Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

データの乱数性に着目した暗号化レベル判定方法の提案

重本倫宏† 鬼頭哲郎† 仲小路博史† 甲斐賢†

本稿では，どの程度データを秘匿すべきかの指針策定の基礎として，暗号化範囲に着目した暗号化レベルを提案する．また，暗号化されたデータは乱数列のように見えるという特徴を利用し，アプリケーションに依存せずに，暗号化レベルを判定する方法について報告する．さらに，実トラフィックを用いた評価実験によって，提案した暗号化レベル判定方法が暗号化レベルを正しく判定していることを示す．

Proposal of Encryption Level Classification method based on Randomness of Communication Data

TOMOHIRO SHIGEMOTO† TETSURO KITO†
HIROFUMI NAKAKOJI† SATOSHI KAI†

This paper proposes a criterion of Encryption Level focusing on ranges of encryption, as a basis for designing a policy which specifies the degree of protection of communication data. Then, this paper also proposes an Encryption Level classification method by computing randomness of communication data. This method is based on observation that randomness of packet payloads in encrypted traffic is much higher than that in plain traffic, and therefore the method is application agnostic. Evaluation experiments using real traffic data shows the effectiveness of the proposed method.

1. はじめに

近年，クラウドコンピューティング（以下，クラウド）環境を利活用したサービスの提供が進展し，国民生活や社会経済活動を支える基盤インフラとなりつつある．その一方で，クラウド内部での細かいデータ運用が不透明であることから，クラウド環境を信頼性のあるプラットフォームとして認識していない利用者も存在する[1][2]. ユーザ企業の情報システム担当者を対象に行われたアンケート[3]でも，「クラウドに関する課題・不安」を問う設問では，セキュリティ・情報漏えいへの不安を非常に意識している回答者が最も多く，セキュリティへの不安がクラウド環境導入の障壁になっていることがわかる．

総務省による「ASP・SaaS における情報セキュリティ対策ガイドライン」[4]では，外部ネットワークを利用した情報交換において，情報を盗聴，改ざん，誤った経路での通信，破壊等から保護するため，通信の暗号化を行うことが推奨されている．クラウド環境に関するセキュリティにおいてもトラフィックを暗号化し，利用者の不安を払拭することが，クラウド環境成長への重要なカギになると考えられる．

一般に，クラウド利用者がクラウド環境を利用する場合には，クラウド提供者と契約を結ぶ. 契約には，提供するサービスの内容と範囲，品質に対する要求水準，また，それらが達成されなかった場合の規則（罰則や補償）などが含まれている．

クラウド提供者は契約通りにクラウド環境を運用しなければ罰則が与えられるため，契約を満たせなくなる可能性（予兆）があれば，早期に発見し，対処したいという思いがある[5]. 契約を満たせなくなる予兆として，意図しない構成変更や設定ミスなどが挙げられる．

例えば，当初は暗号化して通信されていたものが，運用の過程で，構成変更の影響，あるいは設定ミスの影響により，暗号化されずに通信が行われるようになることも考えられる．このような構成変更は，ネットワーク設定ミスや電源が落ちていることと異なり，データ自身はやり取りされるため，クラウド提供者にとって把握しづらく，対処が遅れがちになる. つまり，暗号を利用したシステムの安全性を定量的に評価し，クラウド提供者に通知することが重要な課題となる．

これらの問題を解決する技術の一つとして，暗号化通信の解析に関する研究がなされてきた[6][7][8]．文献[6]では，暗号化された VoIP トラフィックについて，どの言語が利用されているか推測できることを示している．文献[7]では，パケットサイズの時刻変動を利用した，HTTP や SMTP，SSH トラフィックの識別を提案している．文献[8]では，初期パケットの乱数性に着目し，暗号化された P2P 通信の検知を行っている．しかし，これらの研究では，特定のアプリケーションに特化した解析を行っているに過ぎず，暗号を利用したシステムの安全性評価には利用できない．

本稿では，暗号を利用しているシステムの安全性及び定量化ついて検討し，定量化方法のひとつとして，暗号化レベルを提案する．また，提案した暗号化レベルを判定する手法についても報告する．

† (株)日立製作所
Hitachi Ltd.

