新しいタイプの分散型SSHログインブルートフォース攻撃 STBF の攻撃元数の推定

齊藤 聡美[1]　武仲 正彦[1]　鳥居 悟[1]

概要：CSS2015 で我々が報告した, 1 拠点に対し 1 組のユーザ名／パスワードで 1 回のみのログイン試行が断続的に発生する, 新しいタイプの分散型 SSH ログインブルートフォース攻撃 (Brute force attacks with Single Trial, STBF) について, 攻撃に使用された Botnet の規模を推定する. 攻撃元 IP アドレスの重複数から, 確率モデルを構築する推定方式と, 2 種の Mark-Recapture 法を適用した推定方式を用い, 実際の STBF データから攻撃元の推定実験を実施, 比較を行った. その結果, STBF 攻撃には約 29000 の Botnet が利用されていると推定できる.

キーワード：SSH, ログインブルートフォース攻撃, STBF, 攻撃元数, 推定

Estimation of The Number of Attack Sources for Novel Type Distributed SSH Login Bruteforce STBF

Satomi Saito[1]　Masahiko Takenaka[1]　Satoru Torii[1]

Abstract: In this paper, we estimate a botnet scale regarding a novel type distributed SSH login brute force attack (brute force attacks with Single Trial, STBF). A host try to login for a target by using a user name and a password pair in the attack, which has been reported in CSS2015. We estimate the botnet scale by using three estimation methods: one is the method that constructs the probabilistic modes, and the others are two mark-recapture methods. As the result, we can estimate that 29000 hosts are used as the botnet for STBF.

Keywords: SSH, Login Bruteforce Attack, STBF, Attack Source, Estimate

1. はじめに

近年, ネットワークサービスに対するブルートフォース攻撃は巧妙化の一途をたどり, 侵入検知システム (Intrusion Detection System, IDS) だけでは攻撃の発生を検知することも, 効果的な対策を適用することも難しくなってきている．これに対し我々は, SSH(Secure Shell) を対象とした, ログインブルートフォース攻撃の分析を目的とするログインセンサを開発し, 世界 7 拠点に配置した. 本センサを用いて, ログイン試行を観測, センサに到達したログイン試行を 6 か月間, 約 2800 万のログイン試行を観測した. さらに, 観測データを分析した結果, 1 拠点に対し 1 組のユーザ名／パスワードで 1 回のみのログイン試行が断続的に発生する新しい分散型の攻撃事象（Bruteforce attacks with Single Trials, STBF）を抽出した [5].

STBF は, 1 つの攻撃元 IP アドレスからのログイン試行は 1 回のみであり, 正規ユーザによるログイン試行と区別をつけることは困難である．しかし, 正規ユーザのアクセスノイズがないセンサでの観測であること, ログイン試行が短期間に集中していること, 複数拠点にて同様の事象が継続して観測されること, 攻撃元となった IP アドレスの国情報や試行に用いられたユーザ名／パスワードに STBF 間で高い相関がることから, これらは一連の攻撃であると判断した [5].

[1] 株式会社富士通研究所
FUJITSU LABORATORIES LTD.

