Vol.2014-CSEC-65 No.17
Vol.2014-IOT-25 No.17
2014/5/23

筑波大学におけるハニーポットを用いた不適切なSSHアクセスの収集とその解析

佐藤 聡[1,a)]　小川 智也[2]　新城 靖[3]　吉田健一[4]

概要：筑波大学に割り当てられている IP アドレスの中で運用していないセグメント宛の通信は運用上破棄していた．この破棄されたパケットのうち TCP/22 番ポート宛のパケットをハニーポットにて処理することにより，筑波大学の IP アドレス内に設置されている ssh サーバにどのような攻撃があるかの解析を行ったのでその結果を報告する．

キーワード：ハニーポット，ネットワーク解析，大学ネットワーク運用

Analysis of SSH access to IP addresses which not assigned in University of Tsukuba using a honeypod

Abstract: Packets to IP addresses which are not used in University of Tsukuba was dropped at campus-central routers for normal operation. Among these dropped packets, the packets to TCP/22 port are processed with a honeypot. And we analyzed what kind of trend for unsuitable access to ssh servers with the IP address of University of Tsukuba is. In this paper, we report the results of analysis.

Keywords: honeypod, network analysis, campus network operation

1. はじめに

現在，筑波大学には，IP アドレスとしてクラス B の領域が 2 つ割り当てられている．筑波大学において IP アドレスは有効的に利用されているが，それでも未使用の IP アドレスの領域が存在している．これらの利用されていない IP アドレス領域宛のパケットが筑波大学内部にあるルータに到着すると，そのルータではデフォルトルートが大学の境界に設置されているファイアウォールに向けられているため，ルータとファイアウォールの間で何度も送受信が行われてしまう．そのため，未使用の IP アドレス領域宛のパケットは内部のルータにて破棄している．利用されていない IP アドレス宛のパケットであるため，破棄してしまっても問題はない．これらの破棄しているパケットの有効利用として，我々は，ハニーポットを用いて不適切な HTTP リクエストの収集を行った [3].

この研究の結果より，不適切な HTTP リクエストがどのようなものであるかを把握し，Web サーバを安全に運用するために利用した．

また，この研究により，遠隔地にある計算機へのリモート接続をするために広く使われている SSH (Secure Shell) サービスに対するアクセスが多数あることがわかった．SSH サーバをより安全に運用するためには，どのような不適切なアクセスがあるのかを把握することは大変重要である．

本研究では，先行研究を発展させ，SSH サービスへのアクセスに関する情報，特にどのようなパスワードを用いてログイン試行を行っているか情報の収集を行うことを目的とした．この目的を実現するために， ネットワークセグメ

[1] 筑波大学 学術情報メディアセンター
Academic Computing and Communications Center, University of Tsukuba
[2] 筑波大学 情報学群情報科学類
College of Information Science, The School of Informatics, University of Tsukuba
[3] 筑波大学 システム情報工学域 情報工学域
Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba
[4] 筑波大学ビジネスサイエンス系
Faculty of Business Sciences, University of Tsukuba
[a)] akira@cc.tsukuba.ac.jp

ントのエミュレートに優れたハニーポットである Honeyd [7] ，および SSH サービスのエミュレートに優れたハニーポットである Kippo [10] を組み合わせたシステムを新たに開発し，実装した．

入力されたパスワードには，IT 企業等が公表する「セキュリティ強度の低いパスワード」 [8],[9] が多く用いられていたことが明らかになった．

2. 関連研究

2.1 複数のハニーポットを組み合わせたシステムによるネットワークの監視

Hassan ら [6] は，ネットワーク内に設置した Honeyd サーバのプロキシ機能を使用して，Honeyd へのトラフィックを高対話型ハニーポットに中継するハイブリッド型ハニーポットフレームワークを提案している．Honeyd は 1 台のマシンで何千ものホストをネットワークレベルでエミュレートできる．この特徴と，高対話型ハニーポットが持つ高度なエミュレート機能を組み合わせることにより，ネットワーク上にあたかも本物のサーバが何千台も稼働しているかのような振る舞いを可能としている．

本研究においても，1 台のマシンで何千のもホストをエミュレーションする必要があるため，Honeyd のプロキシ機能を用いる．本研究は，SSH サービスへの不適切なアクセス情報を収集することを目的としてしているため，SSH サービスのエミュレーションを行う Kippo と組み合わせている．