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

2. 暗号化レベルの提案

本章では，暗号を利用したシステムの安全性について検討し，暗号化レベルが必要とされる理由ついて述べる．

2.1 暗号を利用したシステムの安全性

暗号を利用したシステムの安全性は，一般に，後述する 3 点の安全性で論じられることが多い[9][10][11][12][13].

(1) 暗号アルゴリズムが安全である

暗号を利用したシステムの安全性を確保するためには，まず，利用する暗号アルゴリズムが安全でなければならない．暗号アルゴリズムが安全であるとは，利用する暗号アルゴリズムの強度が高いことを表す．情報を保護するためには，強力かつ耐用年数の高い暗号アルゴリズムを用いることが重要である．暗号強度の指標としては，NIST(National Institute of Standards and Technology)が採用した標準暗号[9]や電子政府推奨暗号[10]などが存在する．

(2) 暗号モジュールの実装が安全である

暗号アルゴリズムが安全であっても，正しく実装されていなければ，システム全体として，安全であるとは言えない．例えば，暗号を実装した組み込み機器において，その消費する電力や処理にかかる時間から暗号化されたデータを解読されてしまう恐れがある．暗号モジュールの実装評価には，IPA の認証制度[11]や，FIPS 140-2[12]などが存在する．

(3) 鍵管理が安全である

鍵管理についても，暗号を利用したシステムを安全に保つためには重要となる．ある製品が AES（Advanced Encryption Standard）を用いて暗号化を行ったとしても，その際に使用した鍵が生データのままレジストリなどに保存されていたら安全とは言えない．レジストリを見て，その保存されている鍵を用いて暗号化されたデータを復号される恐れがある．鍵の管理については，IPA から「安全な暗号鍵のライフサイクルマネージメントに関する調査」[13]が公表されている．

しかし，上記 3 点の安全性が確保されていたとしても，暗号が適切に利用されていなければシステム全体としての安全性は確保できない．

IPA が 2011 年 5 月に公開した『「暗号をめぐる最新の話題」に関するレポート』[14]に，大手百貨店のオンラインショッピングサイトで，Web サーバの設定ミスにより SSL プロトコルが動作しないまま運用が続いていたという事例が報告されている．本件は約 9 カ月もの間発覚せず，その間に顧客の個人情報が，暗号化されずに顧客の端末から大手百貨店のサーバに送信されていた．

このように，暗号化されるべきデータを暗号化せずに送受信していれば，システム全体として安全であるとは言えない．つまり，暗号化されるべき情報を正しく暗号化しているか否かを評価することが重要である．

オンラインショッピングの例では，顧客の個人情報を暗号化すればよいが，データの価値（漏れた場合の影響の大きさ）に応じて暗号化する範囲にも注意を払わなければならないと考える．そこで，本稿では，どの程度情報を暗号化するべきかの指針策定の基礎として，暗号化範囲に着目した暗号化レベルを提案する．

2.2 暗号化レベル

2.1 節で述べたように，本稿では，暗号化の範囲を暗号化レベルとして提案する．例えば，クレジットカード会社から，カードの利用明細が送られてくることを考える．この時，請求額や，利用明細などの私的な情報は外部に漏れてほしくないが，カード会社から請求が来ているという情報は，外部に漏れても特に問題ないと考える．このようなデータに関しては，請求額や利用明細に関する部分を暗号化していれば問題がない．一方，軍事に関わる通信ではいつ，だれが通信を行ったかという情報が漏れるだけでも，国家の危機となる場合もありえる．このようなデータに関しては，より広い範囲での暗号化が要求される．

このように，通信データの価値に応じで求められる暗号化の範囲は異なる．ここでは，暗号化範囲に着目したレベル分けの一例として，下記に示す 5 段階のレベルを定義する．

