Manageability, Availability and Performance in Porcupine: A Highly Scalable, Cluster-based Mail Service

YASUSHI SAITO, BRIAN N. BERSHAD, and HENRY M. LEVY
University of Washington

This paper describes the motivation, design, and performance of Porcupine, a scalable mail server. The goal of Porcupine is to provide a highly available and scalable electronic mail service using a large cluster of commodity PCs. We designed Porcupine to be easy to manage by emphasizing dynamic load balancing, automatic configuration, and graceful degradation in the presence of failures. Key to the system's manageability, availability, and performance is that sessions, data, and underlying services are distributed homogeneously and dynamically across nodes in a cluster.

Categories and Subject Descriptors: C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Distributed applications*; C.4 [**Performance of Systems**]: Reliability, Availability, and Serviceability; C.5.5 [**Computer System Implementation**]: Servers; D.4.5 [**Operating Systems**]: Reliability—*Fault-tolerance*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Distributed systems*; H.4.3 [**Information Storage and Retrieval**]: Communications Applications—*Electronic mail*

General Terms: Algorithms,Performance,Management,Reliability

Additional Key Words and Phrases: Distributed Systems, E-mail, Cluster, Group membership protocol, Replication, Load balancing

1. INTRODUCTION

The growth of the Internet has led to the need for highly scalable and highly available services. This paper describes the Porcupine scalable electronic mail service. Porcupine achieves scalability by clustering many small machines (PCs), enabling them to work together in an efficient manner. In this section, we describe system requirements for Porcupine and relate the rationale for choosing a mail application as our target.

1.1 System Requirements

Porcupine defines scalability in terms of three essential system aspects: manageability, availability, and performance. Requirements for each follow:

(1) **Manageability requirements.** Although a system may be physically large, it should be easy to manage. In particular, the system must *self-configure* with respect to load and data distribution and *self-heal* with respect to failure and recovery. A system manager can simply add more machines or disks to improve throughput and replace them

This work is supported by DARPA Grant F30602-97-2-0226 and by National Science Foundation Grant # EIA-9870740.
An earlier version of this article appeared at the 17th ACM Symposium on Operating Systems Principles (SOSP), Kiawah Island Resort, SC, Dec., 1999.
Authors' address: MBOX 352350, Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195; email: {yasushi,bershad,levy}@cs.washington.edu.
The Porcupine project web page is at `http://porcupine.cs.washington.edu`.

when they break. Over time, a system's nodes will perform at differing capacities, but these differences should be masked (and managed) by the system.

(2) **Availability requirements.** With so many nodes, it is likely that some will be down at any given time. Despite component failures, the system should deliver good service to *all* of its users at all times. In practice, the failure of one or more nodes may prevent some users from accessing some of their mail. However, we strive to avoid failure modes in which whole groups of users find themselves without any mail service for even a short period.

(3) **Performance requirements.** Porcupine's single-node performance should be competitive with other single-node systems; its aggregate performance should scale linearly with the number of nodes in the system. For Porcupine, we target a system that scales to hundreds of machines, which is sufficient to service a few billion mail messages per day with today's commodity PC hardware and system area networks.

Porcupine meets these requirements uniquely. The key principle that permeates the design of Porcupine is *functional homogeneity*. That is, any node can execute part or all of any transaction, e.g., for the delivery or retrieval of mail. Based on this principle, Porcupine uses three techniques to meet our scalability goals. First, every transaction is *dynamically scheduled* to ensure that work is uniformly distributed across all nodes in the cluster. Second, the system *automatically reconfigures* whenever nodes are added or removed even transiently. Third, system and user data are automatically *replicated* across a number of nodes to ensure availability.

Figure 1 shows the relationships among our goals and key features or techniques used in the system. For example, dynamic scheduling and automatic reconfiguration make the system manageable, since changes to the size or the quality of machines, user population, and workload are handled automatically. Similarly, automatic reconfiguration and replication improve availability by making email messages, user profiles, and other auxiliary data structures survive failures.

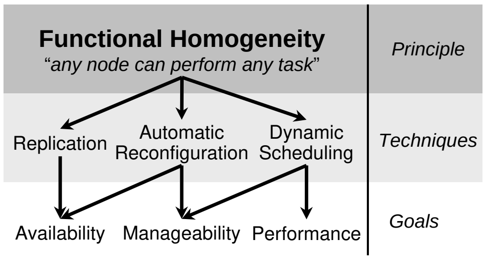


Fig. 1. The primary goal of Porcupine is scalability defined in terms of manageability, availability, and performance requirements. In turn, these requirements are met through combinations of the three key techniques shown above.

Today, Porcupine runs on a cluster of thirty PCs connected by a high-speed network, although we show that it is designed to scale well beyond that. Performance is linear with respect to the number of nodes in the cluster. The system adapts automatically to changes in workload, node capacity, and node availability. Data is available despite the presence of failures.

1.2 Rationale for a Mail Application

Although Porcupine is a mail system, its underlying services and architecture are appropriate for other systems in which data is frequently written and good performance, availability, and manageability at high volume are demanded. For example, Usenet news, community bulletin boards, and large-scale calendar services are good candidates for deployment using Porcupine. Indeed, we have configured Porcupine to act as a web server and a Usenet news node. In this paper, however, we focus on the system's use as a large scale electronic mail server.

We chose a mail application for several reasons. First is need: large-scale commercial services now handle more than ten million messages per day. Anticipating continued growth, our goal with Porcupine is to handle billions of messages per day on a PC-based cluster. Second, email presents a more challenging application than that served by conventional web servers, which have been shown to be quite scalable. In particular, the workload for electronic mail is *write intensive* and most of the Web scaling techniques, such as stateless transformation [Fox et al. 1997] and caching [Chankhunthod et al. 1996; Pai et al. 1998], become useless for write-intensive workloads. Finally, consistency requirements for mail, compared to those for a distributed file or database system, are weak enough to encourage the use of replication techniques that are both efficient and highly available.

1.3 Organization of the Paper

The remainder of this paper describes Porcupine's architecture, implementation, and performance. Section 2 presents an overview of the system's architecture and compares our architecture with alternatives. Section 3 describes how the system adapts to changes in configuration automatically, while Section 4 presents Porcupine's approach to availability. In Section 5 we describe the system's scalable approach to fine-grained load balancing. Section 6 evaluates the performance of the Porcupine prototype on our 30-node cluster. Section 7 discusses some of the system's scalability limitations and areas for future work. In Section 8, we discuss related work, and we draw conclusions in Section 9.

2. SYSTEM ARCHITECTURE OVERVIEW

Porcupine is a cluster–based, Internet mail service that supports the SMTP protocol [Postel 1982] for sending and receiving messages across the Internet. Users retrieve their messages using any mail user agent that supports either the POP or IMAP retrieval protocols [Myers and Rose 1996; Crispin 1996].

A key aspect of Porcupine is its *functional homogeneity*: any node can perform any function. This greatly simplifies system configuration: the system's capacity grows and shrinks with the number and aggregate power of the nodes, not with how they are logically configured. Consequently, there is no need for a system administrator to make specific service or data placement decisions. This attribute is key to the system's manageability.

Functional homogeneity ensures that a service is always available, but it offers no guarantees about the data that the service may be managing. *Replicated state* serves this purpose. There are two kinds of replicated state that Porcupine must manage: hard state and soft state. *Hard state* consists of information that cannot be lost and therefore must be maintained in stable storage. For example, an email message and a user's password are hard state. Porcupine replicates hard state on multiple nodes to increase availability and to survive failures. *Soft state* consists of information that, if lost, can be reconstructed from existing hard state. For example, the list of nodes containing mail for a particular user is soft state, because it can be reconstructed by a distributed disk scan. Most soft state is maintained on only one node at a given instant, and is reconstructed from hard state after failure. The exception is when directories that name and locate other state are themselves soft state. Such directories are replicated on every node to improve performance.

This approach minimizes persistent store updates, message traffic, and consistency management overhead. The disadvantage is that soft state may need to be reconstructed from distributed persistent hard state after a failure. Our design seeks to ensure that these reconstruction costs are low and can scale with the size of the system. In Section 6, we demonstrate the validity of this design by showing that reconstruction has nominal overhead.

The following subsections describe Porcupine's data structures and their management.

2.1 Key Data Structures

Porcupine consists of a collection of data structures and a set of internal operations provided by managers running on every node. The key data structures found in Porcupine are:

Mailbox fragment. The collection of mail messages stored for a given user at any given node is called a *mailbox fragment*; the fragment is also the unit of mail replication. A Porcupine mailbox is therefore a logical entity consisting of a single user's mailbox fragments distributed and replicated across a number of nodes. There is no single mailbox structure containing all of a user's mail. A mailbox fragment is hard state.

Mail map. This list describes the nodes containing mailbox fragments for a given user. The mail map is soft state. For the sake of brevity, we pretend that each user has only one mailbox throughout this paper; in fact, Porcupine supports multiple mailboxes per user, and the mail map actually maps a pair $\langle$user, mailbox$\rangle$ to a set of nodes.

User profile database. This database describes Porcupine's client population, i.e., it contains user names, passwords, etc. It is persistent, changes infrequently for a given user, and is partitioned and replicated across nodes. The user profile database is hard state.

User profile soft state. Porcupine separates the storage and the management of user profile, which is distributed dynamically to improve performance. Each Porcupine node uniquely stores a soft-state copy of a subset of the profile database entries. Accesses and updates to a profile database entry begin at the node holding the soft-state copy of that entry. This data structure is soft state.

User map. The user map is a table that maps the hash value of each user name to a node currently responsible for managing that user's profile soft state and mail map. The user map is soft state and is replicated on each node.