また，溝口らの研究 [4] では，大学のネットワーク内に設置した Honeyd サーバのプロキシ機能を利用して，高対話型ハニーポットおよび低対話型ハニーポットへと通信の中継を行い， 大学ネットワークの監視を行っている．Honeyd には，大学ネットワーク内の DHCP セグメントにて使われていない IP アドレスを割り当てることにより，IP アドレスの衝突問題を回避している．

本研究では，Honeyd にて応答させる未使用 IP アドレスの選定は， 大学のコアルータが行う．Honeyd では送られてきたすべての通信に対して， その送信先 IP アドレスを割り当てられたホストとして応答する．

2.2 SSH 総当たり攻撃における攻撃元 IP アドレスからの通信を遮断する研究

大隅らの研究 [2] や，阿波連らによる研究 [1] は， SSH サーバのログから SSH サービスに対するパスワード総当たり攻撃を検出し，ホスト間で連携して不正アクセスを防止する手法を提案している．各ホストのアクセスログを中央サーバにて解析し，攻撃者と断定された IP アドレスの情報を各ホストへと伝達することによりアクセス制御を行う．

本研究では，SSH サービスへのアクセス情報の収集に，攻撃者に関する情報を収集する目的で設計されている Kippo を用いる．これにより，侵入に関するリスクを限りなく小さくすることが可能となる．また，本研究では，監視対象として未使用 IP アドレスを用いるため， 正常なユーザによるアクセスの可能性を極力排除することが可能である．

3. 不適切な SSH アクセス情報の収集方法

3.1 筑波大学のネットワーク環境

本研究では，筑波大学のネットワークにハニーポットを設置して不適切なアクセス情報を収集する．筑波大学のネットワークは 130.158.0.0/16 と 133.51.0.0/16 とのクラス B の 2 つの IP アドレスが割り当てられている．学内では，所有している IP アドレスをさらに小さいサブネットに分けて利用している．そのため，利用されていないサブネットが存在する．そのサブネット宛の通信は原則として不適切な通信である．

なお，これらの IP アドレスは論理的には，学内と学内の境界にあるファイアウォールに内側に接続されている．したがって，学外からこれらの IP アドレスへのアクセスは，ファイアウォールによる制限を受けている．筑波大学のファイアウォールは，学外から学内への TCP/22 番ポートへのアクセスは許可している．

3.2 収集システムの設置

本研究では，収集システムを内部のコアルータへ接続する．さらに，収集システムは，上記の 2 つのクラス B のネットワーク宛の通信であれば応答するように設定する．さらに，内部のコアルータのルーティングテーブルに対して筑波大学に割り当てられている 2 つのクラス B のネットワークへの next hop として収集システムの IP アドレスを設定する．内部のルータは，最長一致の方式によりルーティングを決定するため，利用されているサブネットは，2 つのクラス B のネットワークよりもネットマスクが長くなるために，正しいルーティングが行われる．使われていないサブネットに対しては，設定したルーティングが用いれるため，結果として学内ネットワークにおいて利用されていないサブネット宛の通信は収集システムに送られる．これにより，収集システムは，それらの通信を監視することが可能になる．

3.3 収集システムの概要

本研究で提案する収集システムは Honeyd [7] と Kippo [10] から構成される．

Honeyd は，仮想ホストをエミュレートし，侵入者に関する情報を収集できるオープンソースの低対話型ハニーポットである．Honeyd は，複数の仮想的なネットワークセグメントのエミュレーションが行える．このため，自分自身に割り当てられた IP アドレス以外の通信であっても，そ

の IP アドレスを割り当てられたホストとして応答し，その IP アドレス上で特定のサービスが運用されているかのように振る舞う． また Honeyd は受け取った通信を別ホストに中継するプロキシ機能を有する．

Kippo は，SSH サービスのエミュレートに特化したオープンソースの低対話型ハニーポットである．Kippo は，SSH サービスに対するパスワードの総当たり攻撃に関する情報を記録できる．Kippo は自分自身に割り当てられた IP アドレス以外の通信には応答できない．