Lv. 1 ：暗号化なし
Lv. 2 ：何の通信を行ったかを暗号化
Lv. 3 ：Lv. 2 に加え，どのように通信を行ったかを暗号化
Lv. 4 ：Lv. 3 に加え，だれが通信を行ったかを暗号化
Lv. 5 ：通信の事実自体を暗号化

表 1 に，暗号化レベルと，暗号化範囲を 5W1H の要素に分解し，どの要素を暗号化しているかの関係を示す．表中の「○」は該当する要素が暗号化されていることを表している．このように，通信データの価値（用途）に応じて求められる暗号化範囲の程度を，暗号化レベルとして表す．

表 1 暗号化レベルと暗号範囲

Lv.	用途	When	Who	Where	What	Why	How
1	掲示板への書込み						
2	請求料金，銀行残高照会の通信				○		
3	企業拠点間の通信				○	○	○
4	契約・M&A の交渉		○	○	○	○	○
5	軍事の通信	○	○	○	○	○	○

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

なお，暗号化 Lv.4 や暗号化 Lv. 5 の判定は通信データを監視するだけでは判定が困難であるため，本稿では，暗号化レベルの範囲を Lv. 1～Lv. 3 に限定した暗号化レベル判定方法について述べる．

3. 暗号化レベル判定手法

本章では，提案した暗号化レベルを判定方法する方法について述べる．また，暗号化通信に対して乱数検定を行った結果を報告し，暗号化レベル判定に適した乱数検定方法及び，検定閾値について検討する．

3.1 暗号化領域と暗号化レベル

ネットワークを介した通信では，IP が広く利用されている．IP とは情報の伝達を行うプロトコルであり，インターネットの基礎部分となる重要な役割を担っている．IP は，IP パケットと呼ばれる単位にデータを分割し，データ伝送を行う．IP では，その上位プロトコルである TCP や UDP などとともに用いられる．

ユーザが TCP/IP を利用してデータを送信する際には，ユーザが送信するデータに TCP ヘッダが付与され，下位プロトコルである IP に渡される．IP では，さらに IP ヘッダを付与し，ネットワークへと伝送する（図 1）．このように，データに次々とヘッダ情報を付与し，カプセル化することにより，プロトコル間の独立性を高め，効率的なデータ転送を実現している．

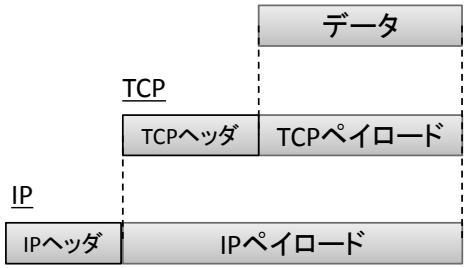

図 1　TCP/IP のデータ構造

この時，どのような通信を行ったかの通信内容については，TCP ペイロードに記載されており，また，どのように通信を行ったかについては，TCP ヘッダに記載されている．そのため，上記カプセル化のどの領域を暗号化しているかを特定することにより，上述した暗号化レベルの判定が可能となる．

図 2 に暗号化レベルと暗号化領域の関係を示す．IPsec のように，IP ペイロード部が暗号文であるものは暗号化 Lv. 3 に，HTTPS のように，TCP ペイロード部が暗号文であるものは暗号化 Lv. 2 に，HTTP のようにすべてのデータが平文であるものは暗号化 Lv. 1 に分類する．