Cluster membership list. Each node maintains its own view of the set of nodes currently functioning as part of the Porcupine cluster. Most of the time, all nodes perceive the same membership, although a node's arrival or departure may cause short-term inconsistencies as the system establishes the new membership. During network partition, inconsistencies may last for a long time. Various system data and services, such as the user map and load balancer, automatically respond to changes in the cluster membership list. The cluster membership list is soft state and is replicated on each node.

2.2 Data Structure Managers

The preceding data structures are distributed and maintained *on each node* by several essential managers shown in Figure 2. The *user manager* manages soft state including user profile soft state and mail maps. By spreading the responsibility for servicing accesses to the user profile database across all nodes in the system, larger user populations can be supported simply by adding more machines.

Two managers, the *mailbox manager* and the *user database manager*, maintain persistent storage and enable remote access to mailbox fragments and user profiles.

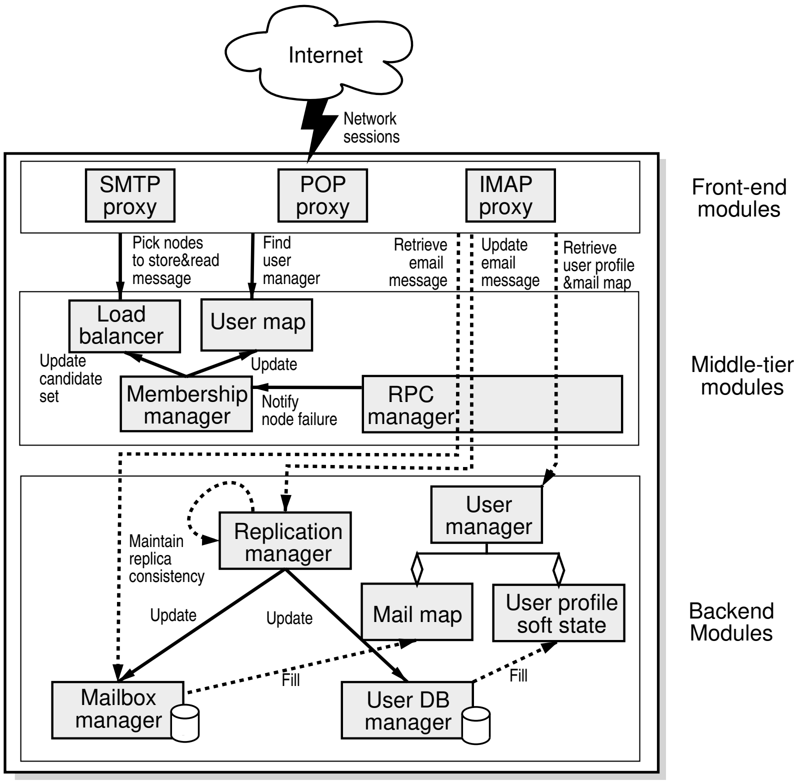


Fig. 2. Each node in Porcupine runs the same set of modules shown in this picture. A solid arrow shows that a module calls another module within the node, and a dotted arrow shows that a module calls another module in a remote node using the RPC module.

The *replication manager* on each node ensures the consistency of replicated objects stored in that node's local persistent storage.

The *membership manager* on each node maintains that node's view of the overall cluster state. It tracks which nodes are up or down and the contents of the user map. It also participates in a membership protocol to track that state. The *load balancer* on each node maintains the load and disk usage of other nodes and picks the best set of nodes to store or read messages. The *RPC manager* supports remote inter-module communication.

On top of these managers, each node runs a *delivery proxy* to handle incoming SMTP requests and *retrieval proxies* to handle POP and IMAP requests.

The Porcupine architecture leads to a rich distribution of information in which mail storage is decoupled from user management. For example, Figure 3 shows a sample Porcupine configuration consisting of two nodes and three users. For simplicity, messages are not shown as replicated. The user manager on node A maintains Alice's and Bob's soft state, which consists of their user profile database entries and their mail maps. Similarly, the user manager on node B maintains Chuck's soft state.

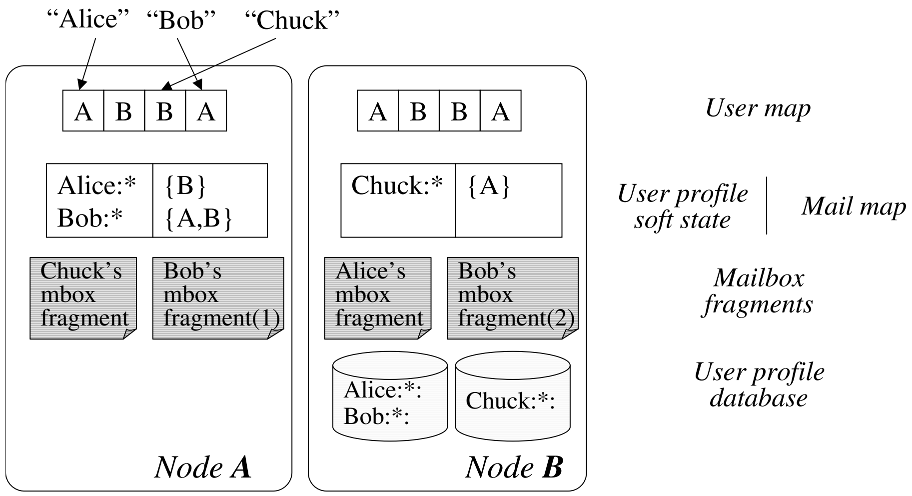


Fig. 3. This picture shows how a two-node cluster might distribute email messages. The user map (shown as four-entry wide in the picture, but 256-entry wide in the implementation) is replicated on each node. For example, a node learns that Bob is managed by node A, because the hash value of the string "Bob" is 3, and the entry number three in the user map is A. To read Bob's messages, the mail client consults the user manager on A to obtain Bob's profile (password is shown as '*') and mail map ($\{A, B\}$) and contacts each node in the mail map to read Bob's messages.

2.3 A Mail Transaction in Progress

In failure-free operation, mail delivery and retrieval work as follows.

2.3.1 *Mail Delivery.* Figure 4 shows the flow of control during mail delivery. An external mail transfer agent (MTA) delivers a message to a user hosted on a Porcupine cluster by discovering the IP address of any Porcupine cluster node using the Internet's Domain Name Service [Brisco 1995] (step 1). Because any function can execute on any node, there is no need for special front-end request routers [Cisco Systems 1999; Foundry Networks 1999], although nothing in the system prevents their use.

To initiate mail delivery, the MTA uses SMTP to connect to the designated Porcupine node, which acts as a delivery proxy (step 2). The proxy's job is to store the message on disk. To do this, it applies the hash function on the recipient's name, looks up the user map, and learns the name of the recipient's user manager (step 3). It then retrieves the mail map from the user manager (steps 4 and 5) and asks the load balancing service to choose the best node from that list. If the list is empty or all choices are poor (for example, overloaded or out of disk space), the proxy is free to select any other node (step 6). The proxy then forwards the message to the chosen node's mailbox manager for storage (step 7). The storing node ensures that its participation is reflected in the user's mail map (step 8). If the message is to be replicated (based on information in the user's profile), the proxy selects multiple nodes on which to store the message.

2.3.2 *Mail Retrieval.* An external mail user agent (MUA) retrieves messages for a user whose mail is stored on a Porcupine cluster using either the POP or IMAP transfer protocols. The MUA contacts any node in the cluster to initiate the retrieval. The contacted node, acting as a proxy, authenticates the request through the user manager for the client

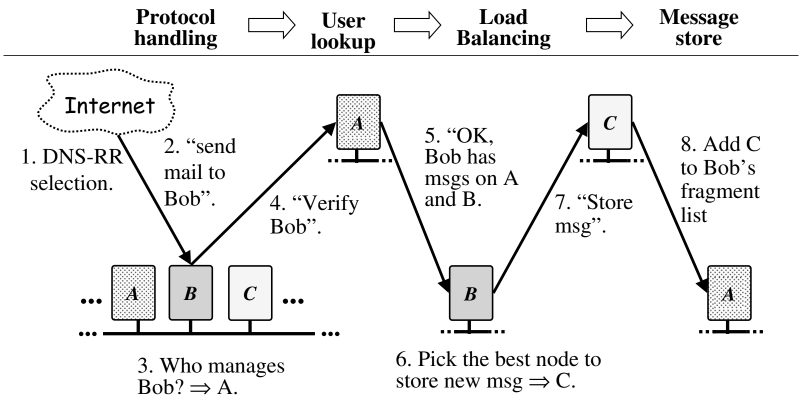


Fig. 4. This picture shows how an external mail transfer agent (MTA) delivers a message to Bob. The MTA picks B, through DNS-RR, as the SMTP session partner (step 1 and 2). B obtains Bob’s mailbox fragment from A (steps 3 through 5) and determines that C is the best node to store the message (step 6). C updates Bob’s mailbox fragment after storing the message (steps 7 and 8).

and discovers the mail map. It then contacts the mailbox manager at each node storing the user’s mail to request mail digest information, which it returns to the MUA. Then, for each message requested, the proxy fetches the message from the appropriate node or nodes. If the MUA deletes a message, the proxy forwards the deletion request to the appropriate node or nodes. When the last message for a user has been removed from a node, that node removes itself from the user’s mail map.

