

# 主 論 文 要 旨

報告番号	㊦ 乙 第 号	氏 名	浅原 理人
主論文題目： Flash Crowd による影響を軽減する複製サーバの配置先決定手法に関する研究			
<p>(内容の要旨)</p> <p>インターネットの利用者数は年々増加傾向にあり、コンテンツの配信にかかる負荷が増加し続けている。それに伴って、安定したサービスの提供を阻害する Flash Crowd と呼ばれる現象がインターネット上で発生するようになった。Flash Crowd が発生すると、インターネットサーバのアクセス数が数分のうちに通常時の数倍から数十倍に達し、サーバは過負荷状態になる。Flash Crowd の発生要因は多岐に渡るため、Flash Crowd の発生を予測することは困難である。</p> <p>大量のクライアントアクセスを処理する一般的な方法として、複製サーバの配置による負荷分散が知られている。しかし、Flash Crowd 対策用に複製サーバを常時稼働させておくためには、Flash Crowd の規模をあらかじめ予測しておかなければならない。また、Flash Crowd の生じていない平常時には余剰サーバが多くなり、サーバの運用コストが増大するという問題がある。</p> <p>本論文では、Flash Crowd による影響を軽減する複製サーバの配置先決定手法 ExaPeer Server Reposition (EPSR) を提案する。EPSR はあらかじめ世界中に設置された共有ホスティングマシン群で動作する。EPSR はこのホスティングマシン群に対して、サービスの需要に応じて複製サーバの配置先を動的に決定する機能を提供する。これにより、Flash Crowd の規模の予測を必要とせずに Flash Crowd への対策が行える。また、需要変動に応じた動的な複製サーバの配置ができるため、余剰サーバの運用コストが生じない。</p> <p>Flash Crowd の影響を軽減するように複製サーバの配置先を決定するには、次の3つの課題がある。第一に、Flash Crowd によるアクセス増加がピークに達するまでの数分以内に配置先を決定する必要がある。第二に、通信負荷の分散とサービス遅延の低減のために、配置先とクライアントとの通信遅延が小さくなるように配置先を決定する必要がある。第三に、複製サーバによってホスティングマシンが過負荷にならないように、複製サーバの配置先を決定する必要がある。既存手法の多くは、Globule のような、全サーバのアクセスログといった収集に時間のかかる大域情報に依存するものや、FCAN や CoralCDN のような、通信遅延やマシンの負荷および性能を考慮せずに配置先を決定するものである。そのため、Flash Crowd 対策に適用するには限界がある。</p> <p>EPSR は複製サーバの配置に関する3つの課題の全てを解決する。EPSR は大域情報を必要としない Peer-to-Peer 技術を用いることで急激な需要の増加に対応する。個々のマシンは自律的に需要の変動を検出し、必要に応じて複製サーバの配置先候補となる。個々のマシンでの需要変動の検出と需要の発生地域の特定を実現するために、EPSR は物理ネットワークのトポロジを考慮したオーバレイネットワークを、ネットワーク座標系と分散ハッシュテーブルを組み合わせて構築する。個々のマシンは、オーバレイ上で隣接するマシンとのみメッセージを交換することでクライアントのアクセス経路を分析し、サービスへの需要が増加もしくは減少している地域を特定する。マシンはアクセス経路の分析から複製サーバの実行による効果を見積もり、配置先の候補となるべきかを判定する。この際、個々のマシンは複製サーバの負荷とマシンの許容負荷を元に、マシンが過負荷にならないように隣接するマシンから追加の配置先候補を独立して選択する。</p> <p>シミュレーションを行い、EPSR が Flash Crowd に対して有効であることを確認した。シミュレーションでは、約3000台のホスティングマシンに対して EPSR が25秒で Flash Crowd の検出と発生地域の特定を行い、配置先の選択を開始することを確認した。また複製サーバの配置先数が300秒で安定することを確認した。その際、EPSR は他方式と比較してよりクライアントとの通信遅延が小さくなるようなホスティングマシンを配置先として選出した。またシミュレーション結果は、EPSR が他方式よりも少ない複製サーバ数で、どのホスティングマシンも過負荷にならないように配置先が決定できることを示した。</p>			

# SUMMARY OF Ph.D. DISSERTATION

School Science for Open and Environmental Systems	Student Identification Number	SURNAME, First name ASAHARA, Masato
<b>Title</b> A Study on the Mechanism for Finding Replica Server Locations for Alleviating Flash Crowds		
<b>Abstract</b> <p>As the Internet population is increasing yearly, the burden of content distribution on the Internet has been augmenting. As the Internet is becoming the widely used media, flash crowd often occurs on the Internet and prevents an Internet server from serving a reliable service. With flash crowds, the load of an Internet server reaches to dozens of times the load of a typical day in a few minutes, and then the server becomes overloaded. It is quite difficult to predict flash crowds since there are various factors of flash crowds.</p> <p>Positioning replica servers is well known as a popular load balancing strategy to provide a reliable service for a massive number of clients. Unfortunately, positioning over-provisioned replica servers to alleviate flash crowds has two drawbacks; one is the necessity for predicting the scale of flash crowds and the other is the expense of maintaining the over-provisioned replica servers.</p> <p>This dissertation proposes a novel mechanism called ExaPeer Server Reposition (EPSR) for finding replica server locations for alleviating flash crowds. EPSR runs on a pool of hundreds or thousands of shared hosting machines all over the world. EPSR provides the hosting machines with the function that determines the number and the location of replica servers on the basis of demand fluctuations. EPSR enables service providers to provide their services in flash crowds without predicting the scale of flash crowds. Also, EPSR needs no over-provisioned replica servers because it enables to dynamically reposition replica servers in short-term demand fluctuations as well as long-term ones.</p> <p>Finding replica server locations has three challenges of alleviating flash crowds. First, we must find replica server locations before flash crowd reaches its peak, e.g., in a few minutes. Second, a machine close to clients on the network, i.e., the network latency between the machine and the clients is short, should be a candidate for running a replica server to distribute network load and suppress service delay. Finally, a machine with enough capacity to run a replica server must be a candidate to prevent a machine from becoming overloaded. Most of existing mechanisms have the limitation of applying to finding replica server locations for alleviating flash crowds. This is because some of them like Globule rely on collecting global information that requires a long period of time such as access logs of all the servers or some of them like FCAN and CoralCDN determine replica server locations without considering the network latency, the capacity of a machine or the load of a machine.</p> <p>EPSR addresses the three challenges. EPSR takes a P2P-based approach to quickly respond to rapidly increasing demands; it does not rely on global information. Each machine autonomously detects demand fluctuations and becomes a candidate if necessary. To enable a machine to detect demand fluctuations and locate a high-demand area, EPSR builds a topology-aware overlay network by combining network coordinate systems and distributed hash tables (DHTs). Each machine analyzes access paths from clients by exchanging messages with only adjacent machines on the overlay and locates the areas in which the demand for a service is increasing or decreasing. It estimates the benefit of running a replica server from this access path analysis. Based on the load of a replica server and machine's capacity, each machine independently chooses more candidates from adjacent machines so that the machine does not become overloaded.</p> <p>Simulation results demonstrate that EPSR is suitable for alleviating flash crowds. In the simulation with 3,000 over hosting machines, EPSR detected demand increase, located high-demand areas and began selecting candidate machines for replica servers within 25 seconds. And then, it correctly estimated the number of required replica servers and the number of candidate machines had been stabilized within 300 seconds. With EPSR, candidate machines were closer to clients than that with other methods. The results also demonstrate that EPSR selected candidate machines as no machines became overloaded even with fewer candidates than that with other methods.</p>		