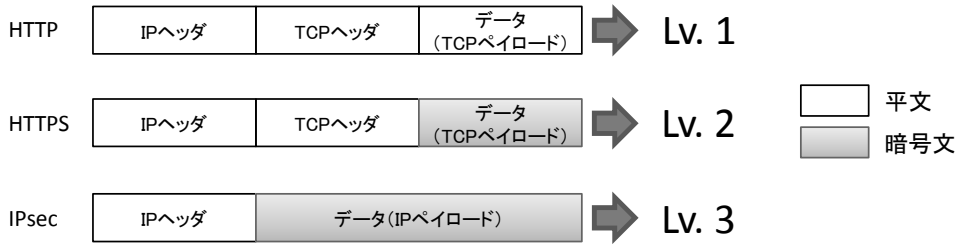

図 2　暗号化レベルと暗号化領域

3.2 暗号化レベル判定手法の設計

(1) 乱数検定

提案手法では，3.1 節で述べたように，パケットのどの領域が暗号化されているのかを特定することで，暗号化レベルを判定する．なお，データが暗号化されているか否かの判定方法には乱数検定手法を用いる．通信が暗号化されている場合には，当該データは何の特徴も存在しない乱数列のように見える．この特徴を利用し，データに乱数検定を適用することで，当該データが暗号化されているか否かの判定を行う．

乱数性を評価するための乱数検定として，次のような検定がある．

- NIST Special Publication 800-22[15]
- FIPS140-2[12]
- DIEHARD による検定法[16]

本稿では，NIST が推奨している乱数検定であり，乱数性の検定に広く利用されていることから，NIST Special Publication 800-22 を利用する．NIST Special Publication 800-22 には 15 種類の乱数検定手法が規定されており，検定ごとに有意確率（p-value）が得られる．p-value とは，検定で出力される統計量の正規分布もしくは，カイ 2 乗分布において，それよりも偏った統計量が発生する確率を表したものである．NIST Special Publication 800-22 では，p-value 値が 0.01 より小さい場合に，そのデータは良い乱数ではないと判断している．

(2) 暗号化レベル判定のための閾値検討

ここでは，暗号化レベル判定において，乱数性ありと判定する閾値を決定するために，トラフィックの調査を行う．調査にあたっては，ネットワーク上を流れる通信を

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

キャプチャしたテストデータと Web からダウンロードしたテストデータを用いる．以下に，トラフィック調査に利用したテストデータを示す．

表 2　テストデータの詳細

プロトコル	乱数性の有無
HTTPS	乱数性ありと判定されるべきトラフィック
IPsec	乱数性ありと判定されるべきトラフィック
HTTP	乱数性なしと判定されるべきトラフィック
ICMP	乱数性なしと判定されるべきトラフィック

テストデータに NIST Special Publication 800-22 で規定されている乱数検定を適用し，乱数と判定された乱数検定手法の割合を図 3 に示す．なお，乱数検定対象のデータサイズ（TCP ペイロード長，あるいは IP ペイロード長）が 0 のパケットを除き，最初の 10 パケットに乱数検定を適用した．

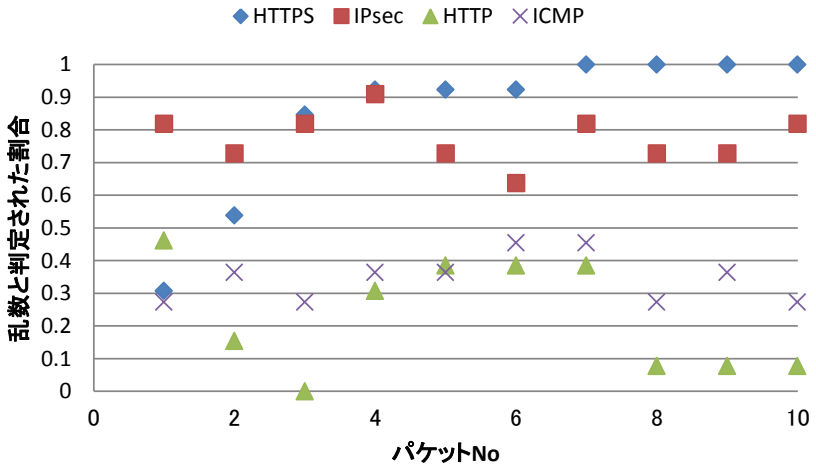

図 3　トラフィック調査結果

HTTPS トラフィックでは，最初の 2 パケットを除き，常に 8 割以上の乱数検定手法で乱数性ありと判定されたが，IPsec トラフィックでは，6〜9 割の間とブレが生じた．

