

準公開 Squid 接続計画におけるキャッシュ間接続の利用統計

2P-2

大家 隆弘  
徳島大学 工学部

1 まえがき

WWWトラフィックによるインターネット回線圧迫問題の有効な解決法として、SquidなどのWWWプロキシキャッシュサーバ（以下、キャッシュ）を運用し相互接続を行なうことが推奨されている。

しかし、セキュリティ上の問題から各組織で運用されるキャッシュは公開キャッシュではなく、接続設定が難しくなる。そこで我々は不完全な公開キャッシュ（以下、準公開キャッシュ）の相互接続を促進するシステムを運用し、キャッシュ間接続の促進を図った [1, 2]。

本稿では、本システムにおいて各キャッシュ管理者が接続の取捨の判断材料としうるICPヒット率の情報提供について報告する。

2 準公開キャッシュ相互接続計画概要

本稿で述べる準公開キャッシュは、公開キャッシュに一定の中継禁止オブジェクトを付加したものである。準公開キャッシュを相互接続には複雑な設定が必要であり、これをを円滑に進めるためにキャッシュの情報登録、情報公開および設定支援情報の提供を行なうシステムを構築、運用している [1, 2]。

本システムでは、各キャッシュ (Squid) の登録情報: (サーバ名, HTTP, ICP ポート番号, 所属ISP, 接続許可範囲, 禁止オブジェクトの範囲など) および各キャッシュの管理インターフェース (cache\_object) を介して動的に得られる情報: (運用状況, 設定パラメータ, 接続キャッシュなど) をもとに、各キャッシュが接続を行なうのに必要な設定を設定支援情報として各キャッシュに提供する。

各キャッシュでは設定支援情報を取得して設定ファイルを自動修正するが、接続するか否かについてはキャッシュ管理者の選択に寄る。

表1 登録キャッシュの諸元 (1999年1月12日現在)

ID #	Swap (MB)	ICP /hour	HTTP /hour	ICP Hit	HTTP Hit	Red. ratio	Ver. #
01	2373	10209.6	376.6	0.01	0.44	-	1.1.20
02	1592	8385.5	1460.3	0.02	0.13	-	1.1.20
03	3114	5741.3	1633.5	0.04	0.35	-	1.1.20
04	12	8978.7	108.9	0.00	0.06	-	1.1.14
05	1670	84.0	1602.4	0.18	0.43	-	1.1.20
06	471	0.0	1830.0	-	0.31	0.84	2.0.P2
07	931	0.0	1587.1	0.00	0.31	-	1.1.14
08	1686	29123.8	2435.1	0.01	0.32	-	1.1.2
09	1260	33500.2	8769.5	0.00	0.07	-	1.1.2
10	1704	25253.1	1109.7	0.00	0.25	-	1.1.2
11	399	2977.4	404.2	0.01	0.25	-	1.1.6
12	9234	9837.0	4280.1	0.07	0.36	-	1.1.22
13	6585	13552.9	5808.0	0.16	0.52	-	1.1.21
14	3173	6849.4	4030.5	0.11	0.34	-	1.1.21
15	1643	6285.4	4134.9	0.07	0.20	-	1.1.19
16	3440	6704.7	3816.2	0.08	0.25	-	1.1.21
17	173	360.9	843.1	0.06	0.28	-	1.1.9
18	2317	3245.1	3419.0	0.15	0.19	-	1.1.20
19	116	37.7	41.8	0.10	0.13	-	1.1.20
20	1695	25.4	163.2	0.00	0.41	-	1.1.20
21	14968	141420.0	14712.0	-	0.28	0.87	2.1.P2
22	14945	148428.0	15462.0	-	0.28	0.83	2.1.P2
23	5878	131202.0	12336.0	-	0.38	0.80	2.1.P2
24	3309	116100.0	6702.0	-	0.40	0.79	2.1.P2
25	1573	93540.0	4740.0	-	0.41	0.80	2.1.P2
26	811	76542.0	3894.0	-	0.20	0.97	2.1.P2
27	1486	213.4	1720.1	0.03	0.30	-	1.1.20
28	8399	24000.0	3456.0	-	0.45	0.71	2.1.P2
29	5984	22104.0	2292.0	-	0.36	0.79	2.1.P2
30	5	0.0	28.1	0.00	0.06	-	1.1.20

表1に登録されているキャッシュの諸元を示す。

3 ICP ヒット率の推定

接続の取捨を考える上で個々の接続のICPヒット率が参考となるが、接続数が増えた場合に新たな接続の取捨には全体としてのICPヒット率の方がより参考となる。ここではキャッシュ ( $i$  番目) が発行するICP要求がいずれかのキャッシュでヒットする確率  $H$  を考える。

$i$  番目のキャッシュから  $j$  番目のキャッシュへのICP要求のヒット率を  $h_{i,j}$  とする。ただし、ICP接続が行なわれていない場合は0とする。 $h_{i,j}$  はSquidから動的に得ることができる。

$H$  は  $i$  番目のキャッシュが行なっている全てのICP接続のヒット率の最大値を最小値とし、

$$H \geq \underline{H} = \max_j \{h_{i,j}\} \quad (1)$$

Statistics of ICP Hit-ratio on Inter-connecting Quasi-public Caches  
Takahiro OIE  
Faculty of Engineering, The University of Tokushima  
2-1 Minamijosanjima, Tokushima 770-8506, Japan

