WWW サーバログ解析から見たセキュリティ問題

名古屋大学大型計算機センター
長谷川 明生

概要

昨今、ネットワーク接続されたホストに対しての、セキュリティ・ホールのスキャンやセキュリティ・ホールをついた攻撃が増加の一途をたどっている。しかしながら、クラッキングの様相は、把握されていない。

このようなクラックのなかで、古典的なもののひとつに CGI サンプルのphfスクリプトをついた攻撃がある。今回、ある WWW サーバの 2 年分のアクセス・ログおよびエラー・ログをセキュリティ問題の観点から解析した。その結果、CGI スクリプトへのアクセスエラーの監視によりクラッキングの様相が把握できることが明らかとなった。

Analysis of WWW server log from the host security

Akiumi Hasegawa
Nagoya University Computation Center

Abstract

Current popularization of the Internet brings wide spread crackling activity to the Internet society. However, activities of cracking or so called security hole scans have not been analyzed quantitatively or qualitatively.

One of famous but obsolete security hole known in WWW server is phf CGI script. Almost all security hole scanners attack this hole first. So because the author analyze the 2 years of logs of our WWW server to find out the crackers activity. And this result shows that the monitoring of the access errors of phf script recorded in WWW server log is useful to watch crackers activity.

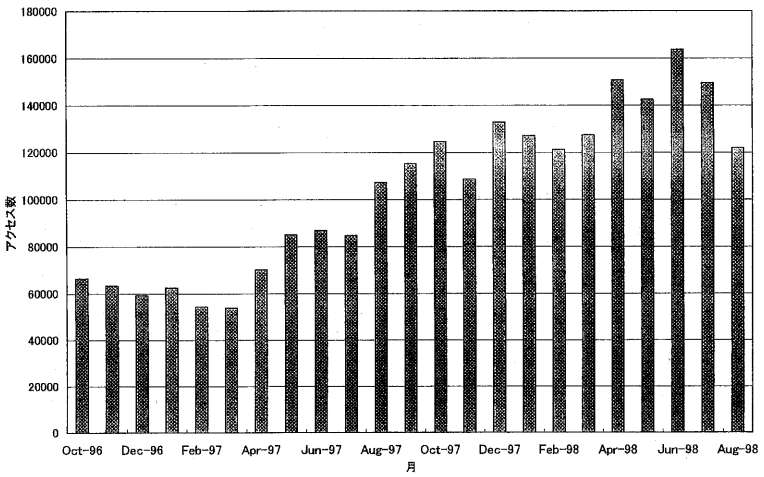
1. はじめに

名古屋大学では WWW 委員会の管理の下で専用の WWW サーバを用意し、WWW による広報を行なっている。この WWW サーバの各種ログは、運用当初から保存されている。今回、このログから WWW サーバへのアクセス動向を調査した。調査に使用したログの採取期間は、1996 年 10 月～1998 年 8 月で、一部 1998 年 9 月のログも含んでいる。資料の解析には、analog[1]および独自作成した perl スクリプトおよび Excel を使用した。

2. アクセス動向

1996 年 10 月～1998 年 8 月の間の月別アクセス数の推移を図1に示す。

図1 月別アクセス動向



1996 年 10 月の暫定運用開始以後アクセス数は漸増している。最高アクセス件数は、1998 年 7 月の 163,549 件であった。

表1に転送量から見た上位10ドメインを示す。

順位	ページ要求数	転送量(%)	ドメイン名
1	2,845,013	50.23	.jp
2	3,737,441	43.00	IP アドレスのみ
3	63,132	1.51	.com
4	50,321	1.11	.net
5	50,388	1.09	.edu
6	12,317	0.26	.de
7	11,321	0.24	.ca
8	10,873	0.24	.kr
9	27,350	0.24	ドメイン名なし
10	10,928	0.22	.uk