一方，HTTP や ICMP トラフィックでは，いずれも乱数性ありと判定した乱数検定手法は 5 割以下となっている．なお，HTTPS トラフィックの最初の 2 パケットにおいて乱数と判定された割合が低かったのは，平文にて通信を行っている SSL ハンドシェイク部分を判定対象としてしまったからである．以上の結果より，通信開始直後の数パケットを除外して，乱数検定手法の 5 割以上で乱数性ありと判定されたパケットを，乱数性ありと判断することで，今回テストに利用したトラフィックでは，暗号化レベル判定の誤判定をなくすことができる．

(3) 暗号化レベル判定フロー

乱数検定手法を用いた暗号化レベル判定フローを図 4 に示す．

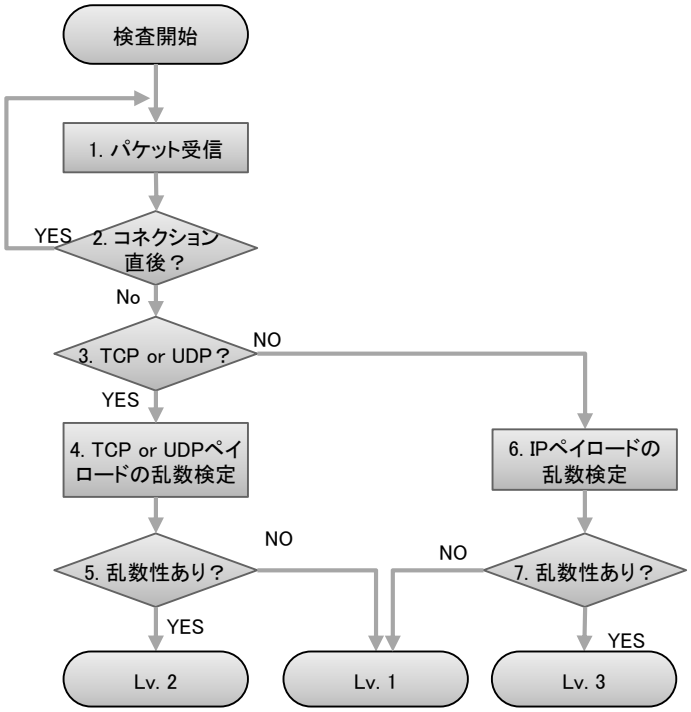

図 4　暗号化レベル判定フロー

暗号化レベル判定の具体的なフローは，次の通りである．

1.　暗号化レベル判定装置が検査対象パケットを受信する．具体的には，判定装置

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

をルータやスイッチのミラーポートに接続したり，判定装置をインラインに設置したりする．

2. コネクション確立直後（10 パケット以内）であれば破棄し，それ以外であれば処理 3 に進む．
3. TCP or UDP パケットであれば，処理 4 に進み，TCP or UDP パケットでなければ処理 6 に進む．
4. TCP or UDP ペイロード部に乱数検定を適用する．
5. TCP or UDP ペイロード部に乱数性があれば，暗号化 Lv. 2 と判定し，乱数性がなければ暗号化 Lv. 1 と判定する．なお，5 割以上の乱数検定手法で乱数性ありと判定されたパケットを乱数性ありと判断する．
6. IP ペイロード部に乱数検定を適用する．
7. IP ペイロード部に乱数性があれば，暗号化 Lv. 3 と判定し，乱数性がなければ，暗号化 Lv.1 と判定する．

4. 評価実験

本章では，暗号化レベル判定機能を実装した提案方式のプロトタイプを用いて実施した有効性の評価実験について述べる．

4.1 評価方法

(1) 評価の想定

暗号化 Lv. 2 の HTTPS 通信が行われている通信路に，暗号化 Lv. 1 の HTTP 通信が混入するケースや，暗号化 Lv. 3 の IPsec 通信が行われている通信路に，暗号化 Lv. 2 の SSH 通信が混入してくるケースを想定し，HTTPS トラフィック，HTTP トラフィック，IPsec トラフィック，SSH トラフィックの暗号化レベル判定精度の評価を行う．

