コンピュータセキュリティ 19－3
（２００２．１２．２０）

動作監視に基づく Web サーバ防御システム

中江　政行*，山形　昌也**，矢野尾　一男*，小川　隆一*

概要

外部からの攻撃を防ぐために，ファイアウォールやコンテンツフィルタなどが広く普及した現在でもなお，Web コンテンツの不正書き換え事件やインターネットワーム感染が後を絶たない．新しい攻撃に対処するために，近年では，保護すべき内部ネットワークまたはサーバにおける異常に対して，適応的に対処しようとする動的防御システムが提案されるようになった．しかし，この方式は異常検知時に，どのサーバに対しても被害が生じないという保証ができなかった．本稿では，内部ネットワークを「監視領域」と「保護領域」に分割し，監視領域内で攻撃検知と動的防御を行うことで適応性を保ちつつ，被害を受けない保護領域を実現しようとする新しい動的防御コンセプト（PDD）を提案する．また，PDD に基づく Web サーバ向けの動的防御システムの試作と評価について報告する．

A Behavior-Based Intrusion Prevention System for Web-servers.

Masayuki NAKAE*, Masaya YAMAGATA**, Kazuo YANOO*, Ryuichi OGAWA*

Abstract

In order to prevent unknown attacks from external networks, the concept of dynamic defense has been proposed. The system employs intrusion detection and filtering technologies to take adaptive measures against anomalies detected on the network. However it cannot guarantee that internal servers are securely protected against attacks. To solve this problem we propose a new dynamic defense model called “prudent dynamic defense (PDD),” that separates the internal network into two parts: monitoring area and protected area. This report describes the outline of the model and its experimental implementation.

1. はじめに

今日，企業・官庁などあらゆる組織で，情報提供のために Web サーバをインターネット向けに運用するようになった．しかし，近年，Web サーバのバグをついた，インターネットワームの感染やコンテンツの改ざんなどの攻撃が増加しつづけている．

こうした外部からの攻撃から Web サーバを保護するために，(a) ファイアウォール，(b) コンテンツフィルタ，(c) 動的防御システムなどを，内部ネットワークの外縁に配置することが一般に行われるようになった．

(a) ファイアウォールは，事前に定められたアクセス制御ルール（ACL）によって，外部からのアクセスの通過可否を決定する．このとき，IP アドレスやプロトコルなどに関する検査しか行われないので，先に挙げたような攻撃を防御できない．

(b) コンテンツフィルタは，外部からのアクセスについて，所定の攻撃パターン（シグネチャ）に合致する場合に，それを遮断する．ファイアウォールと違い，ネットワークデータを検査するので，アプリケーションのバグを突くような攻撃の防御が可能となる．ただし，防御可能な攻撃は，シグネチャとして蓄積された既知のものに限られる．したがって，アプリケーションに新たなバグが発見されるなどして，新たなパターンを持つ攻撃が現れた場合，それを防御できない．

以上 2 種の防御システムは，通過・遮断のルールが一定であることから，本稿では静的防御システムと総称する．こうしたシステムは，攻撃を検知できれば，確実にそれを防御できるという特長をもつが，前述のように，攻撃手法の変化に対して適応性に乏しいという問題をもつ．

* NEC インターネットシステム研究所 Internet Systems Research Laboratories, NEC Corp.
** NEC インターネット基盤開発本部 Internet Solution Platform Development Division, NEC Corp.

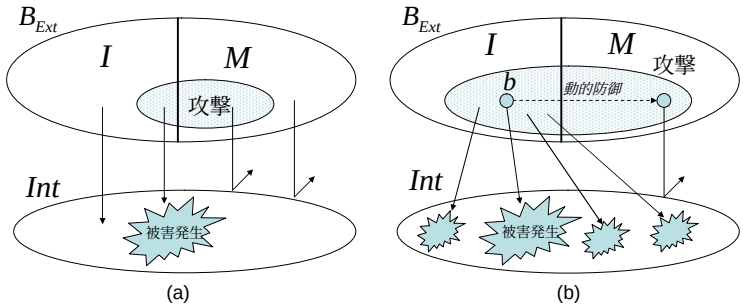

図 1　防御モデル（a：静的，b：動的）

（c）動的防御システムは，OPSEC の SAMP[6]に代表されるように，侵入検知システム（IDS）とファイアウォールとの連携により，内部ネットワーク内で攻撃が発見された際に，その攻撃元を内部ネットワークから排除しようとするものである．静的防御システムがネットワークデータのみを参照するのに対して，動的防御システムは，サーバの挙動など多様な情報源を活用して攻撃検知を行うことができるという特長をもつ．特に，サーバの挙動から攻撃検知するビヘイビアベース IDS[1][2]や統計的手法に基づくアノーマリベース IDS[3]を用いれば，より高い防御性能と適応性を期待できる．