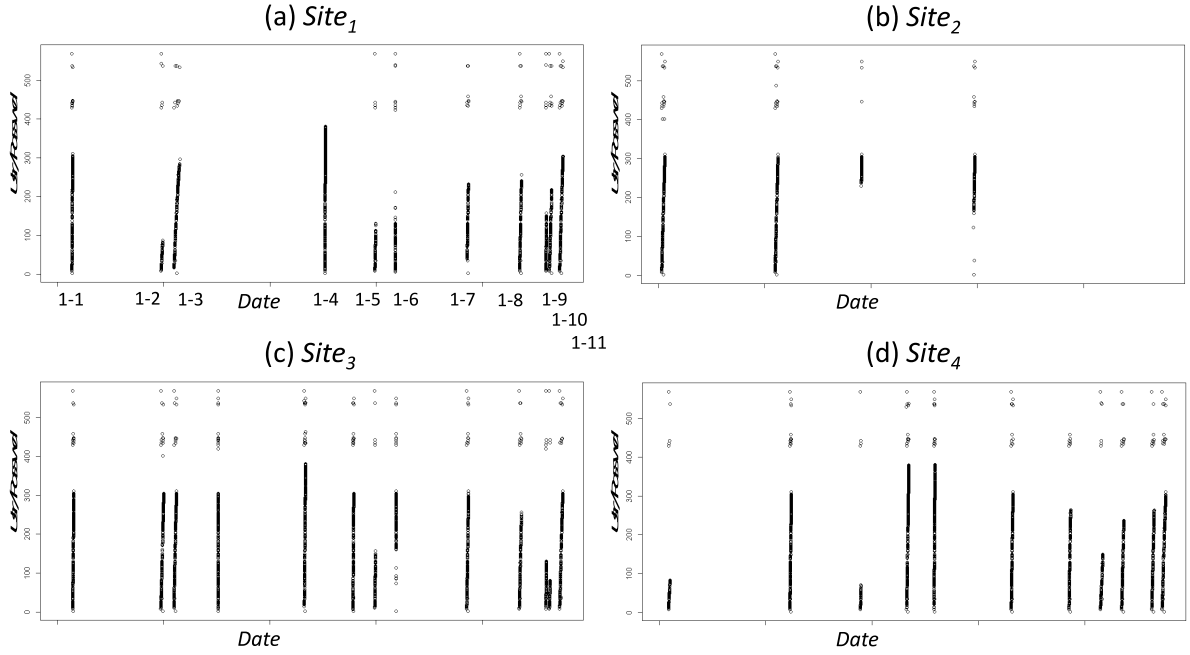

図 1 STBF 事象のグラフのプロット例

STBF は多数の攻撃元 IP アドレスから代わる代わる, 1 回のみのログイン試行が行われる. このことから, 1 セットの STBF は, 一人の攻撃者が Botnet と思われる多数の機器 (以降 Botnet と記載) に指示を与えて攻撃していると推測できる. ここで, 1 セットの STBF では, 攻撃元 IP アドレスに重複は無いが, 複数セットの STBF の攻撃元 IP アドレスを調査をすると, 少数ながら重複する攻撃元 IP アドレスが存在する. そこで本稿では, この攻撃元 IP アドレスの重複から, 攻撃に利用されている Botnet の規模の推定を試みた. 推定方法は次の 3 種類を使用した. 攻撃の確率モデルを構築し, Birthday Paradox から Botnet の規模を推定する方法, 個体群生態学で利用されている mark-recapture 法の一つ Schnabel 法による推定 [8], 別の mark-recapture 法である Jolly-seber 法による推定 [8] である. これらの方法を用いた結果, 前の 2 手法で約 29000 の攻撃元 IP アドレスからなる Botnet が利用されていると推定できた. しかし残念ながら, 最後の推定方法については安定した推定値を得ることが出来なかった.

本稿の貢献は, STBF の攻撃元 IP アドレスの重複から, 確率モデル構築や mark-recapture 法により Botnet の規模を推定する方式を提案したとであり, その方式を用いることにより, 観測した STBF に用いられている Botnet の規模を約 29000 と推定したことである. これは, 直接観測ではなく Botnet の規模を, 攻撃の統計量から推定した始めての成果であると考える.

本稿の構成は以下のとおりである. 第 2 章でこれまで観測した STBF の詳細について述べる. 第 3 章で確率モデルを用いた攻撃元数の推定法を, 第 4 章で mark-recapture 法を用いた推定法を示し, 第 5 章で従来研究も含めた BotNet の規模について議論を行う. 第 6 章でまとめと今後の課題を述べる.

2. STBF (Bruteforce attacks with Single Trials)

本章では, 我々が [5] で報告した STBF について述べる.

2.1 ログインセンサと抽出手順

STBF を抽出したログインセンサは, TCP22 番ポートで通信を待ち受け, ログイン試行を記録する. 試行されたログインに対してはすべてログインエラーと応答する. 記録する情報は, ログイン試行元 IP アドレス, センサが設置される拠点名, 試行発生時刻, 試行されたユーザ名とパスワード 5 つである.

本ログインセンサで収集した試行から, 次の手順により STBF 事象を抽出する.