2.4 Advantages and Tradeoffs

By decoupling the delivery and retrieval agents from the storage services and user manager in this way, the system can balance mail delivery tasks dynamically; any node can store mail for any user, and no single node is permanently responsible for a user’s mail or soft profile information. A user’s mail can be replicated on an arbitrary set of nodes, independent of the replication factor for other users. If a user manager goes down, another will take over for that manager’s users. Another advantage is that the system becomes extremely fault tolerant by always being able to deliver or retrieve mail for a user, even when nodes storing the user’s existing mail are unavailable. The final advantage is that the system is able to react to configuration without human intervention. Newly added nodes will automatically receive their share of mail-session and storage-management tasks. Crashed or retired node will be excluded from the membership list and mail maps automatically, leaving no residual information on other nodes.

The system architecture reveals a key tension that must be addressed in the implementation. Specifically, while a user’s mail may be distributed across a large number of machines, doing so complicates both delivery and retrieval. On delivery, each time a user’s mail is stored on a node not already containing mail for that user, the user’s mail map (a potentially remote data structure) must be updated. On retrieval, aggregate load increases somewhat with the number of nodes storing the retrieving user’s mail. Consequently, it is beneficial to limit the spread of a user’s mail, widening it primarily to deal with load imbalances and failure. In this way, the system behaves (and performs) like a statically partitioned system when there are no failures and load is well balanced, but like a dynamically partitioned system otherwise. Section 5 discusses this tradeoff in more detail.

2.5 Alternative Approaches

Existing large-scale mail systems assign users and their data statically to specific machines [Christenson et al. 1997; Deroest 1996]. A front-end traffic manager directs an external client's request to the appropriate node. We believe that such statically distributed, write-oriented services scale poorly. In particular, as the user base grows, so does service demand, which can be met only by adding more machines. Unfortunately, each new machine must be configured to handle a subset of the users, requiring that users and their data migrate from older machines. As more machines are added, the likelihood that at least one of them is inoperable grows, diminishing availability for users with data on the inoperable machines. In addition, users whose accounts are on slower machines tend to receive worse service than those on faster machines. Finally, a statically distributed system is susceptible to overload when traffic is distributed non-uniformly across the user base. To date, systems relying on static distribution have worked for two reasons. First, service organizations have been willing to substantially overcommit computing capacity to mitigate short-term load imbalances. Second, organizations have been willing to employ people to reconfigure the system manually in order to balance load over the long term. Because the degree of overcapacity determines where short-term gives way to long-term, static systems have been costly in terms of hardware, people, or both. For small static systems, these costs have not been substantial; for example, doubling the size of a small but manageable system may yield a system that is also small and manageable. However, once the number of machines becomes large (i.e., on the order of a few dozen), disparate (i.e., fast/slow machines, fast/slow disks, large/small disks), and continually increasing, this gross overcapacity becomes unacceptably expensive in terms of hardware and people.

An alternative approach is to adopt a typical Web server architecture: use a distributed file system to store all hard state and run off-the-shelf software on a large number of stateless, front-end nodes that serve clients [Fox et al. 1997; Pai et al. 1998]. This approach has been successful in services that deliver mostly read-only data, such as Web servers and search engines, because the front-end nodes can take significant load off the file system by utilizing file caches. Write-intensive services such as email, however, exhibit very low access locality that makes caching nearly useless, and using this approach in email requires the file system itself to be highly scalable under changing workload and system configuration. Such file systems do exist (e.g., xFS [Anderson et al. 1995] and Frangipani [Thekkath et al. 1997]), but they are still in an early research stage due to their sheer complexity. Moreover, even if they were available now, their manageability and availability would not match Porcupine's because the file systems offer generic, single-copy semantics and sacrifice availability along the way. For example, they tolerate only a limited number of node failures, beyond which the entire system stops, and they stop functioning when the network is partitioned. Porcupine, on the other hand, tolerates any number of node failures and continues to serve users after network partition by relaxing the data consistency guarantees.

Another approach is to build an email system on top of a cluster-based operating system that supports membership agreement, distributed locking, and resource fail-over (e.g., [Kronenberg et al. 1986; Vogels et al. 1998; Sun Microsystems 1999; IBM 1998]). While this solution simplifies the architecture of the software, it tends to cost more than previous solutions because these systems run only on proprietary hardware. They also have limited scalability, only up to tens of nodes. More importantly, the primary means of fault tolerance for such systems is shared disks, which statically tie a node to specific data items and create the same manageability and availability problems present in the first approach, albeit to a lesser degree.

Finally, the most obvious solution is to use a large monolithic server with reliable storage (e.g., RAID [Chen et al. 1994]). While this approach is the simplest in terms of architecture and administration, it is rarely employed by Internet services for two main reasons. First, a large server machine is far more expensive than a set of small machines with the same aggregate performance. Moreover, we can scale a single server only up to a certain limit, beyond which we must scrap the machine and buy a faster model. Notice, however, that the problem of making a single node fast and available is orthogonal to the problem of making a cluster fast and available. Porcupine solves only the latter problem, and it is perfectly reasonable to build a Porcupine cluster using large-scale server nodes for those applications in which a single node cannot handle the entire workload.

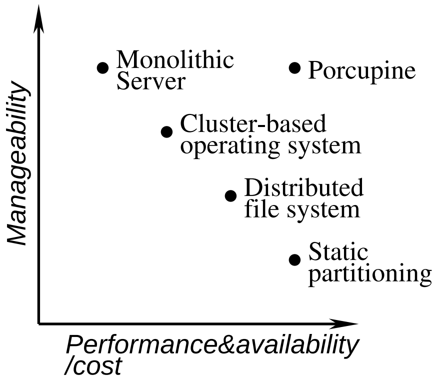


Fig. 5. A schematic view of how different architectures trade off cost, performance, availability and manageability. Porcupine is an architecture that is available, manageable, and cheap at the same time, whereas other solutions need to sacrifice either cost or manageability.

Figure 5 summarizes the cost and manageability trade-offs for these four solutions. Porcupine seeks to provide a system structure that performs well as it scales, adjusts automatically to changes in configuration and load, and is easy to manage. Our vision is that a single system administrator can be responsible for the hardware that supports the mail requirements of one hundred million users processing a billion messages per day. When the system begins to run out of capacity, that administrator can improve performance for all users simply by adding machines or even disks to the system. Lastly, the administrator can, without inconveniencing users, attend to the failure of machines, replacing them with the same urgency with which one replaces light bulbs.

3. SELF MANAGEMENT

Porcupine must deal automatically with diverse changes, including node failure, node recovery, node addition, and network failure. In addition, change can come in bursts, creating long periods of instability, imbalance and unavailability. It is a goal of Porcupine to manage change automatically in order to provide good service even during periods of system flux. The following sections describe the Porcupine services that detect and respond to configuration changes.

3.1 Membership Services

Porcupine's cluster membership service provides the basic mechanism for tolerating changes. It maintains the current membership set, detects node failures and recoveries, notifies other services of changes in the system's membership, and distributes new system state. We assume a symmetric and transitive network in steady state, so that nodes eventually converge on a consistent membership set provided that no new failure occurs for a sufficiently long period (i.e., a few seconds).

The cluster membership service uses a variant of the Three Round Membership Protocol (TRM) [Christian and Schmuck 1995] to detect membership changes. In TRM, the first round begins when any node detects a change in the configuration and becomes the coordinator. The coordinator broadcasts a "new group" message together with its Lamport clock [Lamport 1978], which acts as a proposed epoch ID to identify a particular membership incarnation uniquely. If two or more nodes attempt to become a coordinator at the same time, the one proposing the largest epoch ID wins.

In the second round, all nodes that receive the "new group" message reply to the coordinator with the proposed epoch ID. After a timeout period, the coordinator defines the new membership to be those nodes from which it received a reply. In the third round, the coordinator broadcasts the new membership and epoch ID to all nodes.

Once membership has been established, the coordinator periodically broadcasts probe packets over the network. Probing facilitates the merging of partitions; when a coordinator receives a probe packet from a node not in its current membership list, it initiates the TRM protocol. A newly booted node acts as the coordinator for a group in which it is the only member. Its probe packets are sufficient to notify others in the network that it has recovered.

There are several ways in which one node may discover the failure of another. The first is through a timeout that occurs normally during part of a remote operation. In addition, nodes within a membership set periodically "ping" their next highest neighbor in IP address order, with the largest IP address pinging the smallest. If the ping is not responded to after several attempts, the pinging node becomes the coordinator and initiates the TRM protocol.

3.2 User Map

The purpose of the user map is to distribute management responsibility evenly across live nodes in the cluster. Whenever membership services detect a configuration change, the system must reassign that management responsibility. Therefore, like the membership list, the user map is replicated across all nodes and is recomputed during each membership change as a side effect of the TRM protocol.

After the second round, the coordinator computes a new user map by removing the failed nodes from the current map and uniformly redistributing available nodes across the user map's hash buckets (the user map has many buckets, so a node typically is assigned to more than one bucket). The coordinator minimizes changes to the user map to simplify reconstruction of other soft state, described in the next section.

Each entry in the user map is associated with an epoch ID that shows when the bucket management responsibility is first assigned to a node. In the first phase of the TRM, each node piggybacks on the reply packet the index and the associated epoch IDs of all the user map entries the node manages. For each bucket with a changed assignment, the coordinator assigns the current epoch ID to the entry. On the other hand, for a bucket whose assignment remains unchanged, the coordinator reuses the epoch ID returned by the participant node. The epoch IDs in the user map are used by nodes to determine which entries in the user map have changed.