しかし，JNSA のダイナミックディフェンス WG[5]は，その可能性を検討する中で，(1) 誤検知・検知漏れによるシステムの誤動作，(2) 検知した攻撃そのものは防御できないこと，(3) IP 偽装や DoS などに対して有効な防御が困難なこと，などを問題点として挙げている．(1) については，IDS の精度向上が必要である．(3) については，IP のもつ根本的な脆弱性であり，システムレベルで解決できる問題ではない．一方，(2) については，内部ネットワーク内の異常を防御のトリガとすることに起因し，動的防御システムが解決しうる問題である．

本稿では，特に (2) に対して，高い適応性を備えながら，静的防御システムのように確実な防御を実現できる動的防御システムについて考察する．

2. 防御モデル

前節で述べた静的防御システムと動的防御システムの問題を明確にするために，アクセス制御の観点から，それぞれの防御プロセスをモデル化して考察する（図 1）．

(1) 定義・記法：まず，ネットワーク全体に含まれるエンティティ（ホストまたはサービス）の集合を N とし，その分割 Int（内部ネットワーク）および Ext（外部ネットワーク）が与えられているものとする．また，$e1 \in Ext$ を外部エンティティとよび，$e2 \in Int$ を内部エンティティとよぶ．そして，$e1$ から $e2$ へのアクセスを 3 つ組 $< e1, d, e2 >$ で表す．ここで，d は $e1$ から送信されたネットワークデータとする．以下，単にアクセス $< e1, d, e2 >$ と表した場合，$e1$，$e2$ をそれぞれ外部エンティティ，内部エンティティであるとする．

外部エンティティから内部エンティティに送信されるネットワークデータの集合を D とする．このうち，内部エンティティのいずれかに対して「攻撃」となるものの集合を $A(\subset D)$ とする．

さらに，アクセス $< e1, d, e2 >$ について，それが防御システムによって遮断される場合を $rej < e1, d, e2 >$ と書き，逆に通過が許可される場合を $acc < e1, d, e2 >$ と書く．特に，あるアクセス $< e1, d \in A, e2 >$ について，$rej < e1, d, e2 >$ と判定される場合，これを「防御」と呼び，特にその繰り返しにおいて，いつでも $rej < e1, d, e2 >$ と判定される場合を，「確実な防御」と呼ぶ．

(2) 静的防御のモデル：静的防御システムは，アクセス $< e1, d, e2 >$ について，所定のルール R によって，$rej < e1, d, e2 >$ または $acc < e1, d, e2 >$ を判定するものとみなすことができる．ファイアウォールにおける R は，内部エンティティの区別を省略すれば，$e1$ または d の属性（IP アドレスやプロトコルなど）に基づくものとみなすことができる．また，コンテンツフィルタのルール，d に基づくものとみなすことができる．したがって，静的防御における R は，外部エンティティ $e1$ と d との組 $< e1, d >$ から，そのアクセスの通過・遮断を決めるものとみなしてよい．以下，2 組 $< e1, d >$ を外部エンティティ $e1$ の「振舞い」と呼び，その集合を B_{Ext} と表す．

静的防御システムは，予め B_{Ext} を以下のような「良い振舞い」「悪い振舞い」に分類しておき，それを R に記述しておく仕組みであるといえる．すなわち，R によって，良い振舞い I は（少なくとも 1 つの）内部エンティティへの通過を許可され，悪い振舞い M は一切遮断される（図 1(a)）：

$$I = \{b \in B_{Ext} \mid \exists e \in Int; acc < b, e >\}$$

$$M = \{b \in B_{Ext} \mid \forall e \in Int; rej < b, e >\}$$

もし，任意の $< e, a \in A >$ が M に属するように R が記述されていれば，その静的防御システムは任意の攻撃を確実に防御できる．しかし，現実には，それまで知られていない脆弱性を突くような振舞い $< e, d >$ は，I に分類される．その後，脆弱性が発見されてから，初めて攻撃であると認識され，$< e, d > \in M$ と分類される．