(2) 評価データ

評価にあたって，HTTP トラフィック，HTTPS トラフィック，SSH トラフィック，IPsec トラフィックを用意し，それぞれ 120 パケット分判定を行った．以下に，評価に利用したテストデータを示す．

表 3　評価データの詳細

プロトコル	暗号化レベル
HTTP	暗号化 Lv. 1 と判定されるべきトラフィック
HTTPS	暗号化 Lv. 2 と判定されるべきトラフィック
SSH	暗号化 Lv. 2 と判定されるべきトラフィック
IPsec	暗号化 Lv. 3 と判定されるべきトラフィック

4.2 評価結果と考察

(1) 評価結果

表 4 に 120 パケットの暗号化レベル判定結果を，図 5～図 8 に各トラフィックの乱数と判定された割合を示す．なお，SSH トラフィックでは 100%の精度で，それ以外のトラフィックでも 99%以上の精度で暗号化レベルを正しく判定することができた．

また，HTTP トラフィック，HTTPS トラフィック，IPsec トラフィックの誤判定を行ったパケットについても，乱数と判定された割合は閾値の 0.5 に近い値となっており，提案手法が有効であることが分かる．

表 4　暗号化レベル判定結果

プロトコル	暗号化 Lv. 1	暗号化 Lv. 2	暗号化 Lv. 3
HTTP	119	1	0
HTTPS	1	119	0
SSH	0	120	0
IPsec	0	1	119