Figure 6 shows an example of a user map reconfiguration. In this example, node C crashes. A new membership is computed on node A, but the packet containing the new membership fails to reach node B. Next, C recovers, and B receives a new membership

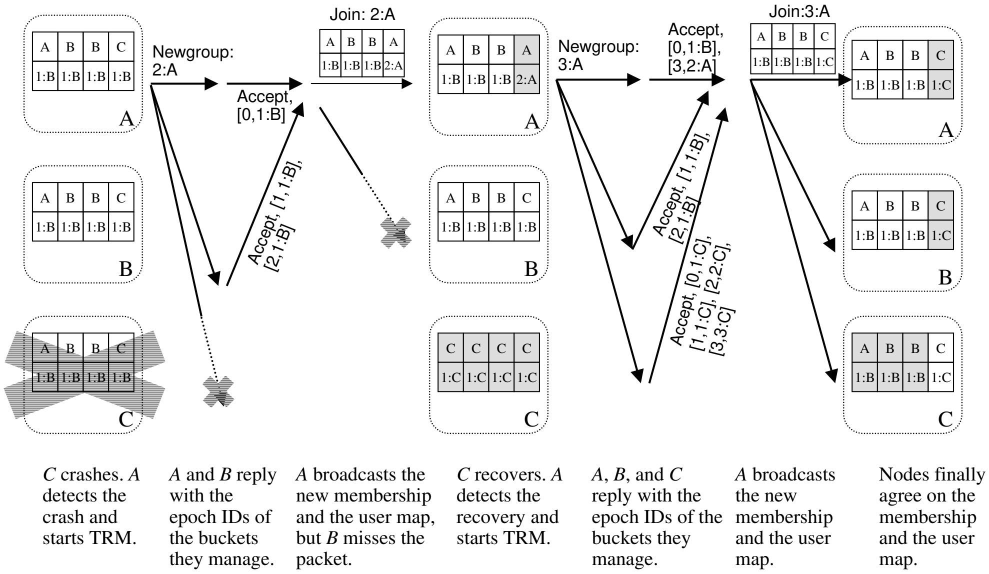


Fig. 6. Example of membership reconfiguration. Arrows show messages exchanged among the nodes. Upper boxes in each user map show the assignments of buckets to nodes, and lower boxes show the epoch IDs of buckets. In this example, the node C crashes and then recovers. The node B fails to receive the membership renewal after C's crash. Shaded area in user maps show the entries that nodes recognize as changed.

and a new user map that are identical to the old one (on B) except that the epoch ID for the bucket managed by C is renewed. Without epoch IDs in the user maps, B would be unable to detect that assignment for the last bucket of the user map has changed.

3.3 Soft State Reconstruction

Once the user map has been reconstructed, it is necessary to reconstruct the soft state at user managers with new user responsibilities. Specifically, this soft state is the user profile soft state and the mail map for each user. Essentially, every node pushes soft state corresponding to any of its hard state to new user managers responsible for that soft state.

Reconstruction is a two-step process, completely distributed, but unsynchronized. The first step occurs immediately after the third round of membership reconfiguration. Here, each node compares the previous and current user maps to identify any buckets having new assignments. A node considers a bucket assignment new if the bucket's previous epoch ID does not match the current epoch ID. Recall that the user map associates nodes with hash buckets, so the relevant soft state belonging on a node is that corresponding to those users who hash into the buckets assigned to the node.

Each node proceeds independently to the second step. Here, every node identifying a new bucket assignment sends the new manager of the bucket any soft state corresponding to the hard state for that bucket maintained on the sending node. First, the node locates any mailbox fragments belonging to users in the newly managed bucket and requests that the new manager include this node in those users' mail maps. Second, the node scans its portion of the stored user profile database and sends to the new manager all pertinent user profiles. As the user database is replicated, only the replica with the largest IP address among those functioning does the transfer. The hard state stored on every node is "bucketed" into directories so that it can be quickly reviewed and collected on each change to the corresponding bucket in the user map.

The cost of rebuilding soft state is intended to be constant per node in the long term, regardless of cluster size for the following reason. First, the cost of reconfiguration per node after a failure is roughly proportional to the total number of mailboxes to be discovered on the node, because the disk scan is by far the most expensive operation in the entire reconfiguration process. Second, the number of mailboxes to be discovered is determined by the number of reassignments to the user map, assuming that mailboxes are evenly distributed in each hash bucket. Third, the number of user map reassignments per single node crash or recovery is inversely proportional to cluster size, because each node manages $1/\mathit{cluster\text{-}size}$ of the user map. Consequently, the cost of reconfiguration per node per failure is *inversely proportional* to the cluster size. Finally, because the frequency of reconfiguration increases linearly with cluster size (assuming independent failures), the two factors cancel each other out, and the reconfiguration cost per node over time remains about the same regardless of the cluster size.

3.4 Effects of Configuration Changes on Mail Sessions

When a node fails, all SMTP, POP, and IMAP sessions hosted on the node abort—an unavoidable result given the difficulty of TCP session fail-over. Among them, the abortion of the SMTP sessions is transparent to the senders and the recipients except for delay and possible duplicate message delivery, because the remote MTAs retry delivery later. For the aborted POP and IMAP sessions, the users must reconnect to the cluster. An SMTP session that is hosted on another node and is about to store messages on the failed node reselects another node for storage until it succeeds. Thus, the node failure is masked from the remote server (and the sender) and the recipient of mail. A POP or IMAP session hosted on another node may report an error when it tries to read a message on the failed node, but the session itself continues running and is able to retrieve messages stored on other nodes.

The combination of the mail-map update mechanism (Section 2.3) and the automatic reconfiguration mechanism makes each user's mail-map consistent with respect to mailbox fragments locations without introducing the complexity of solutions based on atomic transactions [Gray and Reuter 1993]. We argue that sessions that are affected by node failures keep mail maps consistent by considering four different failure scenarios.

(1) A node fails just after a message is stored in a new mailbox fragment on its disk, but before the corresponding mail map is updated. This case causes no problem because this copy of the message becomes non-retrievable after the node failure. The replication service (Section 4) ensures that another copy of the message is still available.

(2) A node fails just after the last message in a mailbox fragment on its disk is deleted, but before the corresponding mail map is updated. Each node periodically scans the mail maps it manages and removes all "dangling" links to nodes not in the membership. The links will be restored when the failed nodes rejoin the cluster.

(3) A node stores a message in a new mailbox fragment on its disk, but the corresponding user manager node fails before the mail map is updated. The message will be discovered by the disk scan algorithm that runs after membership reconfiguration and will be added to the mail map on a new user manager node.

(4) A node deletes the last message in a mailbox fragment on its disk, but the corresponding user manager node fails before the mail map is updated. The same argument as

above is applied: a new user manager will receive the result of a disk scan that excludes the deleted mailbox.

3.5 Node Addition

Porcupine's automatic reconfiguration procedure makes it easy to add a new node to the system. A system administrator simply installs the Porcupine software on the node. When the software boots, it is noticed by the membership protocol and added to the cluster. Other nodes see the configuration change and upload soft state onto the new node. To make the host accessible outside of Porcupine, the administrator may need to update border naming and routing services. Occasionally, a background service rebalances replicated email messages and user database entries across the nodes in the cluster[1].

3.6 Summary

Porcupine's dynamic reconfiguration protocols ensure that the mail service is always available for any given user and allow the reconstruction and distribution of soft state with constant overhead. Client activities are affected minimally by a failure; after the ensuing reconfiguration process, the soft state is restored correctly regardless of ongoing client activities. The next section discusses the maintenance of hard state.

4. REPLICATION AND AVAILABILITY

This section describes object replication support in Porcupine. As in previous systems (e.g., [Fox et al. 1997]), Porcupine defines semantics tuned to its application requirements. This permits a careful balance between behavior and performance.

Porcupine replicates the user database and mailbox fragments to ensure their availability. Our replication service provides the same guarantees and behavior as the Internet's electronic-mail protocols. For example, Internet email may arrive out of order, on occasion more than once, and may sometimes reappear after being deleted. These anomalies are artifacts of the non-transactional nature of the Internet's mail protocols. Porcupine never loses electronic mail unless all nodes on which the mail has been replicated are irretrievably lost.

4.1 Replication Properties

The general unit of replication in Porcupine is the *object*, which is simply a named byte array that corresponds to a single mail message or the profile of a single user. A detailed view of Porcupine's replication strategy includes these five high-level properties:

Update anywhere. An update can be initiated at any replica. This improves availability, since updates need not await the revival of a primary. This strategy also eliminates the requirement that failure detection be precise, since there need not be agreement on which is the primary node.

Eventual consistency. During periods of failure, replicas may become inconsistent for short periods of time, but conflicts are eventually resolved. We recognize that single-copy consistency [Gray and Reuter 1993] is too strong a requirement for many Internet-based services, and that replica inconsistencies are tolerable as long as they are resolved eventually. This strategy improves availability, since accesses may occur during reconciliation or even during periods of network partitioning.

[1] In the current implementation, the rebalancer must be run manually.

Total update. An update to an object totally overwrites that object. Since email messages are rarely modified, this is a reasonable restriction that greatly simplifies update propagation and replica reconciliation, while also keeping overheads low.

Lock free. There are no distributed locks. This improves performance and availability and simplifies recovery.

Ordering by loosely synchronized clocks. The nodes in the cluster have loosely synchronized clocks [Mills 1992; 1994] that are used to order operations on replicated objects.

The update-anywhere attribute, combined with the fact that any Porcupine node may act as a delivery agent, means that incoming messages are never blocked (assuming at least one node remains functional). If the delivery agent crashes during delivery, the initiating host (which exists outside of Porcupine) can reconnect to another Porcupine node. If the candidate mailbox manager fails during delivery, the delivery agent will select another candidate until it succeeds. Both of these behaviors have the same potential anomalous outcome: if the failure occurs after the message has been written to stable storage but before any acknowledgement has been delivered, the end user may receive the same message more than once. We believe that this is a reasonable price to pay for service that is continually available.