表 1 転送量上位10ドメイン

表 1 から、国内からのアクセスが大半を占めることが推定できる。

3. アクセスエラー動向

アクセス・エラー件数を図2に示す。ただし、アクセス・エラーのうちタイムアウトやプロトコ

ル・エラーについては除外した。除外したエラー数は約 3 万件でエラー総数の 3 分の 1 にあたる。

図2 アクセス・エラー件数

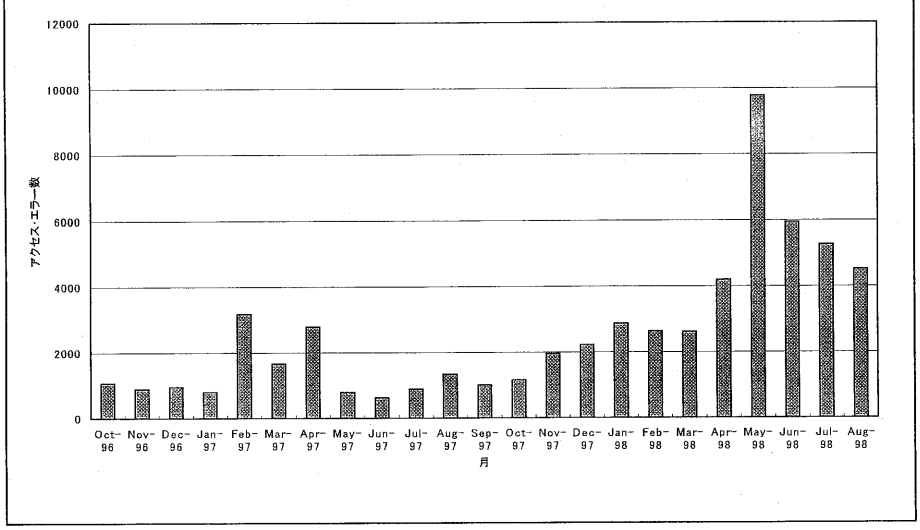
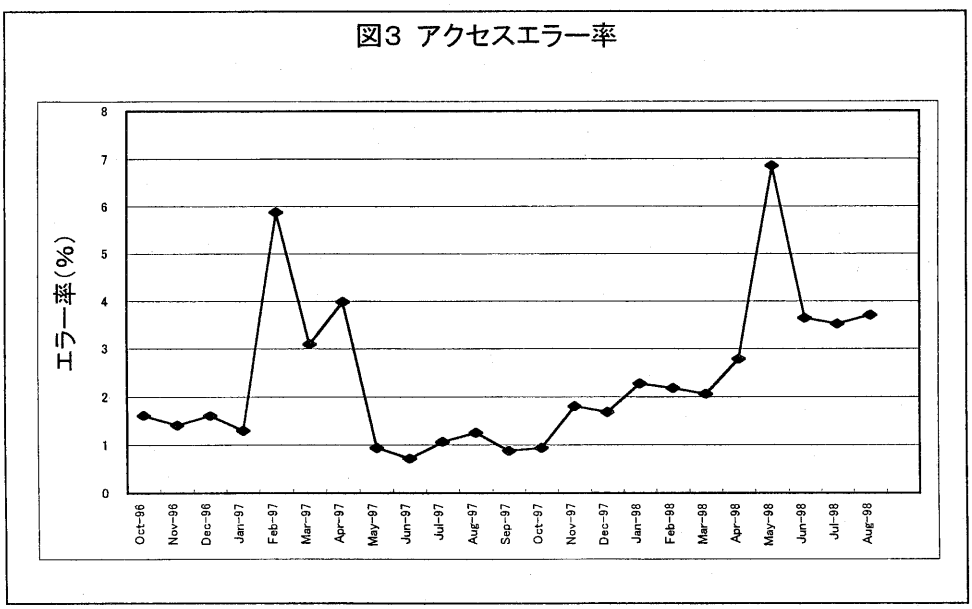