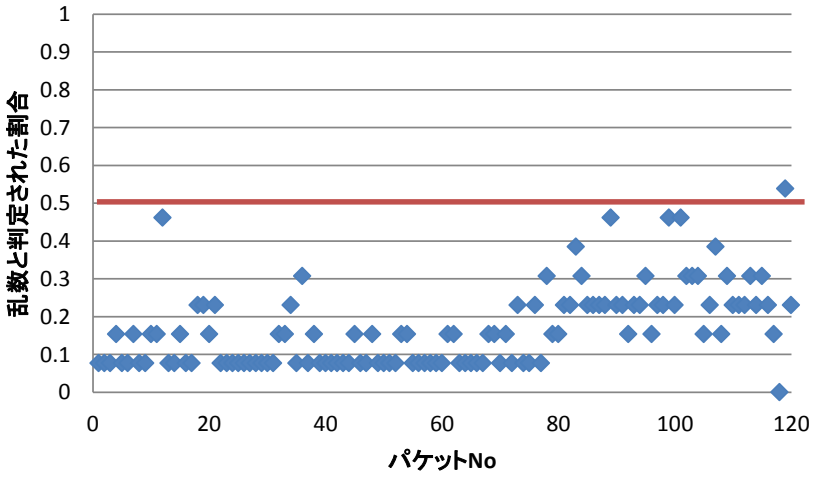

図 5　HTTP トラフィック

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

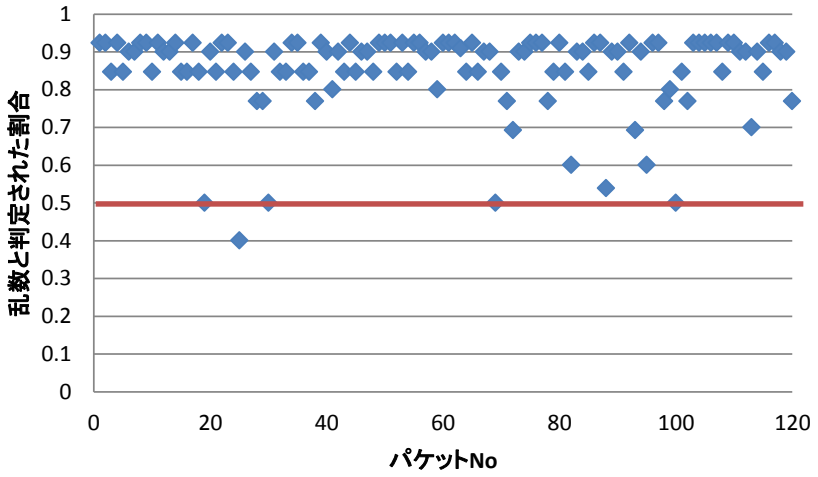

図 6　HTTPS トラフィック

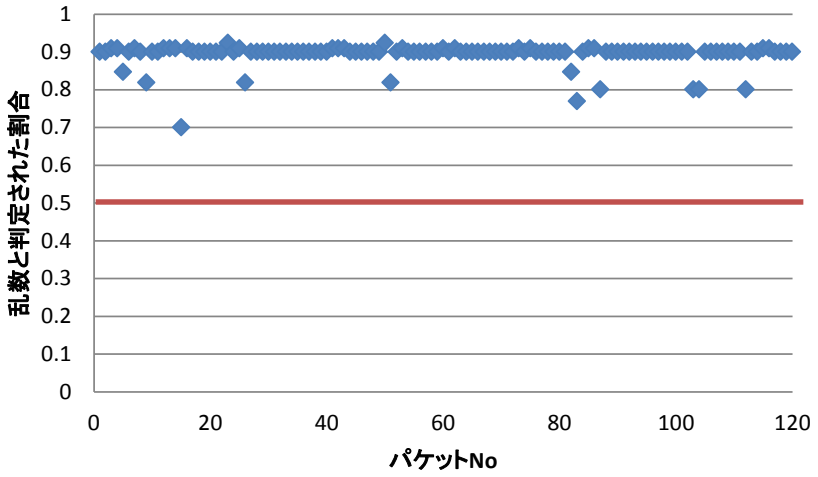

図 7　SSH トラフィック

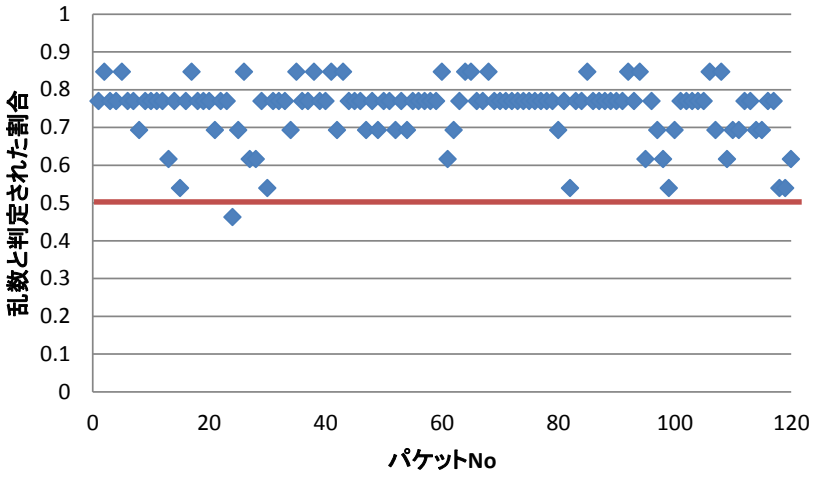

図 8　IPsec トラフィック

(2) 考察

(a) 判定率について

評価実験において，100%の判定精度は達成できなかった．しかし，構成変更を検出するために提案手法を利用するのであれば，例えば，10 パケット中 9 パケットで同じ暗号化レベルが算出されれば，その暗号化レベルを出力するといった多数決システムを導入するという運用でもよい．多数決システムを導入することにより，暗号化レベル算出までに時間を要してしまうが，その分正確になる．人が対応しなければならない構成変更等のアラートであれば，数秒単位の遅延は許容されるため，十分活用可能であると考える．

(b) 閾値について

評価実験に利用しなかった様々なアプリケーションの通信が含まれているため，適切な閾値は変化すると考えられる．今後は，様々なアプリケーションの通信に対して評価を行い，適切な閾値を検討する．