静的防御システムが，こうした新たな攻撃に対応するには I,M の再定義（=R の更新）が必要となるが，その作業にセキュリティベンダやシステム運用者の人手を要する．このことが静的防御システムの適応性を損なう要因である．

(3) **動的防御のモデル：**動的防御システムは，静的防御システムの基本動作に加えて，通過を許されたアクセス $< e1, d, e2 >$ について，内部ネットワーク上の IDS によって侵入が検知された際に，その振舞い $< e1, d >$ を M に属するものと判定する．すなわち，B_{Ext} の分割 {I,M} において，ある時点 t での振舞い b∈I について，次の時点 t+1 で b∈M としようとするものである（図 1(b)）：

$$I(t+1) = I(t) \setminus \{b \in I(t)\}$$

$$M(t+1) = M(t) \cup \{b\}$$

時刻 t+1 以降，振舞い b は防御されるので，適応性があるといえる．しかし，前節の問題点（2）の通り，b は少なくとも一度通過を許可されているので，ある内部エンティティ $e2$ はその被害を受ける可能性がある．すなわち，検知可能である攻撃に対して，確実な防御を保証できない．

3. PDD コンセプト

従来の静的・動的防御システムで達成できない，防御の適応性と確実性を両立させたい．そこで，「用心深い動的防御（Prudent Dynamic Defense; PDD）」というコンセプトを提案する（図 2）．

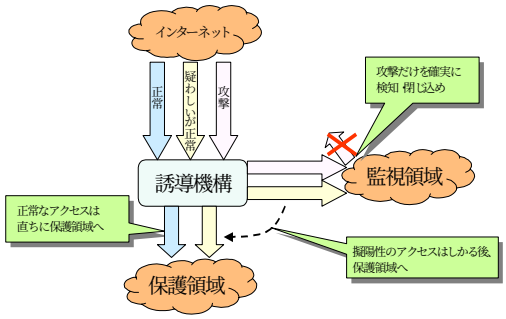

図 2　PDD コンセプト

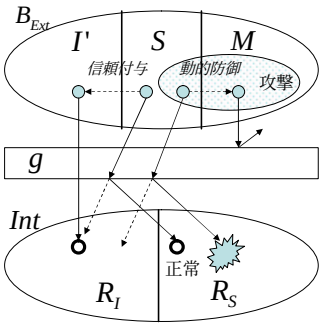

図 3　PDD の防御モデル

(1) **概念：**PDD は，従来の静的・動的防御システムと異なり，内部ネットワークを，信頼できるアクセスを処理するための「保護領域」と，不審アクセスを処理するための「監視領域」とに分割することを特徴とする．

誘導機構によって，保護領域へ向かうアクセスは，最初すべて「不審」であるものとし，監視領域へ強制的に誘導する．監視領域には，保護領域と同じサービスを提供するサーバを配置し，不審アクセスに対する動作状況を監視して，攻撃の有無を検査する．

検査結果は，誘導機構にフィードバックされ，異常動作が見られない場合に限り，そのアクセスを信頼して，保護領域へのアクセス権を与える．また，攻撃があった場合には，以降，そのアクセスを一切遮断する．

こうすることで，新しい攻撃であっても，必ず監視領域で検査され，異常が認められれば防御される（適応性）．さらに検知可能な攻撃については，一度たりとも保護領域に到達しないことを保証できる（確実性）．これらの性質によって，被害を受けない保護領域が実現される．

(2) PDD **モデル：**以下，前節のモデルを用いて PDD をあらわす（図 3）．まず，内部ネットワーク Int が，保護領域 R_I と，監視領域 R_S に分割されている．また，外部エンティティの振舞い集合 B_{Ext} の分割を以下のように再定義する．