図3 アクセスエラー率

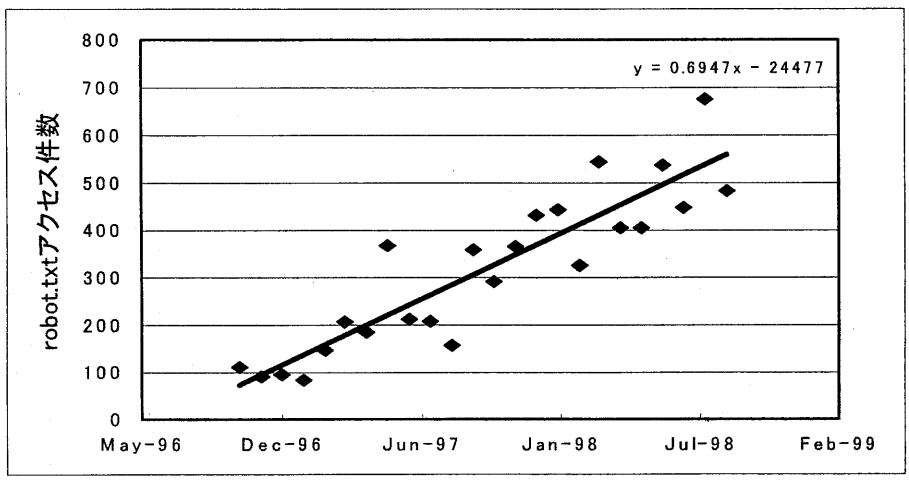


アクセスエラー率を図3に示す。1998 年 5 月のアクセス・エラー率の増加は、直前のページ構成変更に起因すると考えられるが、1997 年 2 月から 4 月にかけてのエラー率増加