The eventual-consistency attribute means that earlier updates to an object may "disappear" after all replica inconsistencies are reconciled. This behavior can be confusing, but we believe that this is more tolerable than alternatives that block access to data when replica contents are inconsistent. In practice, eventual consistency for email means that a message once deleted may temporarily reappear. This is visible only if users attempt to retrieve their mail during the temporary inconsistency, which is expected to last at most a few seconds.

The lock-free attribute means that multiple mail-reading agents, acting on behalf of the same user at the same time, may see inconsistent data temporarily. However, POP and IMAP protocols do not require a consistent outcome with multiple clients concurrently accessing the same user's mail.

The user profile database is replicated with the same mechanisms used for mail messages. Because of this, it is possible for a client to perceive an inconsistency in its (replicated) user database entry during node recovery. Operations are globally ordered by the loosely synchronized clocks; therefore, a sequence of updates to the user profile database will eventually converge to a consistent state. We assume that the maximum clock skew among nodes is less than the inter-arrival time of externally initiated, order-dependent operations, such as Create-User and Change-Password. In practice, clock skew is usually on the order of tens of microseconds[Mills 1994], whereas order-dependent operations are separated by networking latencies of at least a few milliseconds. Wall clocks, not Lamport clocks [Lamport 1978], are used to synchronize updates, because wall clocks can order events that are not logically related (e.g., an external agent contacting two nodes in the cluster serially).

We now describe the replication manager, email operations using replicas, and the details of updating replicated objects.

4.2 Replication Manager

A replication manager running on each host exchanges messages among nodes to ensure replication consistency. The manager is oblivious to the format of a replicated object and does not define a specific policy regarding when and where replicas are created. Thus, the replication manager exports two interfaces: one for the creation and deletion of objects, which is used by the higher level delivery and retrieval agents, and another for interfacing to the specific managers, which are responsible for maintaining on-disk data structures. The replication manager does not coordinate object reads; mail retrieval proxies are free to pick any replica and read them directly.

4.3 Sending and Retrieving Replicated Mail

When a user's mail is replicated, that user's mail map reflects the set of nodes on which each fragment is replicated. For example, if Alice has two fragments, one replicated on nodes A and B and another replicated on nodes B and C, the mail map for Alice records $\{\{A, B\}, \{B, C\}\}$. To retrieve mail, the retrieval agent contacts the least-loaded node for each replicated mailbox fragment to obtain the complete mailbox content for Alice.

To create a new replicated object (as would occur with the delivery of a mail message), an agent generates an object ID and the set of nodes on which the object is to be replicated. An *object ID* is simply an opaque, unique string. For example, mail messages have an object ID of the form ⟨*type*, *username*, *messageID*⟩, where *type* is the type of object (mail message), *username* is the recipient, *and messageID* is an unique mail identifier found in the mail header.

4.4 Updating Objects

Given an object ID and an intended replica set, a delivery or retrieval agent can initiate an update request to the object by sending an update message to any replica manager in the set. A delivery agent's update corresponds to the storing of a message. The retrieval agent's update corresponds to the deletion and modification of a message.

The receiving replica acts as the update coordinator and propagates updates to its peers. The replication manager on every node maintains a persistent update log, used to record updates to objects that have not yet been accepted by all replica peers maintaining that object. Each entry in the update log is the tuple ⟨*timestamp*, *objectID*, *target-nodes*, *remaining-nodes*⟩:

—*Timestamp* is the tuple ⟨*wallclock time*, *nodeID*⟩, where *wallclock time* is the time at which the update was accepted at the coordinator named by *nodeID*. Timestamp uniquely identifies and totally orders the update.

—*Target-nodes* is the set of nodes that should receive the update.

—*Remaining-nodes* is the set of peer nodes that have not yet acknowledged the update. Initially, *remaining-nodes* is equal to *target-nodes* and is pruned by the coordinator as acknowledgments arrive.

The coordinating replication manager works through the log, attempting to push updates to all the nodes found in the *remaining-nodes* field of an entry. Once contact has been made with a remaining node, the manager sends the replica's contents and the log entry to the peer. Since updates to objects are total, multiple pending updates to the same object on a peer are synchronized by discarding all but the one with the newest timestamp. If no pending update exists, or if the update request is the newest for an object, the peer adds the update to the log, modifies the replica, and sends an acknowledgement to the coordinator. Once the coordinator receives acknowledgements from all replica peers, it notifies all the participants of the update (including itself) of the completion of the update. Finally, the participants *retire* the completed update entry in their log (freeing that log space) after waiting for a sufficiently long period to filter out updates that arrive out of order. The wait period we use, 3 minutes in our prototype, is set to the sum of the maximum clock skew among nodes and maximum network-packet *lifetime*; i.e., the time long enough for most packets to reach the destination. This retirement mechanism is a variant of the at-most-once messaging algorithm using synchronized clocks [Liskov et al. 1991].

If the coordinator fails before responding to the initiating agent, the agent will select another coordinator. For updates to a new object, as is the case with a new mail message, the initiating agent will create another new object and select a new, possibly overlapping, set of replicas. This helps to ensure that the degree of replication remains high even in the presence of a failed coordinator. This design may deliver a message to the user more than once. This duplicate delivery problem, however, is already fairly common in the Internet today; it may happen after a network transmission failure or simply by a user pressing the "Send" button twice. Message duplication due to node failures is far rarer than duplication due to other causes.

The coordinators and participants force their update log to disk before applying the update to ensure that the replicas remain consistent. As an optimization, a replica receiving an update message for which it is the only remaining node need not force its log before applying the update. This is because the other replicas are already up to date, so the sole remaining node will never have to make them current for this update. In practice, this means that only the coordinator forces its log for two-way replication.

Should the coordinator fail after responding to the initiating target but before the update is applied to all replicas, any remaining replica can become the coordinator and bring others up to date. Multiple replicas can become the coordinator in such case, since replicas can discard duplicate updates by comparing timestamps.

In the absence of node failures, the update log remains relatively small for two reasons. First, the log never contains more than one update to the same object. Second, updates are propagated as quickly as they are logged and are deleted as soon as all replicas acknowledge. Timely propagation also narrows the window during which an inconsistency could be perceived.

When a node fails for a long time, the update logs of other nodes could grow indefinitely. To prevent this, updates remain in the update log for at most a week. If a node is restored after that time, it must reenter the Porcupine cluster as a "new" node, rather than as a recovering one. A node renews itself by deleting all of its hard state before rejoining the system.

4.5 Summary

Porcupine's replication scheme provides high availability through the use of consistency semantics that are weaker than strict single-copy consistency, but strong enough to service Internet clients using non-transactional protocols. Inconsistencies, when they occur, are short lived (the update propagation latency between functioning hosts) and, by Internet standards, unexceptional.

5. DYNAMIC LOAD BALANCING

Porcupine uses dynamic load balancing to distribute the workload across nodes in the cluster in order to maximize throughput. As mentioned, Porcupine clients select an initial contact node either to deliver or to retrieve mail. That contact node then uses the system's load-balancing services to select the "best" set of nodes for servicing the connection.

In developing the system's load balancer, we had several goals. First, it must be fine-grained, making good decisions at the granularity of message delivery. Second, it must support a heterogeneous cluster, since not all the nodes are of equivalent power. Third, it must be automatic and minimize the use of "magic constants," thresholds, or tuning parameters that needs to be manually adjusted as the system evolves. Fourth, with throughput as the primary goal, it needs to resolve the tension between load and affinity. Specifically, in order to best balance load, messages should be stored on idle nodes. However, it is less expensive to store (and retrieve) a message on nodes that already contain mail for the message's recipient. Such *affinity-based scheduling* reduces the amount of memory needed to store mail maps, increases the sequentiality of disk accesses, and decreases the number of inter-node RPCs required to read, write, or delete a message.

In Porcupine, delivery and retrieval proxies make load-balancing decisions. There is no centralized load-balancing node service; instead, each node keeps track of the load on other nodes and makes decisions independently.

Load information is collected in the same ways we collect liveness information (Section 3.1): (1) as a side-effect of RPC operations (i.e., each RPC request or reply packet contains the load information of the sender), and (2) through a virtual ring in which load information is aggregated in a message passed along the ring. The first approach gives a timely but possibly narrow view of the system's load. The second approach ensures that every node eventually discovers the load from every other node.

The load on a node has two components: a boolean, which indicates whether or not the disk is full, and an integer, which is the number of pending remote procedure calls that might require a disk access. A node with a full disk is always considered "very loaded" and is used only for operations that read or delete existing messages. After some experimentation, we found that it was best to exclude diskless operations from the load to keep it from becoming stale too quickly. Because disk operations are so slow, a node with many pending disk operations is likely to stay loaded for some time.

A delivery proxy that uses load information alone to select the best node(s) on which to store a message will tend to distribute a user's mailbox across many nodes. As a result, this broad distribution can actually reduce overall system throughput for the reasons mentioned earlier. Consequently, we define for each user a *spread*; the spread is a soft upper bound on the number of different nodes on which a given user's mail should be stored. The bound is soft to permit the delivery agent to violate the spread if one of the nodes storing a user's mail is not responding. When a mailbox consists of fewer fragments than its spread limit, the delivery proxy adds a random set of nodes on message arrival to make up a candidate set. Adding a random set of nodes helps the system avoid a "herd behavior" in which a herd of nodes all choose the same node that is idle at one moment and instantly overloading the node the next moment [Mitzenmacher 1998].