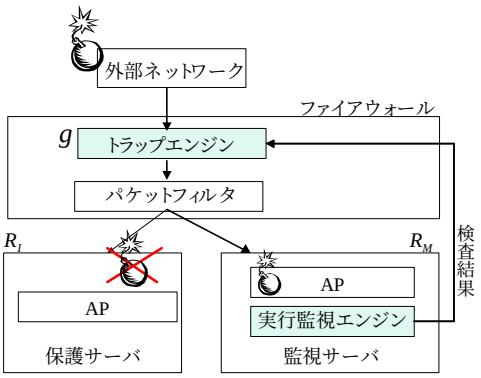

図 4　システムアーキテクチャ

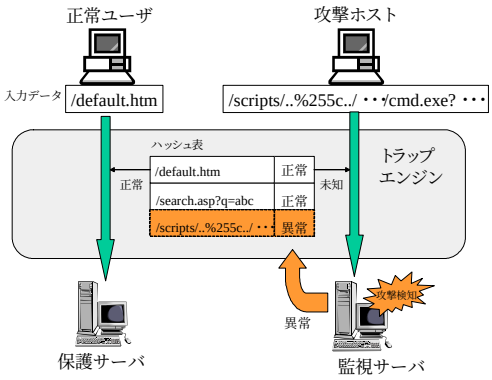

図 5　トラップ方式

$$I'=\{b\in B_{Ext}\mid \exists e\in R_I; acc<b,e>\}$$
$$S=\{b\in B_{Ext}\mid \forall e\in R_I; rej<b,e>\}$$
$$M=\{b\in B_{Ext}\mid \forall e\in Int; rej<b,e>\}$$

これは，外部エンティティの振舞いを，「信頼（I'）「不審（S）」「攻撃（M）」の 3 種類に分類することを表す．

また，R_I，R_S との分離は，以下のアクセス制御ルールによって表される：

$$\forall e\in R_S \forall e'\in R_I; rej(e,d,e') \cdots(3.1)$$

(3) 誘導：誘導機構は保護領域へ向かうアクセスを，その振舞いの分類に応じて，通過許可，誘導および遮断を行うものであり，以下のような誘導関数 g で表される．

$$g(<b,e\in R_I>)=\begin{cases}<b,e>(b\in I')\\<b,e'\in R_M>(b\in S)\\<b,f>(b\in M)\end{cases}$$

ここで，$<b,f>$は遮断を示す．

(4) 信頼付与：最初に，すべての振舞いを不審とみなす：

$$I'(0)=f,\quad S(0)=B_{Ext} \cdots(3.2)$$

そして，任意の時点 t で，$S(t)$ に属する不審な振舞い b について，R_S で正常動作のみが観測されたとき，b は「信頼」される：

$$S(t+1)=S(t)\setminus\{b\in S(t)\}$$
$$I'(t+1)=I'(t)\cup\{b\} \cdots(3.3)$$

(5) 防御の適応性：信頼付与のルール(3.1), (3.2)によって，任意の時点 t で新しい攻撃があった場合，その振舞い $<e1,d\in A>$ は必ず $S(t)$ に属する．そして，R_S で異常動作が観測されたとき，その振舞いは「攻撃」に分類され，それ以降，防御される：

$$S(t+1)=S(t)\setminus\{b\in S(t)\}$$
$$M(t+1)=S(t)\cup\{b\}$$

このように，PDD は適応性をもつ．

(5) 防御の確実性：R_S に到達した攻撃は，分離のルール(3.1)により，R_S を経て R_I に到達することはない．また，R_S で検知されれば，それ以降，防御される．したがって，PDD は検知可能な R_I への攻撃全てを，確実に防御できる．

4. Web サーバ防御システム

4.1. アーキテクチャ

本節では，PDD コンセプトに基づいた，Web サーバ向けの攻撃防御システム（以下 PDD プロトタイプ）について述べる．

本システムでは，保護領域に保護すべき Web サーバ（保護サーバ）を配置し，監視領域に，ミラーサーバ（監視サーバ）を配置する．また，誘導機構に相当するトラップエンジンを，ファイアウォールに組み込む（図 4）．

監視サーバでは，保護サーバと同じサービスを提供するアプリケーションプログラムと共に，実行監視エンジンを動作させる．実行監視エンジンは，4.3節で述べる実行監視方式に基づいて，アプリケーションプログラムの動作を監視しながら，その正常・異常を判定する．また，判定結果については，随時トラップエンジンにフィードバックする．

4.2. トラップ方式

トラップエンジンは，その内部に，過去に外部ネットワークから受信したデータと，それに対する監視サーバでの検査結果に関するハッシュ表を備える．ハッシュ表は，前節のモデルにおける I', S, M に相当する．