の原因は不明である。

本学では、ロボットによるアクセスの制限を行っていない。したがって robot.txt の要求はすべてエラーとなる。その件数を図 6 に示す。

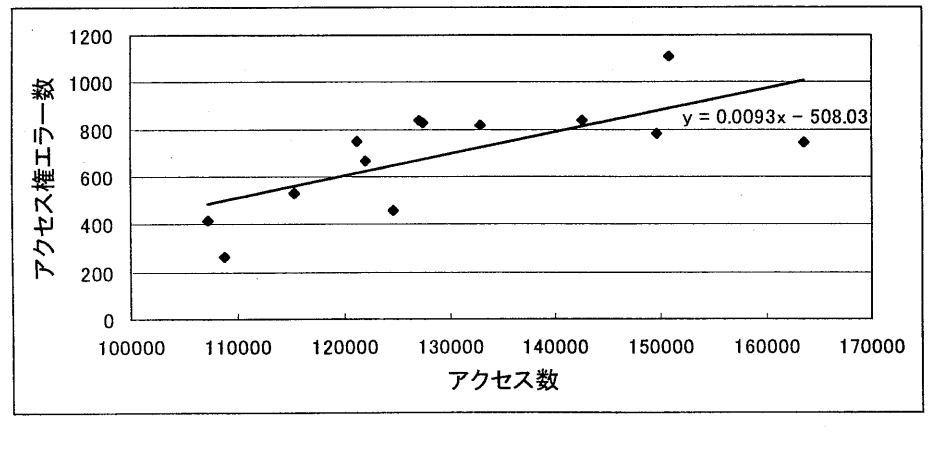
図4 ロボット・アクセス件数の推移



ロボットによるアクセス数は、図4のように月ごとに増加しており一次式でフィットできる。

つぎに、アクセス保護されている領域への外部からのアクセスの試みについて検討する。図5にアクセス権侵害件数とアクセス数の関係を示す。

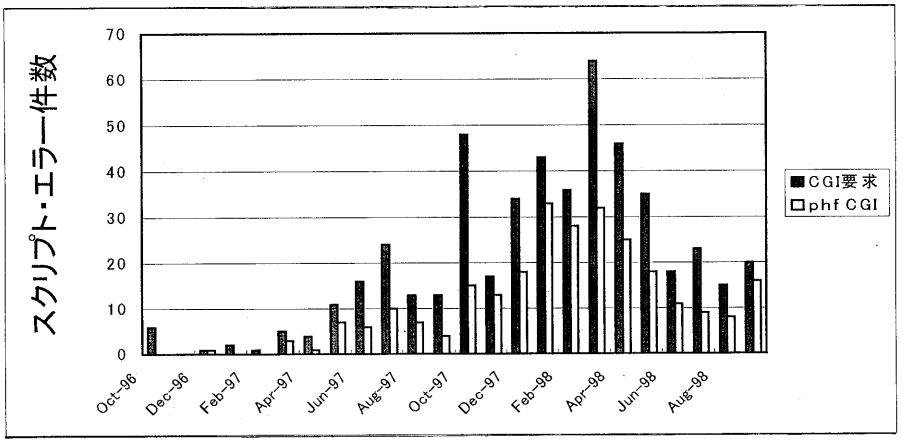
図5 アクセス権エラー数とアクセス数



アクセス権のないページへのアクセスの試みは、アクセス件数に対して常に一定の割合で存在していることがわかる。

古い WWW サーバには、不適切な CGI スクリプト例が用意されていて、セキュリティ・ホールとして知られていた。特に、phf スクリプトは、その典型である。本サーバには、これらのスクリプトはインストールされていないので、アクセスを行うと自動的にエラーとなる。CGI のエラーおよび phf エラーの月別件数を図6に示す。

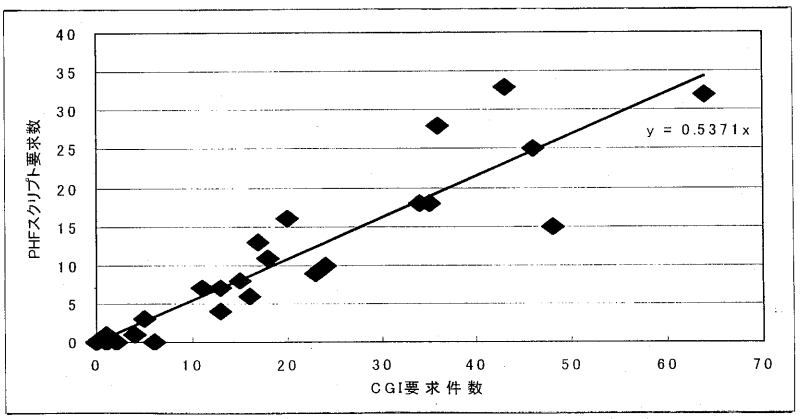
図6 CGI スクリプトエラー件数



図から、CGI スクリプトに関するアクセス・エラーの約半分が phf スクリプトに対してのものであることがわかる。

CGI スクリプトと phf スクリプトのエラーの関係を図7に示す。ここでは、phf スクリプトについてのみ調べたが、phf について CGI カウンターへのアクセスが多いことがわかっている。

図7 CGI アクセス件数と phf アクセス件数



4. 各種統計量間の関係と考察

ログに記録されている各種の値間の相関関係を表2に示す。

表2. 統計値間の相関

	年月	CGI	phf	アクセス権	ロボット	エラー数	アクセス数
年月	1						
CGI要求	0.669	1					
phf CGI	0.675	0.922	1				
アクセス権	0.884	0.771	0.791	1			
ロボット	0.904	0.600	0.616	0.780	1		
エラー数	0.786	0.467	0.427	0.738	0.645	1	
アクセス数	0.929	0.721	0.693	0.924	0.813	0.793	1

表から、時間の経過に依存するものとして、全アクセス数およびロボットによるアクセス件数が考えられる。また、表からアクセス数とアクセス権侵害数の強い関係が読み取れるが、これは、保護されたページに対して、出来心でアクセスする利用者が常に一定の割合で存在していることがわかる。

問題は、CGI スクリプトに対するアクセスと悪意を持った phf スクリプトに対するアクセスの強い依存関係である。全アクセス件数と CGI(および phf)関係は、アクセス保護されたページに対するアクセスと全アクセス数の関係に比べると薄弱で、これは phf 等のスクリプトへのアクセスがサーバへの侵入の意図を持ってなされていることを強く示唆している。

5. おわりに

WWW サーバの 2 年分のログを解析することにより、以下のようなことが明らかになった。

- WWW のアクセスはサービス開始以来徐々に増加していること。
- ロボットによるアクセスが増加傾向にあること。
- 保護されたページへのアクセスは、アクセス件数と比例していること。
  保護されたページを見ようと試みる利用者の割合は常に一定であること。
- CGI を介したクラッキングの試みが、相変わらずなされており、phf スクリプトや CGI カウンタを狙った攻撃がなくなっていないこと。

今回の解析によって、WWW ログの解析により、セキュリティ・ホールをスキャンしようとしているクラッカーの動向がある程度把握できることが明らかにできた。

6. 参考文献

1) S. Turner, http://www.statslab.cam.ac.uk/~sret1/analog/