(1) 1 つの試行元 IP アドレスから 1 つのサイトに対し 1 回のみ実施されたログイン試行を抽出
(2) 抽出したログイン試行をユーザ名/パスワードの組み合わせで時系列にグラフにプロット (図 1)
(3) ログイン試行が連続してに観測された箇所を目視により抽出, STBF とする

2.2 抽出結果

本手順により, 2015 年の 8 ヶ月間, 世界 7 拠点で観測を実施し, STBF の抽出作業を行った. その結果, 合計 91 回の STBF の抽出に成功した. ただし, 内 3 回はログイン試

表 1　攻撃元 IP アドレスの回数の分布 (%)

出現回数	[5] の分布	今回の分布
1	90.2	83.9
2	8.3	11.5
3	1.2	2.9
4	0.3	1.0
5	0.0	0.5
6 回以上	0.0	0.3

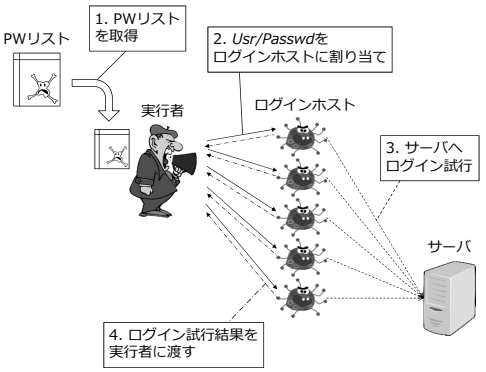

図 2　推測できる STBF の実行手順

行元 IP アドレス (以降, 攻撃元 IP アドレス) の数が 4〜6 個と小規模な STBF であった.

今回の観測期間は, [5] の 6 ヶ月を含む 8 ヶ月となっている. 抽出された STBF の特徴は, 出現回数の分布以外 [5] とは大きく変わらなかった. 出現回数の分布については, 表 2.2 に示すように, 観測期間が延びた分, 同じ攻撃元 IP アドレスが再使用される割合が大きくなっている.

2.3　STBF の攻撃方法に関する推測

[5] では STBF の特徴から, 攻撃方法について次のように推測を行っている. 1 セットの STBF は, 1 人の攻撃者がログイン試行を行うためのホスト群を用意し, 各ホストに対し 1 回ずつのログイン試行を指令する. この攻撃手順を複数の攻撃者が実施することで, 観測されている現象が発生すると考えられる (図 2).

3.　確率モデルを用いた攻撃元ホスト数の推定

抽出結果から, STBF の観測数が増加するに従い, 再使用される攻撃元 IP アドレスの割合が増加している. これは Birthday Problem の確率と同様の振る舞いである. このことから, STBF のログイン試行を行うホストは, 非常に大きな Botnet 等のホスト群 (以下, Botnet) からランダムに選択されていると考えられる.

そこで, 攻撃元 IP アドレスの重複に注目し, STBF の確率モデルを構築, ログイン試行を行うホストを選択する Botnet の規模の推定を行う.

3.1　確率モデルの構築

STBF の確率モデルを次のように構築する. Botnet を構成するホストの数を n 台, 攻撃者はそこらランダムで重複なしに x 台のホスト選択し 1 回のログイン試行に使用, a 回の攻撃を行ったときの, 同じホストを使う期待値を Ex_a とする. すると a 回の攻撃時には, 重複回数により $ax+1$ 個の確率状態を持つ. これを $\mathrm{St}_{a,j}$ とすると, 期待値は以下のように表すことが出来る.

$$\mathrm{Ex}_a = \sum_{j=0}^{ax} j\mathrm{St}_{a,j}$$

a 回の攻撃時の確率状態 $\mathrm{St}_{a,j}$ は, $a-1$ 回の攻撃時の確率状態 $\mathrm{St}_{a-1,j}$ と遷移確率 $\mathrm{Pr}_{n,m_{a,j},x,i}$ から次のように表すことが出来る.

$$\mathrm{St}_{a,j} = \sum_{k=j-x,0\le k\le(a-1)x}^{j} \mathrm{Pr}_{n,ax-k,x,i}\cdot\mathrm{St}_{a-1,k}$$

ここで, $m_{a,j}$ はある確率状態 $mboxSt_{a,j}$ のときの使用済みホストの数, i はその状態で x 台のホスト選択したときの (使用済みでない) 新しいホストの数を表す. また遷移確率 $\mathrm{Pr}_{n,m_{a,j},x,i}$ は次の式で表すことが出来る.

$$\mathrm{Pr}_{n,m_{a,j},x,i} = \frac{x!}{(x-i)!i!}\cdot\frac{m!(n-x)!(n-m)!}{n!(m-x+i)!(n-m-i)!}$$

3.2　Botnet を構成するホスト数 n の推定

本確率モデルを用いて, Botnet を構成するホスト数 n を推定する. 観測から, 攻撃の回数 $a=88$ 回, 1 セットの STBF 当たりの平均ログイン試行回数 $x=146$, のべ重複数を 2477 として計算機実験を行った. なお, ログイン試行回数を平均値としたため, 3 回の小規模な STBF は対象から除外している.