収集システムでは，初めに Honeyd　が SSH アクセス，すなわち，TCP/22 番ポートへの通信を受け取る．Honeyd は SSH サービスへの通信を受けたとき，Honeyd のプロキシ機能を用いて別サーバ上の Kippo に通信を中継する．Kippo は，中継された通信に対して SSH サービスの擬態を行い，入力されたユーザ名やパスワードを収集する．この連携により，Honeyd による仮想的なネットワークセグメントの利用，および Kippo による高度な SSH サービスの擬態という両者の利点を活かしたシステムを実装した．

Honeyd および Kippo におけるログの照合には，Honeyd 側のポート番号を用いる．Honeyd のプロキシ機能を利用して Kippo に SSH アクセスを中継すると，Kippo のアクセスログには，送信元 IP アドレスおよびポート番号として Honeyd サーバの IP アドレスおよび ポート番号が，送信先 IP アドレスとして Kippo サーバの IP アドレスがそれぞれ記録される． このため，学外からの攻撃者の IP アドレスや，学内の攻撃先の IP アドレスは，Kippo のアクセスログからは判別できない．

一方，Honeyd によるプロキシ機能を利用する際に出力されるログには， 送信元や送信先，中継先に関する情報が記録されるが，中継の際に用いた Honeyd 側のポート番号は記録されない．

本研究では， Honeyd がプロキシ機能を使用する際に Honeyd 自身のポート番号もログに出力するようにプログラムを修正した．これより，Honeyd と Kippo のそれぞれのログを照合することを可能とした．

本研究では SSH アクセスに関して，通信の開始時刻，終了時刻，送信元 IP アドレス，送信元ポート番号， 送信先 IP アドレス，セッション番号を記録する．収集システムがパスワード入力を受けた場合は，入力されたユーザ名やパスワードを，セッション番号や Honeyd 側ポート番号と紐付けて記録する．これらの情報は MySQL に格納した．

4. 収集したデータの解析

3 章で述べた収集方法により，2013 年 12 月 1 日から 31 日までの 1 ヶ月間に収集した情報をもとに解析を行った結果を示す．本研究では収集システムを収集期間の前日にあたる 11 月 30 日より稼働させた．収集システムを稼働させる以前には，未使用 IP アドレス上で応答するサーバは存在していない．なお，本章では，送信元 IP アドレスについては上位 16 ビットをアルファベット等に置き換えて表記する．また，本研究においては，SSH アクセスの送信元をホスト，送信先をサーバと呼ぶことにする．

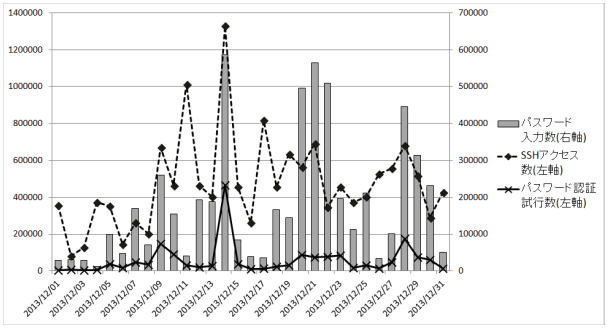

図 1　Honeyd および Kippo で収集した情報の推移

4.1 収集したデータの概要

図 1 に， 本研究にて収集した 1 ヶ月間の SSH アクセスの概要を示す．ここで，SSH アクセス数とは，Honeyd が 22 番ポートに関する通信を受けて，Kippo へ通信を中継した回数である．また，パスワード認証試行数とは，Kippo がホストに対して パスワード入力要求を送信した回数である．この期間における SSH アクセス数は，14,696,933 セッションであり，1 日平均 47 万セッションであった．このうち，パスワード入力が行われたセッション数は，1,813,911 セッションであり，これは全体の 12% に当たる．ほとんどがパスワードを入力しない，すなわち，ポートスキャンであることがわかる．

この期間におけるサーバは 65,003 台であり，これは未使用 IP アドレスすべてが攻撃を受けていることを表している．

4.2 ホストに対しての解析

Honeyd が受信した 22 番ポート宛の通信に関して，ホストの分布がどのようになっているかを調べた．期間内に 22 番ポート宛の通信を行ったホストは 3,373 台である．