まず，ハッシュ表の初期状態を空とする．

その後，ファイアウォールが，新たなアクセスを受ける度に，ハッシュ表を検索し，以下のようなアルゴリズムで転送先の選択または遮断を行う（図 5）．

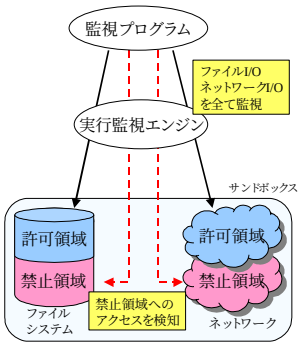

図 6　実行監視方式

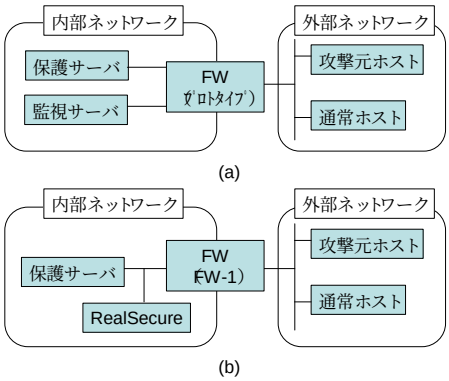

図 7　実験システムの構成

1. ハッシュ表にエントリされており，かつ過去の検査結果が「正常」であれば，保護サーバへ転送する．
2. ハッシュ表にエントリされていないものは監視サーバに転送する．
3. エントリされていてかつ過去の検査結果が「異常」であるものは（パケットフィルタに働きかけて）遮断する．

新たな振舞いがハッシュ表にエントリされるのは，監視サーバから検査結果を受けた時点である．したがって，エントリされている振舞いは，検査済みであることを保証できる．

4.3. 実行監視方式

実行監視エンジンは，監視対象となるアプリケーションの動作をシステムコールレベルで監視しながら，アプリケーションの正常動作または異常動作に関する知識を用いて，攻撃を検知する．こうした方式は一般に「ビヘイビアベース IDS」と呼ばれる[4]．

今回，コンテンツ改ざんやインターネットワームなどの「不正書き込み」を捉えるために，Web サーバによるファイル I/O またはネットワーク I/O について，所定のアクセス範囲に収まっているかどうかを監視する方法を用いた．

すなわち，実行監視エンジンに，予めファイル I/O・ネットワーク I/O に関する許可・禁止ルールの集合（サンドボックス）を設定しておき，Web サーバが行う個々の I/O を監視しながら，サンドボックス内の各ルールと照合する．許可された I/O のみから成る動作を正常とし，少なくとも 1 つ禁止された I/O が観測されたときに，それを異常とみなす（図 6）．

同様の方法は Chari ら[1]の研究にも見られ，比較的負荷の低いことが利点として挙げられている．より高度な方法として，事前のプログラム解析を利用する手法が提案されている [2]．こうした方式は，I/O だけでなく一般のシステムコールについて，発行順序の異常を検出できることから，より広範な攻撃を検知できるとされているが，特に空間計算量の点で実用性に乏しい．今後，リアルタイム性やメモリ効率と，検知性能とを両立した方式が求められる．

5. 実装と評価

5.1. 実装

ファイアウォールは，Linux の ipchains をベースにした．トラップエンジンは透過型プロキシとして実装し，保護サーバへ向かうアクセスをトラップエンジンへリダイレクトするようにした．また，監視サーバから，異常動作が通知された際，攻撃元ホストを遮断するために，ipchains の ACL を変更するようにした．

監視サーバは，Windows のファイルシステムおよびネットワークドライバを新たに作成し，Web サーバの I/O を監視できるようにした．実行監視エンジンは，独立したアプリケーションとして実装し，上記ドライバから定期的にイベントを取得しながら，所定のサンドボックス定義との照合を行い，その結果をトラップエンジンに通知するようにした．

5.2. 評価方法

PDD プロトタイプの防御性能を，定量的に評価するため，実際に Microsoft 社の Web サーバ「Internet Information System バージョン 5」（以下 IIS5）の Extended Unicode Vulnerability（以下 EUV）[7]を用いて，不正なファイル書き出しを伴う攻撃を行い，防御に成功・失敗したアクセス数を測定する実験を行った（図 7(a)）．