計算機実験では, Botnet を構成するホストの数 n を変更しながら, 重複するホストを使う期待値 Ex_{88} を計算, $\mathrm{Ex}_{88}\sim 2477$ となる n を特定した. その結果, $n\sim 29000$ となった. このことから, STBF は 29000 台のホストからなる Botnet を利用しているものと推定できる.

4.　Mark-Recapture 法を用いた攻撃元数の推定

前節では, STBF の確率モデルを構築し, 攻撃元 IP アドレスの重複確率から, 攻撃が利用している Botnet の規模を推定した. この推定方式では, モデルの単純化のため, 攻撃回数にばらつきがあるにもかかわらず, 各攻撃の回数を測定結果の平均値としている. また, 88 回の攻撃全体での重複確率の計算を行っているため, 途中状態を十分活用していない.

そこで, Mark-Recapture 法 (標識差異捕獲法, 以下 MR 法) を導入する. MR 法は主に個体群生態学で使用される, 個体群を構成する個体数を推定する為の方法の一つである.

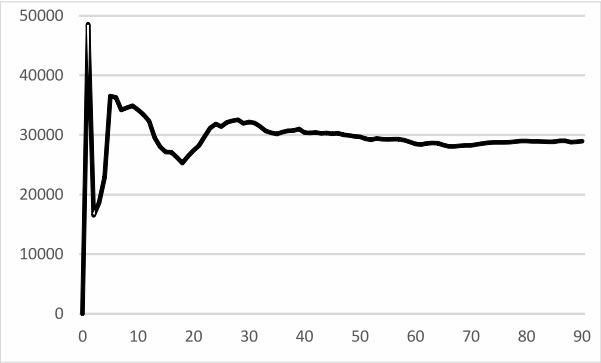

図 3 Schnabel 法による Botnet を構成するホスト数 N の推定

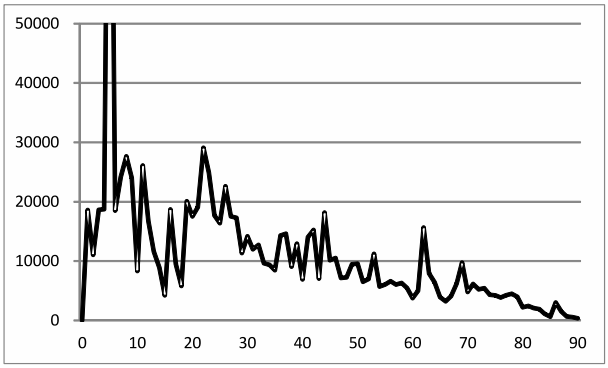

図 4 Jolly-Seber 法による Botnet を構成するホスト数 N_t の推定

全個体を集計することが困難な場合に, 一部の個体を捕獲・標識を付けた後開放し, 再度捕獲したときに標識を付けた個体がどれくらい含まれるかから全体の個体数を把握するものである. MR 法にはさまざまな個体数推定法が提案されている [2], [4], [6], [7], [8].

本章ではその中から, 捕獲を複数回実施して個体数を推定する 2 方式の適用を行う. Botnet を構成するホスト数を個体数, 攻撃を捕獲, IP アドレスの記録を標識付けとみなして MR 法を適用することで, Botnet を構成するホスト数を推定する.

4.1 Petersen の標識再捕獲法

最初に最も基本的で単純なモデルである Petersen の標識再捕獲法を紹介する [4]. これは最捕獲 1 回の方式である.

個体数が変動せず, 個体が短時間に十分混じり合えるフィールドにおいて, 次の処理を実施する.

(1) フィールドから M 個体を捕獲し, それらに標識をつけてフィールドに戻す
(2) 再度 C 個体の捕獲を行う

再捕獲数 C 個体中, R 個体に標識が付いていたとすると, 全体個体数 n は次のように推定できる.

$$N = \frac{MC}{R}$$

本方式は, 全体個体数 N と比較して, 捕獲数 M, C が大きいほど誤差が小さくなる.