4.2.1 SSH アクセス数による解析

表 1 は，期間内におけるそれぞれのサーバが受信した 22 番ポート宛のアクセス数のうち，ホスト上位 10 台を示している．表 1 より，SSH アクセスの多いホストが必ずしもパスワード入力まで積極的に行うわけではないことがわかる．これは，SSH サーバを探す行為と，パスワード総当たり攻撃を行う行為とが，別々のホストにて行われていることを示している．

4.2.2 SSH アクセス先サーバ数による解析

表 2 は，期間内におけるそれぞれのサーバが受信した 22 番ポート宛のレコード数のうち，接続先サーバ数の多いホスト上位 10 台を示している．期間内にすべての未使用 IP

表 1 期間内における SSH アクセス数を軸としたホスト上位 10 台

順位	ホスト	SSH アクセス数	アクセス先サーバ数	パスワード入力数	入力先サーバ数	パスワード種類数	ユーザ名種類数
1	A.14.140	2022025	10736	11306	1251	673	1
2	B.185.73	827789	8072	357	121	53	1
3	C.74.223	703451	36089	1131374	35368	140	1
4	D.144.190	632105	39144	332	15	6	1
5	E.141.32	278040	8703	23	6	10	1
6	F.236.18	185001	42930	0	0	0	0
7	G.239.70	174998	50973	443587	32105	602	1
8	H.37.114	171587	3805	82	14	54	1
9	I.168.170	167563	22044	0	0	0	0
10	J.74.239	166438	7003	141034	7003	163	119

表 2 期間内における SSH アクセス先サーバ数を軸としたホスト上位 10 台

順位	ホスト	SSH のアクセス数	SSH のサーバ数	パスワード入力数	入力先サーバ数	パスワード種類数	ユーザ名種類数
1	G.239.70	174998	50973	443587	32105	602	1
2	W.231.190	63847	50932	0	0	0	0
3	Q.183.158	89842	43149	0	0	0	0
4	F.236.18	185001	42930	0	0	0	0
5	g.192.40	43751	41441	0	0	0	0
6	D.144.190	632105	39144	332	15	6	1
7	d.32.48	48252	37075	0	0	0	0
8	k.54.172	41364	36997	4	4	4	1
9	F.224.66	36200	36197	0	0	0	0
10	C.74.223	703451	36089	1131374	35368	140	1

表 3 期間内におけるパスワード入力数を軸としたホスト上位 10 台

順位	ホスト	SSH のアクセス数	SSH のサーバ数	パスワード入力数	入力先サーバ数	パスワード種類数	ユーザ名種類数
1	C.74.223	703451	36089	1131374	35368	140	1
2	G.239.70	174998	50973	443587	32105	602	1
3	P.184.109	90709	4812	410596	4199	2288	4
4	G.239.72	81736	26820	278223	13847	584	1
5	G.239.133	75269	16356	160057	8938	546	1
6	G.239.75	83393	33241	156719	17636	610	1
7	C.113.77	23993	10273	147076	5865	339	1
8	J.74.239	166438	7003	141034	7003	163	119
9	C.113.85	45468	14510	133238	10286	400	1
10	Y.133.51	49262	23469	116355	10189	456	1

表 4 期間内におけるパスワード入力先サーバ数を軸としたホスト上位 10 台

順位	ホスト	SSH のアクセス数	SSH のサーバ数	パスワード入力数	入力先サーバ数	パスワード種類数	ユーザ名種類数
1	C.74.223	703451	36089	1131374	35368	140	1
2	G.239.70	174998	50973	443587	32105	602	1
3	c.102.182	49321	32268	44070	29477	2	2
4	G.239.75	83393	33241	156719	17636	610	1
5	G.239.72	81736	26820	278223	13847	584	1
6	C.116.5	58545	27224	93974	12298	495	1
7	C.113.85	45468	14510	133238	10286	400	1
8	Y.133.51	49262	23469	116355	10189	456	1
9	G.239.133	75269	16356	160057	8938	546	1
10	J.74.239	166438	7003	141034	7003	163	119