本実験に際して，既存の動的防御システムの典型例として，ISS 社の RealSecure Network Sensor 6.5 と CheckPoint 社の FireWall-1 を組み合わせた動的防御システム（以下 RS+FW-1）を挙げ，ベンチマークとした（図 7(b)）．

公平な比較のため，攻撃を行うにあたっては，所定のスクリプトにしたがって，一定の速度で攻撃用トラフィックを生成するツール（トラフィックジェネレータ）を開発し，利用した．また，本ツールを動作させるホストと，攻撃対象である正規サーバは同じものを使用した．

本実験では，Nimda の感染手段を参考に，以下のような 2 種類のテストケースを策定し，スクリプトを作成した．

- ケース１：トップページ改竄
  （1）EUV を用いて，cmd.exe を CGI 領域にコピーし，バックドアを準備．
  （2）バックドアを用いて，10 バイト程度の不正な文字列でトップページを置換．
- ケース２：脆弱な CGI を介したワーム感染
  10 数 K バイトの不正プログラムを 6 分割したものを，ケース１のバックドアを用いて，順次不正に書き出す．

実験結果の解析については，保護サーバのログファイルから，到達した攻撃アクセスの数（＝防御に失敗した回数）を，試行ごとに計数した．この値から，次式によって得られる防御率 r を，各試行におけるスコアと定義した．

$$r = (1 - Ng / N) \times 100$$

ただし，N はアクセス総数，Ng は保護サーバに到達した攻撃アクセス数を示す．保護サーバに到達する攻撃が少なければ少ないほど，r は 100(%)に近づいていくので，これを防御性能の指標とした．

5.3. 評価結果

ケース１，ケース２の実験結果を表 1に示す．これは，スクリプトを 100 回繰り返したものを 1 回の試行とし，5 回の試行を行って得たスコアの平均である．

どちらのケースにおいても，PDD プロトタイプが 100%に近い防御率を示した．特に，ケース１で，RS+FW-1 の防御率が最良でも 99.5%であったのに対し，PDD プロトタイプは常に 100%であった．したがって，本防御方式が，従来方式に比べて，より高い確実性をもつことが確かめられた．

表 1　実験結果

	ケース１	ケース２
PDD プロトタイプ	100	99.7
RS+FW-1	28.3	0

また，ケース２で，RS+FW-1 がバックドア経由の攻撃をまったく検知できなかったが，PDD プロトタイプは 100%検知できた．一般に RealSecure のようなシグネチャベース IDS では，このような変種ワームの検知が難しい．これに対して，ビヘイビアベース IDS を用いることにより，高い適応性を実現できることが確かめられた．

6. おわりに

従来の静的防御システムや動的防御システムにない，「被害を受けない保護領域の実現」という特徴をもつ PDD コンセプトを提案した．

そして，PDD コンセプトに基づいた Web サーバ向け防御システムのアーキテクチャ，トラップ方式および実行監視方式を示し，その実装と評価実験について述べた．評価の結果，不正ファイル書き出しを伴う典型的な攻撃にについて，本防御システムが，従来方式に比べて，より高い確実性と適応性を備えていることを確認した．

しかし，各不審アクセスにつき１回の検査しか行わないトラップ方式では，対応できる攻撃は限られている．今後，より広範な攻撃に対応できるよう検討を重ねていきたい．

参考文献

[1]S.N. Chari, P. Cheng, "BlueBoX: A Policy–driven, Host–Based Intrusion Detection system", NDSS 2002.

[2]Wagner, D., and Dean, D., “Intrusion Detection via Static Analysis”, Proc. IEEE Security & Privacy, pp. 156-168, 2001.

[3]竹内,山西,”データマイニングにおける統計的外れ値検出”,応用数理,vol. 11, no.2, pp. 71-75, .

[4]SANS, IDS FAQ(ver.1.60), http://www.sans.org/newlook/resources/IDFAQ/ID_FAQ.htm

[5] JNSA ダイナミックディフェンス WG ，http://www.jnsa.org/active/houkoku/dd.doc.pdf.

[6]OPSEC Technology Overview, http://www.sofaware.com/pdf/tech_opsec.pdf.

[7]IIS Extended Unicode Vulnerability, http://www.sans.org/newlook/digests/unicode.htm.