4.2 Schnabel 法による推定

前節の確率モデルによる推定より, Botnet を構成するホストの数は, 1 回の STBF においてログイン試行を行うホスト数と比較すると, 十分大きいとはいえない. そこで, 捕獲を複数回って全体個体数の推定を行う Schnabel 法 [6] を適用する. 本手法は Petersen 法の拡張であるため, 個体数が変動せず, 個体が短時間に十分混じり合えるフィールドを前提とする.

Schnabel 法を以下に示す.

(1) フィールドから n_0 個体を捕獲, 標識に付けてフィールドに戻す
(2) フィールドから n_i 個体を再捕獲
(3) その中の再捕獲個体数 x_i を記録
(4) 捕獲された未標識個体に標識を付ける
(5) すべての捕獲個体を開放する
(6) (2)〜(5) を r 回繰り返す

全体個体数 N は次の式で推定できる.

$$N = \frac{\sum_{i=1}^{r} n_i X_i}{\sum_{i=1}^{r} x_i}$$

ここで, X_i は第 i 回再捕獲時のフィールドの標識個体数で, 以下のように表すことが出来る.

$$X_i = X_{i-1} + (n_{i-1} - x_{i-1})$$

本推定方式を用いて, STBF における Botnet を構成するホスト数 N の推定を行った. 本推定方式では, 攻撃の回数の平均値を用いないため, 3 回の小規模な STBF も対象としている. Botnet を構成するホスト数 N の推定結果を図 3 に示す. この結果より, Schnabel 法でも Botnet を構成するホスト数 $N \sim 29000$ と推定できた.

4.3 Jolly-Seber 法による推定

Jolly-Seber 法は, 3 回以上の調査により, 2 回以上の標識付けと開放を行った場合に使用できる手法である [2], [7]. この手法は, 全体個体数の推定を行うだけでない. 全体個体数の変動を前提とし, 新規加入数や消失数, 生存率などの推定も可能である. 本方式が STBF の観測結果に適用できれば, Botnet を構成するホストの増減まで明らかにすることが可能となると考えられる. Jolly-Seber 法による推定手法を示す.

推定に必要な値は, 調査時点 t の捕獲数 n_t, そのうち標識個体数 m_t, 非標識個体数 u_t, 開放個体数 s_t(捕獲から開放までの間の死亡・除去などを考慮している), t 回目の調査で初めて捕獲・標識付け・開放され, それ以降の調査で再捕獲された個体数を R_t, t 回目以前の調査で捕獲され, t 回目の調査で捕獲されず, それ以降の調査で再捕獲された

個体数を Z_t である.

また本手法で推定する対象は, 調査時点 t での全体個体数 N_t, 生存率 ϕ_t, 加入数 B_t, 標識個体数 M_t, 標識個体比率 α_t である. なお, 本来は真の値と推定値は別の記号 (e.g. 全体個体数 N_t, 推定全体個体数 $\hat{N}_t$) とすべきところであるが, 単純化のため, 本稿では同じ記号を用いる. これらの推定値以下のように算出される.

$$\alpha_t = \frac{m_t + 1}{n_t + 1}$$

$$M_t = \frac{(s_t + 1)Z_t}{R_t + 1} + m_t$$

$$N_t = \frac{M_t}{\alpha_t}$$

$$\phi_t = \frac{M_{t+1}}{M_t + (s_t - m_t)}$$

$$B_t = N_{t+1} - \phi_t(N_t - (n_t - s_t))$$

Schnabel 法と同様に, 本推定方式を用いて, STBF における Botnet を構成するホスト数 N_t の推定を行った. STBF の観測では, 開放個体数は捕獲数と同じ $s_t = n_t$ としている. Botnet を構成するホスト数 N_t の推定結果を図 4 に示す. 図 4 より, 本推定方式では, Botnet を構成するホスト数について安定した推定値を得ることが出来なかった.

5. Botnet の規模についての議論

本章では, これまでに示した 3 つの推定手法と, それらにより推定された Botnet を構成するホストについて議論を行う.

5.1 推定法の比較

確率モデルによる推定と Schnabel 法による推定では, STBF に用いられている Botnet を構成するホスト数がほぼ同じ推定値 29000 となった. これは, 両推定法が共に次の 2 つの前提を元に推定を行っているためだと考えられる.