As shown in Section 6, the use of a spread-limiting load balancer has a substantial effect on system throughput even with a relatively narrow spread. The benefit is that a given user's mail will be found on relatively few nodes, but those nodes can change entirely each time the user retrieves and deletes mail from the server.

6. SYSTEM EVALUATION

This section presents measurements from the Porcupine prototype running synthetic workloads on a 30-node cluster. We characterize the system's scalability as a function of its size in terms of the three key requirements:

Performance. We show that the system performs well on a single node and scales linearly with additional nodes. We also show that the system outperforms a statically partitioned configuration consisting of a cluster of standard SMTP and POP servers with fixed user mapping.

Availability. We demonstrate that replication and reconfiguration have low cost.

Manageability. We show that the system responds automatically and rapidly to node failure and recovery, while continuing to provide good performance. We also show that incremental hardware improvements can automatically result in system-wide performance improvements. Lastly, we show that automatic dynamic load balancing efficiently handles highly skewed workloads.

6.1 Platform and Workload

The Porcupine system runs on Linux-based PCs with all system services on a node executing as part of a multi-threaded process. For the measurements in this paper, we ran on a cluster of thirty nodes connected by 1Gb/second Ethernet hubs. As would be expected in any large cluster, our system contains several different hardware configurations: six 200MHz machines with 64MB of memory and 4GB SCSI disks, eight 300 MHz machines with 128MB of memory and 4GB IDE disks, and sixteen 350 MHz machines with 128MB of memory and 8GB IDE disks.

Some key attributes of the system's implementation follow:

—The system runs on Linux 2.2.7 and uses the ext2 file system for storage [Ts'o 1999].

—The system consists of fourteen major components written in C++. The total system size is about forty-one thousand lines of code, yielding a 1MB executable.

—A mailbox fragment is stored in two files, regardless of the number of messages contained within. One file contains the message bodies, and the other contains message index information.

—The user map contains 256 buckets.

—The mailbox fragment files are grouped and stored in directories corresponding to the hash of user names (e.g., if Ann's hash value is 9, then her fragment files are `spool/9/ann` and `spool/9/ann.idx`). This design allows discovery of mailbox fragments belonging to a particular hash bucket – a critical operation during membership reconfiguration – to be performed by a single directory scan.

—Most of a node's memory is consumed by the soft user profile state. In the current implementation, each user entry takes 76 bytes plus 44 bytes per mailbox fragment. For example, in a system with ten million users running on 30 nodes, about 50 MB/node would be devoted to user soft state.

We developed a synthetic workload to evaluate Porcupine because users at our site do not receive enough email to drive the system into an overload condition. We did, however, design the workload generator to model the traffic patterns we have observed on our departmental mail servers. Specifically, we model a mean message size of 4.7KB, with a fairly fat tail up to about 1MB. Mail delivery (SMTP) accounts for about 90% of the transactions, with mail retrieval (POP) accounting for about 10%. Each SMTP session sends a message to a user chosen from a population according to a Zipf distribution with $\alpha = 1.3$, unless otherwise noted in the text.

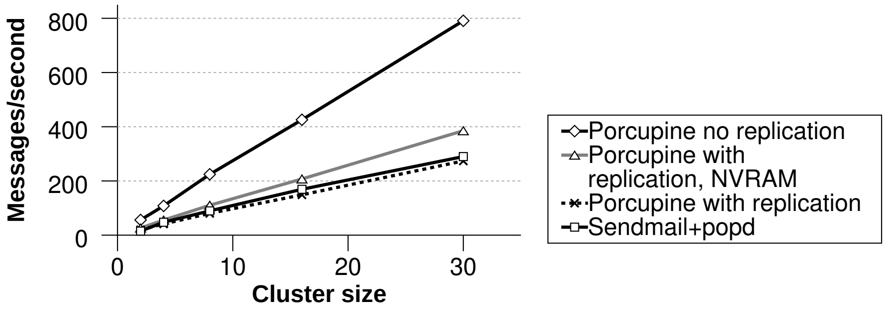


Fig. 7. Throughput scales with the number of hosts. This graph shows how Porcupine and the sendmail-based system scale with respect to cluster size.

For purposes of comparison, we also measure a tightly configured conventional mail system in which users and services are statically partitioned across the nodes in the cluster. In this configuration, we run SMTP/POP redirector nodes at the front end. At the back end, we run modified versions of the widely used Sendmail-8.9.3 and ids-popd-0.23 servers. The front-end nodes accept SMTP and POP requests and route them to back-end nodes by way of a hash on the user name. To keep the front ends from becoming a bottleneck, we determined empirically that we need to run one front end for every fifteen back ends. The tables and graphs that follow include the front ends in our count of the system size. Based on *a priori* knowledge of the workload, we defined the hash function to distribute users perfectly across the back-end nodes. To further optimize the configuration, we disabled all security checks, including user authentication, client domain name lookup, and system log auditing.

For both Porcupine and the conventional system, we defined a user population with size equal to 160,000 times the number of nodes in the cluster (or about 5 million users for the 30-node configuration). Nevertheless, since the database is distributed in Porcupine, and no authentication is performed for the conventional platform, the size of the user base is nearly irrelevant to the measurements. Each POP session selects a user according to the same Zipf distribution, collects and then deletes all messages awaiting the user. In the Porcupine configuration, the generator initiates a connection with a Porcupine node selected at random from all the nodes. In the conventional configuration, the generator selects a node at random from the front-end nodes. By default, the load generator attempts to saturate the cluster by probing for the maximum throughput, increasing the number of outstanding requests until at least 10% of them fail to complete within two seconds. At that point, the generator reduces the request rate and resumes probing.

We demonstrate performance by showing the maximum number of messages the system receives per second. Only message deliveries are counted, although message retrievals occur as part of the workload. Thus, this figure really reflects the number of messages the cluster can receive, write, read, and delete per second. The error margin is smaller than 5%, with 95% confidence interval for all values presented in the following sections.

Resource	No replication	With replication
CPU utilization	15%	12%
Disk utilization	75%	75%
Network send	2.5Mb/second	1.7Mb/second
Network receive	2.6Mb/second	1.7Mb/second

Table I. Resource consumption on a single node with one disk.

6.2 Scalability and Performance

Figure 7 shows the performance of the system as a function of cluster size. The graph shows four different configurations: without message replication, with message replication, with message replication using NVRAM for the logs, and finally for the conventional configuration of sendmail+popd. Although neither replicates, the Porcupine no-replication case outperforms and outscales conventional sendmail. The difference is primarily due to the conventional system's use of temporary files, excessive process forking, and the use of lock-files. With some effort, we believe the conventional system could be made to scale as well as Porcupine without replication. However, the systems would not be functionally identical, because Porcupine allows users to read incoming messages even when some nodes storing the user's existing messages are down.

For replication, the performance of Porcupine scales linearly when each incoming message is replicated on two nodes. There is a substantial slowdown relative to the non-replicated case, because replication increases the number of synchronous disk writes three-fold: once for each replica and once to update the coordinator's log. Even worse, in this hardware configuration the log and the mailbox fragments share the same disk on each node.

One way to improve the performance of replication is to use non-volatile RAM for the log. Since updates usually complete propagation and retire from the log quickly, most of the writes to NVRAM never need go to disk and can execute at memory speeds. Although our machines do not have NVRAM installed, we can simulate NVRAM simply by keeping the log in standard memory. As shown in Figure 7, NVRAM improves throughput; however, throughput is still about half that of the non-replicated case, because the system must do twice as many disk operations per message.

Table I shows the CPU, disk, and network load incurred by a single 350Mhz Porcupine node running at peak throughput. For this configuration, the table indicates that the disk is the primary impediment to single-node performance.

To demonstrate this, we made measurements on clusters with one and two nodes with increased I/O capacity. A single 300MHz node with one IDE disk and two SCSI disks delivered a throughput of 105 messages/second, as opposed to about 23 messages/second with only the IDE disk. We then configured a two node cluster, each with one IDE disk and two SCSI disks. The machines were each able to handle 38 messages/second (48 assuming NVRAM). These results (normalized to single-node throughput) are summarized in Figure 8.

Lastly, we measured a cluster in which disks were assumed to be infinitely fast. In this case the system does not store messages on disk but only records their digests in main memory. Figure 9 shows that the simulated system without the disk bottleneck achieves a six-fold improvement over the measured system. At this point, the CPU becomes the bottleneck. Thus Porcupine with replication performs comparatively better than on the real

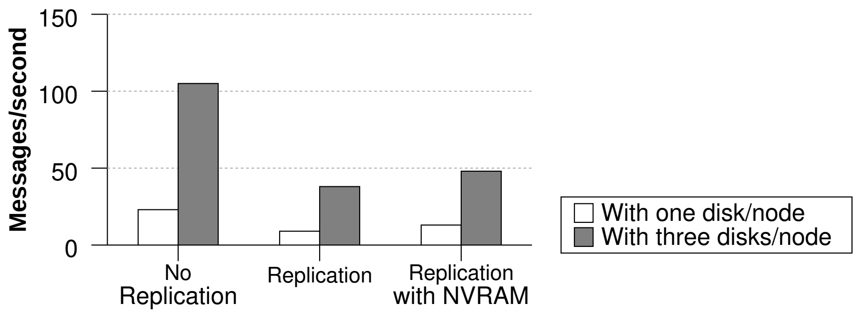


Fig. 8. Summary of single-node throughput in a variety of configurations.