表 5 期間内における入力数を軸としたパスワード上位 10 種類

順位	パスワード	のべ入力数	入力元ホスト数	入力先サーバ数	ユーザ名との組み合わせ数	WORM: Conficker	adobe-top100
1	admin	306671	547	60872	5	hit	
2	qm21	266472	42	35382	1		
3	qm22	253331	42	35163	1		
4	123456	227798	362	53296	11	hit	1
5	root	124180	519	46628	4	hit	
6	1234	56012	367	30557	8	hit	14
7	----------	49096	1	1397	2		
8	abc123	43772	573	25480	10	hit	13
9	963963369	43742	17	21679	2		
10	1qaz2wsx	41788	41	23572	3		46

表 6 期間内における入力元ホスト数を軸としたパスワード上位 10 種類

順位	パスワード	のべ入力数	入力元ホスト数	入力先サーバ数	ユーザ名との組み合わせ数	WORM: Conficker	adobe-top100
1	test	22078	672	15605	5	hit	85
2	testing	20062	654	10964	3		
3	12341234	7443	638	5462	1		
4	backup	6612	631	4832	3	hit	
5	abc123	43772	573	25480	10	hit	13
6	redhat	12200	561	10122	3		
7	superman	12403	559	7752	3		39
8	rootroot	12356	559	9843	2	hit	
9	webmaster	8220	553	7166	2		
10	shell	15993	548	10201	2		

表 7 期間内における入力数を軸としたユーザ名上位 10 種類

順位	ユーザ名	のべ入力数	入力元ホスト数	入力先サーバ数	パスワードとの組み合わせ数
1	root	5323185	1735	63432	5842
2	admin	131747	5	33018	644
3	support	15296	1	11727	1
4	debbie	12733	1	6324	2
5	apache	9303	3	409	23
6	test	9205	1	5991	8
7	backup	7922	2	477	57
8	anke	6802	1	6789	1
9	anja	6454	1	5615	1
10	anna	6338	1	6204	1

アドレスへ SSH アクセスを行うホストはみられなかった．また，表 2 より，SSH アクセス先サーバ数の多いホストが必ずしもパスワード入力まで行っていないことがわかる．

4.2.3 パスワード入力数による解析

表 3 は，期間内におけるそれぞれのサーバが受信した 22 番ポート宛のレコード数のうち，パスワード入力数の多いホスト上位 10 台を示している．表 3 における “C.74.223” ホストは，パスワード入力数が 1,131,374 件であり，これはパスワード入力に関するレコード全体のおよそ 20%にあたる．“P.184.109”以外のホストにみられるように，パスワード入力数とパスワードの種類数は比例しないことがわかる．また，表 3 によると，上位 10 台のうち 4 台のホストの IP アドレスが “G.239. ∗ ”であった．レコードの各項目にも類似点がみられることから，この 4 台のホストには何らかの関係性があると考えられる．

4.2.4 パスワード入力先サーバ数による解析

表 4 は，期間内におけるそれぞれのサーバが受信した 22 番ポート宛のレコード数のうち，パスワード入力先サーバ数の多いホスト上位 10 台を示している．表 4 における SSH アクセス先サーバ数とパスワード入力先サーバ数とを

比較すると，“G.239.72”ホストのように SSH アクセス先サーバ数の半分程度のサーバにパスワード入力を行うホストや，“J.74.239”ホストのように SSH アクセス先サーバ数とパスワード入力先サーバ数が一致するホストが存在することがわかった．

4.3 パスワードに対しての解析

どのパスワードが多く利用されたかについて調査するため，期間内にそれぞれのサーバが受信したパスワード入力のログを解析した．なお，期間内に入力されたパスワードは 6150 種類である．

4.3.1 入力数による解析

表 5 は，期間内におけるそれぞれのサーバが受信したパスワードの入力数上位 10 種類を示している．“WORM:Conficker”カラムにて，2008 年に初めて検出されたワームである Conficker が利用するパスワード [9] に該当するかを示す．このワームは，Windows OS を標的とし，感染したマシンが属するネットワーク上の他のマシンに接続しようとする際に，ワームに組み込まれたパスワード群を利用する．Conficker が利用するパスワードは全部で 248 種類であり，本研究にて収集したパスワードでは 198 種類が該当した．また，“adobe-top100”カラムにて，Adobe Systems 社が被害を受けた不正アクセスによるユーザ情報流出事件 [5] に関して，Stricture Consulting Group(以下，SCG 社) が公表したパスワード 100 種類 [8] に該当するかを示す．本研究にて収集したパスワードでは 76 種類が該当した．“adobe-top100”の数値は，SCG 社が公表した流出アカウントに使用されていたパスワード数よりランキングされた順位である．この 2 項目から，攻撃者が用いるパスワードは，セキュリティ企業が公表した「セキュリティ強度の低いパスワード」を多く含むことがわかる．これはホスト側が攻撃の際にこれらの情報を参考に入力するパスワードを決定していることが推測される．