(1) Botnet を構成するホストの数は一定で, そこからランダムに選択 (捕獲)
(2) 再利用 (最捕獲) の確率から全体数を推定

確率モデルでは, これを Birthday Paradox と見なしてモデル化・推定を行っているのに対し, Schnabel 法では, 全体個体数 : 全体標識個体数 = 捕獲数 : 再捕獲数 との仮定から推定を行っているところが異なる. 同じ仮定の下の異なる推定方法で, ほぼ同じ推定値が得られたことから, この仮定が正しければ, STBF に用いられている Botnet を構成するホスト数は 29000 程度である可能性が高い.

一方, Jolly-Seber 法はより個体群生態学に特化されたもので, 上記条件 (1) を前提としていない. 例えば, 「一度標識付けされた個体は次の捕獲時に再捕獲されやすい」という条件があってもかまわない. 特に STBF への適用に適しないものは, 生存率という概念である. Jolly-Seber 法では, この概念を導入するために, 全体個体数が捕獲数の関数になっている (全体個体数が多いほど捕獲しやすいという考え方). これを STBF に適用すると, 「Botnet を構成する規模が大きいほど, ログイン試行に使用されるホストが多い」となってしまう. そのため, STBF の観測情報を用いた実験では安定した推定結果が出なかったものと考えられる.

表 2 Botnet Statistics [1]

Botnet	#IPaddress
Ponmocup	1,203,691
Artro	167,532
TDSS	90,330
Gbot	59,338
Carberp	25,622
Gozi	22,855
Ramnit	20,599
ZeuS	11,354
PsyEye	4,855

5.2 関連研究

Botnet の規模は直接観測を行い, 観測された IP 数とされることがほとんどである. 例えば, 文献**?**では, 自身で設けた Shinkhole で 24 時間に観測された攻撃基 IP アドレスから各種 Botnet の規模について述べている (表 5.2). また, さまざまな Botnet の特徴とその検出について研究が行われている [3] が, その規模について言及されているものはほとんど存在しない.

直接観測では, 確実に Botnet を構成するホストを特定できる一方で, STBF のような回数が少ない攻撃を行う Botnet の場合は観測が困難なため, その規模を明らかにすることは出来ない.

それに対し, 本研究は攻撃の統計的性質から, Botnet を構成するホスト数を推定するという全く新しい報告である.

6. まとめと今後の課題

本稿では, 我々が CSS2015 で報告した, 新しいタイプの分散型 SSH ログインブルートフォース攻撃 STBF について, 攻撃に利用されていると考えられる, Botnet を構成するホスト数の推定を行った. 推定には, 攻撃元 IP アドレスの重複数から確率モデルを構築する推定方式と, 2 種の Mark-Recapture 法を適用した推定方式を用いた. STBF の観測結果から推定を行ったところ, そのうち 2 種類の推定手法で, Botnet を構成するホスト数が約 29000 であると推定できた. また, 上手く推定できなかった手法について, STBF 適用の観点から原因について見当を行った.

今後は, STBF が観測された拠点別, および STBF で使用されているパスワード辞書の種類別で同様の Botnet を構成するホスト数の推定を行い, 全ての STBF が同じ Botnet

を利用しているのかの検証を行う予定である.

参考文献

[1] abuse.ch: How Big is Big? Some Botnet Statistics. The Swiss Security Blog, available from ⟨https://www.abuse.ch/?p=3294⟩ (2011.05.23).

[2] G. M. Jolly: Explicit estimates from capture-recapture data with both death and immigration – Stochastic model. Biometrika 52, pp. 225–247 (1965).

[3] F. Feily, A. Shahrestani, S. Ramadass: ASurvey of Botnet and Botnet Detection. The Third International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2009), pp. 268-273 (2009)

[4] C. G. J. Petersen: The Yearly Immigration of Young Plaice Into the Limfjord From the German Sea, (1896).

[5] 齊藤, 武仲, 鳥居: SSH ログインセンサによる STBF (Brute Force attacks with Single Trials) の観測, CSS2015, 3E2-4 (2015).

[6] Z. E. Schnabel: The estimation of total fish populations of a lake. Am. Math. Monthly 45, pp. 348-352 (1938).

[7] G. F. A. Seber, A note on the multiple recapture census. Biometrika 52 pp. 249–259 (1965).

[8] W. J. Sutherland: Ecological Census Techniques: A Handbook, Cambridge University Press (1996).