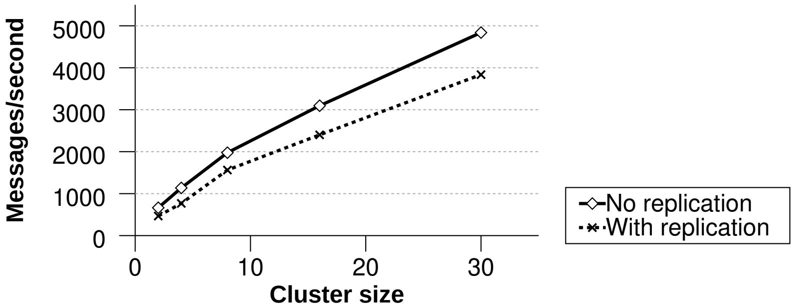


Fig. 9. Throughput of the system configured with infinitely fast disks.

system. The high performance observed in 2- and 4- node clusters is due to the shortcutting of inter-node RPCs into function calls that happens often in small clusters.

With balanced nodes, the network clearly becomes the bottleneck. In the non-replicated case, each message travels the network four times ((1) Internet to delivery agent (2) to mailbox manager (3) to retrieval agent (4) to Internet). At an average message size of 4.7KB, a 1Gb/second network can then handle about 6500 messages/second. With a single "disk loaded" node able to handle 105 messages/second, roughly 62 nodes will saturate the network as they process 562 million messages/day. With messages replicated on two nodes, the same network can handle about 20% fewer messages (as the message must be copied one additional time to the replica), which is about 5200 messages/second, or about 450 million messages/day. Using the throughput numbers measured with the faster disks, this level of performance can be achieved with 108 NVRAM nodes, or about 137 nodes without NVRAM. More messages can be handled only by increasing the aggregate network bandwidth. We address this issue further in Section 7.

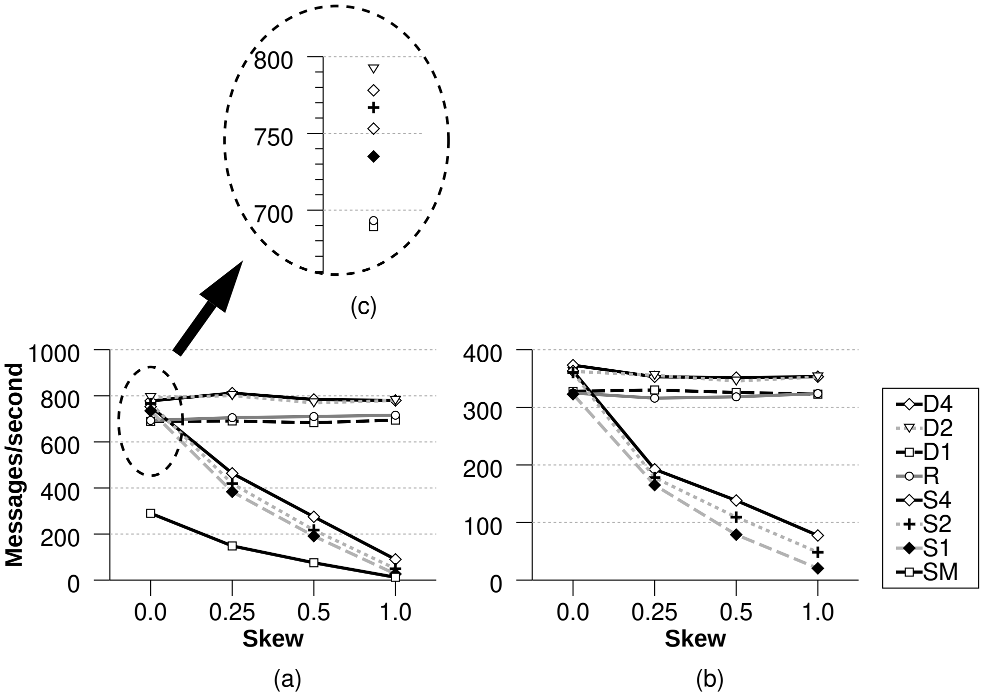


Fig. 10. (a) Non-replicated and (b) replicated throughputs on a 30-node system with various degrees of workload skew. Graph (c) shows a close-up view of the non-replicated throughputs under a uniform workload.

6.3 Load Balancing

The previous section demonstrated Porcupine's performance assuming a uniform workload distribution and homogeneous node performance. In practice, though, workloads are not uniformly distributed and the speeds of CPUs and disks on nodes differ. This can create substantial management challenges for system administrators when they must reconfigure the system manually to adapt to the load and configuration imbalance.

This section shows how Porcupine automatically handles workload skew and heterogeneous cluster configuration.

6.3.1 *Adapting to Workload Skew.* Figure 10 shows the impact of Porcupine's dynamic spread-limiting, load-balancing strategy on throughput as a function of workload skew for our 30-node configuration (all with a single slow disk). Both the non-replicated and replicated cases are shown. Skew along the x-axis reflects the inherent degree of balance in the incoming workload. When the skew equals zero, recipients are chosen so that the hash distributes uniformly across all buckets. When the skew is one, the recipients are chosen so that they all hash into a single user map bucket, corresponding to a highly non-balanced workload.

The graphs compare random, static, and dynamic load balancing policies. The random policy, labeled R on the graph, simply selects a host at random to store each message received; it has the effect of smoothing out any non-uniformity in the distribution. The static spread policy, shown by the lines labeled S1, S2, and S4, selects a node based on a hash of the user name spread over 1, 2 or 4 nodes, respectively. The dynamic spread policy – the one used in Porcupine – selects from those nodes already storing mailbox fragments for the recipient. It is shown as D1, D2 and D4 on the graph. Again, the spread value (1, 2, 4) controls the maximum number of nodes (in the absence of failure) that store a single user's mail. On message receipt, if the size of the current mail map for the recipient is smaller than the maximum spread, Porcupine increases the spread by choosing an additional node selected randomly from the cluster.

Static spread manages affinity well but can lead to a non-balanced load when activity is concentrated on just a few nodes. Indeed, a static spread of one (S1) corresponds to our sendmail+popd configuration in which users are statically partitioned to different machines. This effect is shown as well on the graph for the conventional sendmail+popd configuration (SM on Figure 10). In contrast, the dynamic spread policy continually monitors load and adjusts the distribution of mail over the available machines, even when spread is one. In this case, a new mailbox manager is chosen for a user each time his/her mailbox is emptied, allowing the system to repair affinity-driven imbalances as necessary.

The graphs show that random and dynamic policies are insensitive to workload skew, whereas static policies do poorly unless the workload is evenly distributed. Random performs worse than dynamic because of its inability to balance load and its tendency to spread a user's mail across many machines.

Among the static policies, those with larger spread sizes perform better under a skewed workload, since they can utilize a larger number of machines for mail storage. Under uniform workload, however, the smaller spread sizes perform better since they respect affinity. The key exception is the difference between spread=1 and spread=2. At spread=1, the system is unable to balance load. At spread=2, load is balanced and throughput improves. Widening the spread beyond two improves balance slightly, but not substantially. The reason for this has been demonstrated previously [Eager et al. 1986] and is as follows: in any system where the likelihood that a host is overloaded is p, then selecting the least loaded from a spread of s hosts will yield a placement decision on a loaded host with probability p^s. Thus, the chance of making a good decision (avoiding an overloaded host) improves exponentially with the spread. In a nearly perfectly-balanced system, p is small, so a small s yields good choices.

The effect of the loss of affinity with larger spread sizes is not pronounced in the Linux ext2 file system because it creates or deletes files without synchronous directory modification [Ts'o 1999]. On other operating systems, load balancing policies with larger spread sizes will be penalized more by increased frequency of directory operations.

6.3.2 *Performance under Uniform Workload.* Figure 10 (c) shows the system throughput under uniform workload. It is interesting to see that Porcupine's load balancing service can improve system performance even when the workload is uniform. D4, D2, S4 and S2 all perform well; the difference among them is statistically insignificant. S1, which emulates a statically partitioned system, performs about 5 to 10% worse than the rest because of the lack of load balancing. Under uniform workload, the load balancing service improves the performance mainly by avoiding nodes that are undergoing periodic buffer flush activities (`bdflush`) that stall all other disk I/O operations for a few seconds. R and D1 both perform about 15 to 20% worse, but for different reasons. R performs worse because it lacks load balancing, and because it ignores message affinity. D1 performs worse because it lacks load balancing, and because it tends to overload a few nodes that happen to host hyper-active users. On the other hand, D2 and D4 host hyper-active users on multiple

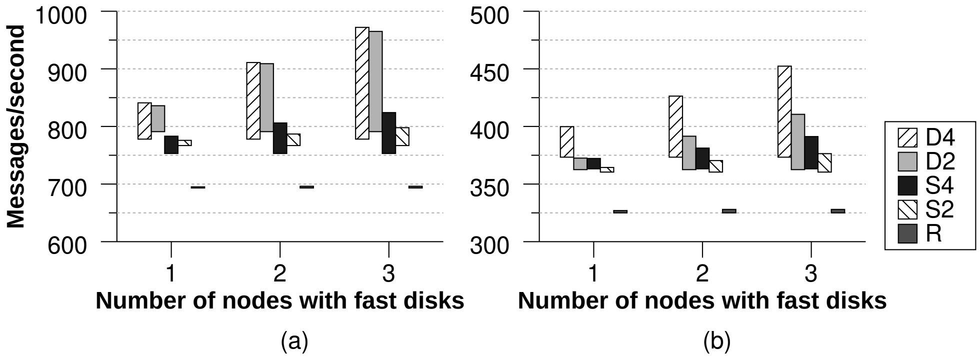


Fig. 11. Performance improvement by the Porcupine load balancing mechanism, without replication (a) and with replication (b). X axis is the number of nodes with fast disks. The bottom of each bar each bar shows the performance on the baseline system with a particular load balancing mechanism, and the height of the bar shows the relative improvement over the baseline system.

nodes, and the load balancer is able to split the workload at fine grain to keep the load on these nodes low.