表 5 における 10 種類のパスワードによる入力数の総計は 1,412,862 件であり，これは全パスワードの入力数のおよそ 25%にあたる．

4.3.2 入力元ホスト数による解析

表 6 は，期間内におけるそれぞれのサーバが受信したパスワードのうち，入力元ホスト数による上位 10 種類を示している．本研究における期間内に最も多くのホストから入力されたパスワードは “test” であり，605 台のホストから入力された．

4.4 ユーザ名に対しての解析

Kippo が受信したパスワード入力に関して，どのユーザ名が多く利用されたかを明らかにするため，期間内でそれぞれのサーバが受信したユーザ名の入力数を計測した．なお，期間内に入力されたユーザ名は 126 種類である．

表 7 は，期間内におけるそれぞれのサーバが受信したユーザ名の入力数上位 10 種類を示している． 表 7 より，大多数のホストが”root”というユーザ名を用いてログインを試みていることがわかる．

5. おわりに

本研究では，筑波大学のネットワークにおける未使用 IP アドレスを利用し，SSH アクセスの収集を行い，収集した情報を元に解析を行った．

本研究の今後の課題について，収集した情報をさらに精細に調べる必要がある．たとえば，集計したデータを数時間ごとに区切った場合，どのサーバに対しても SSH アクセスが来ない期間や，数分の間に数十万もの SSH アクセスが来た期間などがみられた．短い期間における SSH アクセスの解析を行うことに，SSH アクセスに関する傾向の把握に可能性がある．

参考文献

[1] 阿波連良尚, 新城靖, 佐藤聡, 中井央, 板野肯三: ホスト協調による SSH パスワード総当たり攻撃の防御. 第 21 回 コンピュータシステム・シンポジウム (ComSys 2009), ポスター・デモセッション, (2009).
[2] 大隅淑弘, 山井成良: ホスト間連携を可能にするパスワード総当たり攻撃対策手法 (セッション 3). 情報処理学会研究報告. DSM, [分散システム/インターネット運用技術], Vol. 2007, No. 93, pp. 49–54,(2007).
[3] 佐藤聡, 三田尚貴, 新城靖, 板野肯三：ハニーポットを利用した筑波大学の未使用 IP アドレス宛ての HTTP リクエストの解析, 情報処理学会研究報告. IOT, [インターネットと運用技術] Vol. 2013, No. 8, pp.1–6,(2013).
[4] 溝口誠一郎, Erwan Le Malecot, 堀良彰, 櫻井幸一: DHCP によって管理されたセグメントに存在する未使用 IP アドレスの監視手法. 情報処理学会研究報告. CSEC, [コンピュータセキュリティ], Vol. 2008, No. 45, pp. 55–60, (2008).
[5] Adobe Systems: Important Customer Security Announcement. available from ⟨http://blogs.adobe.com/conversations/2013/10/important-customer-security-announcement.html⟩. accessed: 2014/01/24.
[6] Hassan Artail, Haidar Safa, Malek Sraj, Iyad Kuwatly, and Zaid Al-Masri: A hybrid honeypot framework for improving intrusion detection systems in protecting organizational networks. Computers & Security, Vol. 25, No. 4, pp. 274–288, (2006).
[7] Niels Provos: A Virtual Honeypot Framework. USENIX Security Symposium, Vol. 173, (2004).
[8] Stricture Consulting Group: Top 100 Adobe Passwords with Count. available from ⟨http://stricture-group.com/files/adobe-top100.txt⟩. accessed: 2014/01/24.
[9] Symantec: W32.Downadup.B Technical Details. available from ⟨http://www.symantec.com/security_response/writeup.jsp?docid=2008-123015-3826-99&tabid=2⟩. accessed: 2014/01/24.
[10] The Honeynet Project: Kippo - SSH honeypot. available from ⟨http://www.honeynet.org/project/Kippo⟩. accessed: 2014/01/24.