(c) 動画・画像データについて

例えば，暗号化 Lv. 1 と判定されるべき HTTP 通信においても，動画データや画像データ，音声データがやり取りされている場合には，乱数性があると判定される恐れがある．この点に関しても今後様々な通信を評価することで，適切な閾値を検討する．

Vol.2011-CSEC-54 No.11
Vol.2011-SPT-1 No.11
2011/7/12

5. おわりに

本稿では，暗号化レベルと，暗号化レベルの判定を行う方式を提案した．また，評価用データを，提案方式のプロトタイプを搭載した PC で判定処理を行うことで，提案方式の有効性を検証した．その結果，データの乱数性に着目することで，暗号化レベルを判定できることを示した．

今後は，クラウド環境への適用を視野に入れ，判定速度の評価や，判定制度の向上を目指す．また，クラウド環境上の所定の端末（あるいは所定のポート）で送受信されるパケットを常時監視し，暗号化レベルが低下した場合にクラウド提供者にアラートを挙げることで, 構成変更や設定ミスが起きた可能性を検出できることを確認する．

謝辞 本研究は総務省から受託した「クラウド対応型セキュリティ対策技術の研究開発」の成果の一部です．本研究を進めるにあたって有益な助言と協力を頂いた関係者各位に深く感謝いたします．

参考文献

1) IPA, “クラウドコンピューティング社会の基盤に関する研究会報告書”, http://www.ipa.go.jp/about/research/2009cloud/index.html
2) IPA, “IT コーディネータが見た中小企業等におけるクラウドサービス利用上の課題・導入実態調査報告書”, http://sec.ipa.go.jp/reports/20110331.html
3) ITPro, “「クラウドのセキュリティ不安」は食わず嫌い”, http://itpro.nikkeibp.co.jp/article/NEWS/20110128/356606/
4) 総務省, “ASP・SaaS における情報セキュリティ対策ガイドライン”, http://www.soumu.go.jp/menu_news/s-news/2008/pdf/080130_3_bt3.pdf
5) 甲斐賢，重本倫宏，鬼頭哲郎，武本敏，鍛忠司, “クラウドコンピューティング環境に適したセキュリティ状態定量化手法の提案”, JSSM, 2011 年 6 月
6) Charles V. Wright, Lucas Ballard, Fabian Monrose, and Gerald M. Masson, "Language identification of encrypted VoIP traffic: Alejandra y Roberto or Alice and Bob?", In Proceedings of 16th USENIX Security Symposium, pp. 1-12, 2007.
7) Charles V. Wright, Fabian Monrose, and Gerald M. Masson, "Using visual motifs to classify encrypted traffic", In Proceedings of the 3rd international workshop (VizSEC '06), pp. 41-50, 2006.
8) 重本倫宏，仲小路博史，寺田真敏, “初期パケットの乱数性に着目した P2P 通信検知方式”, 情報処理学会研究報告, Vol. 2010-CSEC-50, No. 38, pp. 1-6, 2010.
9) NIST, FIPS 197, “Advanced Encryption Standard”, http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
10) IPA, “電子政府推奨暗号の仕様書”, http://www.ipa.go.jp/security/enc/CRYPTREC/fy15/cryptrec20030425_spec01.html
11) IPA, “暗号モジュール試験及び認証制度”, http://www.ipa.go.jp/security/jcmvp/index.html
12) NIST, FIPS PUB 140-2, “Security requirements for cryptographic modules”, http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf
13) IPA, “安全な暗号鍵のライフサイクルマネージメントに関する調査”, http://www.ipa.go.jp/security/fy19/reports/Key_Management/index.html
14) IPA, “暗号をめぐる最近の話題に関するレポート”, http://www.ipa.go.jp/about/technicalwatch/20110511.html
15) NIST, “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”, http://csrc.nist.gov/groups/ST/toolkit/rng/index.html
16) G. Marsaglia, “DIEHARD”, http://www.stat.fsu.edu/pub/diehard/