表2 キャッシュ間接続のICP要求のヒット率 ( $h_{i,j}$ ) と推定総ヒット率 (1999年1月12日現在)。縦列のC\*はクライアント側のID (表1参照), 横列のS\*はサーバ側のIDを示す。ヒット率は百分率表示で示し, 枠有りは親接続 (parent), 枠無しは隣接続 (sibling) であることを示す。Q<sub>i</sub>はC\*が発行したICP要求数。

c/s	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	Q <sub>i</sub>	H	H <sub>1</sub>	H <sub>2</sub>	
C01	.	.	2	.	.	.	.	1	0	0	0	3	9	.	.	.	.	.	.	.	.	12	12	7	6	1	1	.	.	.	.	.	4914	11.0	41.5	-
C02	0	.	1	0	.	.	0	0	0	0	0	1	0	.	.	.	.	.	.	.	.	6	6	1	1	1	0	.	.	.	.	60012	6.4	14.3	14.3	
C03	0	1	.	0	.	.	0	0	0	1	2	5	.	.	.	.	.	.	.	.	.	13	14	4	3	1	1	.	.	.	.	75028	13.6	30.4	30.4	
C04	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	-	-	-
C05	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	3	.	.	.	.	.	.	121340	4.3	7.3	-
C06	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	-	-	-
C07	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	10	4	.	.	.	.	.	429087	12.5	23.4	46.6	
C08	.	.	.	.	.	.	0	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	8	6	.	.	.	.	.	8393	18.1	32.9	-	
C09	.	.	.	.	.	.	1	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	25	23	20	.	.	.	.	.	24621	33.6	69.0	-	
C10	.	.	.	.	.	.	2	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	10	8	.	.	.	.	.	5262	23.2	39.4	-	
C11	.	.	15	0	.	.	3	2	3	.	11	.	.	.	.	6	.	.	.	.	.	.	20	14	12	9	.	.	.	.	8761	15.8	56.7	-		
C12	0	.	1	0	.	.	1	0	.	.	.	.	5	.	.	.	.	.	.	.	.	.	7	8	3	2	.	.	.	.	28762	5.8	20.9	-		
C13	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	9	9	4	3	2	1	.	.	.	.	138374	7.9	38.3	-
C14	.	.	.	.	.	.	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	16	16	10	7	6	2	.	.	.	.	145438	23.4	60.1	-
C15	.	.	.	.	.	.	.	.	.	.	7	.	.	.	.	.	.	.	.	.	.	.	16	16	9	7	5	2	.	.	.	.	188184	24.0	61.7	-
C16	.	.	.	.	.	.	.	.	.	.	13	.	.	.	.	.	.	.	.	.	.	.	23	22	17	15	12	7	.	.	.	.	154508	30.0	80.4	-
C17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	-	-	-
C18	1	2	4	.	.	.	1	0	0	1	5	7	.	.	.	.	.	.	.	.	.	.	.	8	5	2	1	.	.	.	.	98470	7.9	29.2	44.8	
C19	1	.	1	.	.	.	1	0	.	.	1	3	.	.	.	.	.	.	.	.	.	.	5	5	2	1	1	.	.	.	.	44299	5.3	19.6	-	
C20	.	.	.	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2257	6.5	9.7	-
C21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	73778	4.5	4.5	-
C22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	77307	3.7	3.7	-
C23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16	16	.	.	.	.	.	.	.	89256	16.2	29.7	29.7
C24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17	17	.	.	.	.	.	.	.	44078	17.1	30.9	30.9
C25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22	22	.	.	.	.	.	.	.	29904	21.8	38.8	38.8
C26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	24	.	.	.	.	.	.	.	31052	23.8	41.2	41.2
C27	3	.	7	0	.	.	1	0	1	2	4	14	.	.	4	.	.	.	.	.	.	.	18	17	14	10	4	2	.	.	.	3727	16.7	67.5	-	
C28	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	331776	7.4	22.6	-
C29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	8	.	.	.	.	19	.	.	212551	18.7	35.2	-
C30	.	.	20	0	.	.	.	.	.	12	21	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	763	21.1	48.1	63.4

である。いま各キャッシュ間で蓄積オブジェクトに相関が無いと仮定すると、HはICPがどのキャッシュにもヒットしない確率から、

$$\hat{H}_1 = 1 - \prod_j (1 - h_{i,j}) \quad (2)$$

で推定することができる。

また、当該キャッシュがいずれかのキャッシュに親接続を行なっている場合、ICPでヒットしなかったオブジェクトに関しては親キャッシュのいずれかにHTTP要求の後、その親キャッシュから再びICPメッセージが発行されることになる。親キャッシュの選択が無作為に行なわれる場合、2段目のICP要求まで考慮した場合のヒット率は

$$\hat{H}_2 = 1 - (1 - \hat{H}_1) \frac{1}{N} \sum_{k=1}^N (1 - \hat{H}'_{1,k}) \quad (3)$$

で推定される。ただし、Nは接続している親キャッシュの数であり、 $\hat{H}'_{1,k}$ は1段目のICP接続で利用したキャッシュを除外した残りのキャッシュへの親キャッシュのICPヒット率である。

表2に登録キャッシュ間の $h_{i,j}$ および各キャッシュのH,  $\hat{H}_1$ ,  $\hat{H}_2$ を示す。C07, C30の例からICP接続数が少ない場合でも親キャッシュを適切に選択することによって飛躍的なICPヒット率の上昇が見込めることを示唆している。

これらの情報は定期的 (現在, 1時間毎) に作成されWWW (計画のWWWサーバ[1]) を介して各キャッシュ管理者に提供されている。

#### 4 むすび

本稿では、準公開キャッシュ間の相互接続を促進する計画において、キャッシュ接続選択の際に参照する情報の作成について述べた。

#### 文献

- [1] 大家, “WWW マルチプロキシ計画 Version3.0 — JP-wide object cache project”, URL: <http://www.orions.ad.jp/project/multiproxy3/> (1997-7) .
- [2] 大家, “準公開 Squid 相互接続支援システムの運用”, 分散システム運用技術情報処理研究会, DSM (1998-7) .