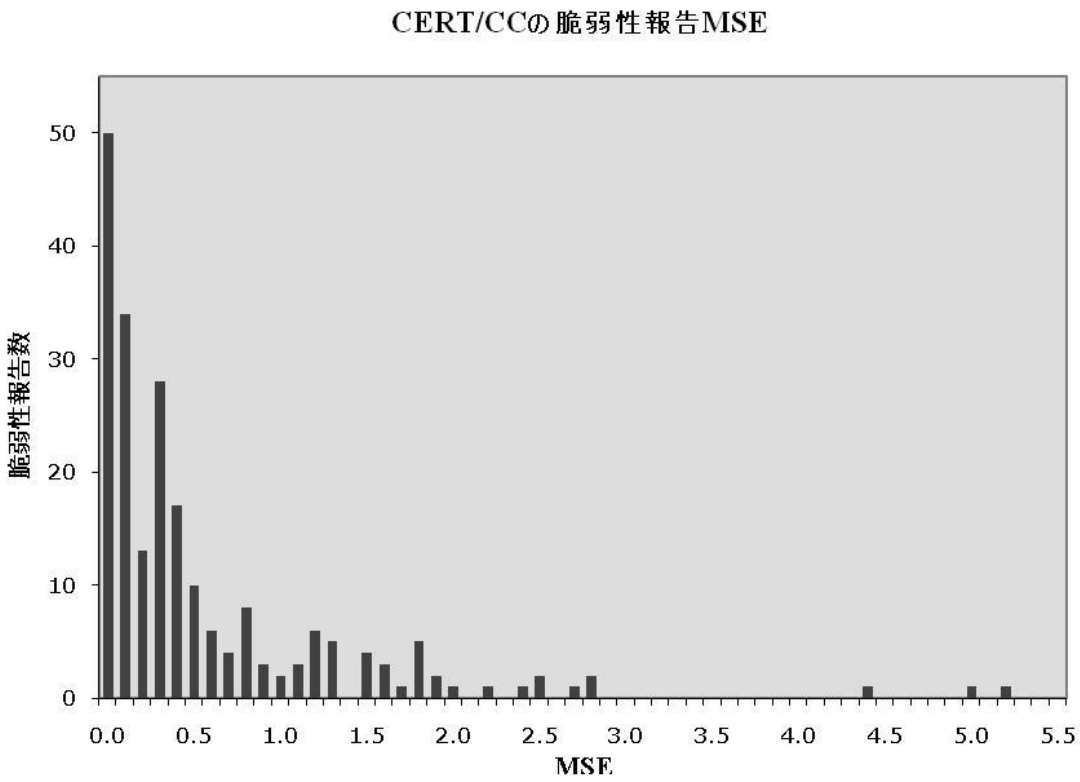


図 9：CERT/CC の脆弱性情報の MSE の分布

付録 G：脆弱性情報の深刻度

深刻度は、予想の正確性の評価に関係なく、それぞれの報告がエンコードされた FACT 値の合計としておおまかに見積られた。一例として、影響 FACT の深刻度のコードを表 23 に示す。

表 23：FACT のエンコード

影響	深刻度
高	4
中－高	3
低－中	2
低	1

値が大きくなるほど深刻度が高くなる。脆弱性の深刻度の範囲は、12～47 である。CERT/CC での脆弱性報告の深刻度の分布を図 10 に示す。

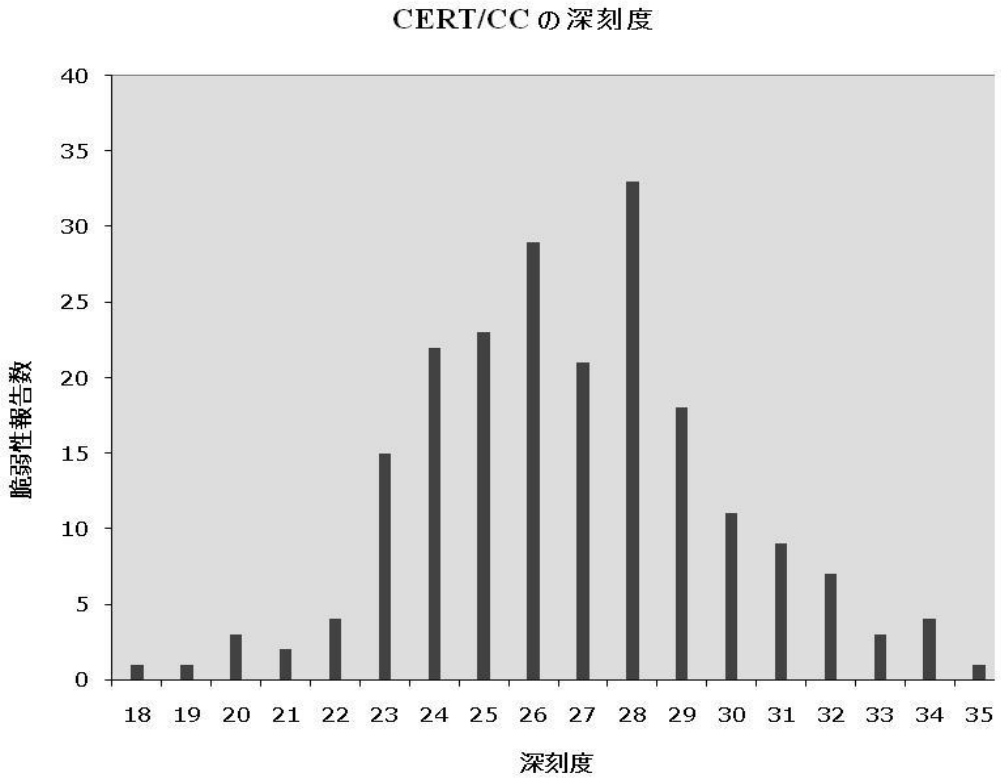


図 10：CERT/CC の深刻度分布