6.3.3 *Adapting to Heterogeneous Configurations.* As mentioned in the previous section, the easiest way to improve throughput in our configuration is to increase the system's disk I/O capacity. This can be done by adding more machines or by adding more or faster disks to a few machines. In a statically partitioned system, it is necessary to upgrade the disks on all machines to ensure a balanced performance improvement. In contrast, because of Porcupine's functional homogeneity and automatic load balancing, we can improve the system's *overall* throughput for all users simply by improving the throughput on a few machines. The system will automatically find and exploit the new resources.

Figure 11 shows the absolute performance improvement of the 30-node configuration when adding two fast SCSI disks to each of one, two, and three of the 300Mhz nodes, with and without replication. The improvement for Porcupine shows that the dynamic load balancing mechanism can fully utilize the added capacity. Here, a spread of four slightly outperforms a spread of two, because the former policy is more likely to include the faster nodes in the spread. When a few nodes are many times faster than the rest, as is the case with our setting, the spread size needs to be increased. On the other hand, as described in Section 5, larger spread sizes tend to reduce the system efficiency. Thus, spread size is one parameter that needs to be revisited as the system becomes more heterogeneous.

In contrast, the statically partitioned and random message distribution policies demonstrate little improvement with the additional disks. This is because their assignment improves performance for only a subset of the users.

6.4 Failure Recovery

As described previously, Porcupine automatically reconfigures whenever nodes fail or restart. Figures 12 and 13 depict an annotated timeline of events that occur during the failure and recovery of 1, 3, and 6 nodes in a 30-node system without and with replication.

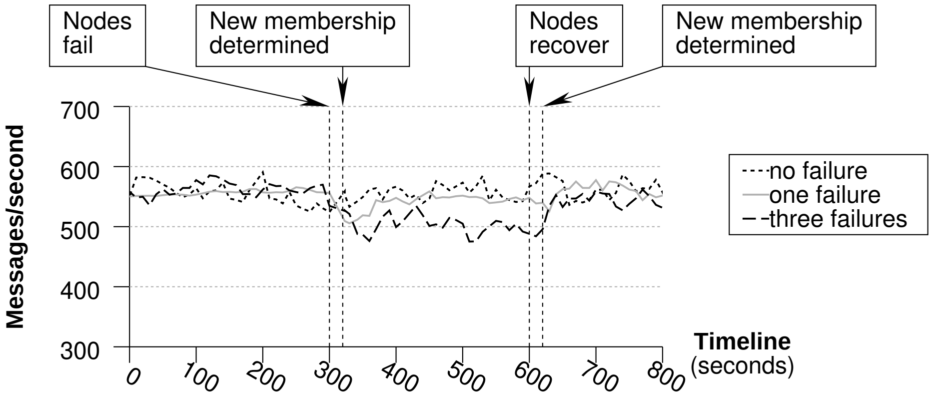


Fig. 12. Reconfiguration timeline without replication.

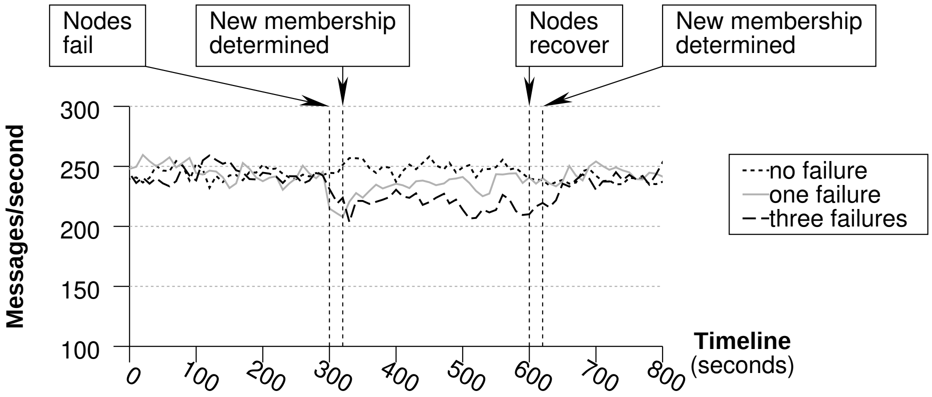
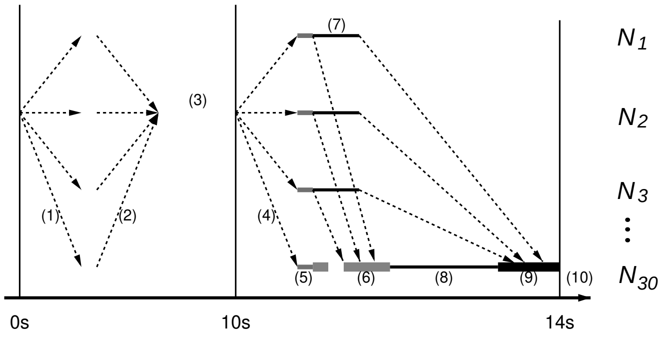


Fig. 13. Reconfiguration timeline with replication.

Both figures show the same behavior. Nodes fail and throughput drops as two things occur. First, the system goes through its reconfiguration protocol, increasing its load. Next, during the reconfiguration, SMTP and POP sessions that involve the failed node abort. After ten seconds, the system determines the new membership, and throughput increases as the remaining nodes take over for the failed ones. The failed nodes recover 300 seconds later and rejoin the cluster, at which time throughput starts to rise. For the non-replicated case, throughput increases back to the pre-failure level almost immediately. With replication, throughput rises slowly as the failed nodes reconcile while concurrently serving new requests.

Figure 14 shows the timing of events that take place during a reintegration of one node (N_{30}) to a 29-node cluster. Overall, fourteen seconds are spent to reconfigure the membership and to recover the soft state. The first ten seconds are spent in the membership protocol. Ongoing client sessions are not blocked during this period because the computational and the networking overheads of the membership protocol is minimal. The next four seconds are spent to recover the soft state. Again, ongoing client sessions on existing nodes are not affected during this period because the soft state recovery affects nodes other



(1) N_{30} recovers. N_2 starts TRM.
(2) Nodes reply to N_2.
(3) N_2 waits timeout.
(4) N_2 computes the new user map (4ms) and broadcasts the new membership.

(5) User profile disk scan. 6ms/node.
(6) User profile updates on N_{30}. 7ms/RPC.

(7) Mbox fragment scan. 20ms on existing nodes.
(8) Mbox fragment scan. 3300ms on N_{30}.

(9) Mail map updates on N_{30}. 7ms/RPC.
(10) Soft state reconstruction completes. Replica reconciliation begins.

Fig. 14. Time breakdown of failure recovery procedure. The timeline is not to scale.

than N_{30} only in a limited way — 6ms to scan the user profile and 20ms to scan mailbox fragments. On the other hand, N_{30} needs to scan its entire email spool directories to discover mailboxes and fill other nodes' mail maps (step 8). In addition, N_{30} needs to receive its assigned portions of the user profile database and mail map from other nodes (steps 6 and 9). However, notice that the cost of step 8 is orders of magnitude larger than that of all the other steps combined and depends only on the node's disk capacity and not on the number of nodes in the cluster. Thus, this analysis demonstrates that Porcupine's failure recovery scales with the cluster size.

7. LIMITATIONS AND FUTURE WORK

Porcupine's architecture and implementation have been designed to run well in very large clusters. There are, however, some aspects of its design and the environment in which it is deployed that may need to be rethought as the system grows to larger configurations.

First, Porcupine's communication patterns are flat, with every node as likely to talk to every other node. A 1Gb/second heavily switched network should be able to serve about 6500 messages/second (or 560 million messages/day) without replication. With replication, the network can handle 5200 messages/second, or 450 million messages/day. Beyond that, faster networks or more network-topology-aware load balancing strategies will be required to continue scaling.

Our membership protocol may also require adjustments as the system grows. Presently, the membership protocol has the coordinator receiving acknowledgment packets from all participants in a very short period of time. Although participants currently insert a randomized delay before responding to smooth out packet bursts at the receiver, we still need to evaluate whether this works well at very large scale. In other work, we are experimenting with a hierarchical membership protocol that eliminates this problem. In time, we may use this to replace Porcupine's current protocol.

Our strategy for reconstructing user profile soft state may also need to be revisited for systems in which a single user manager manages millions of users (many users, few machines). Rather than transferring the user profile soft state in bulk, as we do now, we could modify the system to fetch profile entries on use and cache them. This would reduce node recovery time (possibly at the expense of making user lookups slower, however).

8. RELATED WORK

The prototypical distributed mail service is Grapevine [Schroeder et al. 1984], a wide-area service intended to support about ten thousand users. Grapevine users are statically assigned to (user-visible) registries. The system scales through the addition of new registries having sufficient power to handle their populations. Nevertheless, Grapevine's administrators are often challenged to balance users across mail servers. In contrast, Porcupine implements a flat name space managed by a single cluster and automatically balances load. Grapevine provided a replicated user database based on optimistic replication, but it did not replicate mail messages. Porcupine uses optimistic replication for both mail and the user database.

As described earlier, contemporary email cluster systems deploy many storage nodes and partition the user population statically among them, either using a distributed file system [Christenson et al. 1997] or protocol redirectors [Deroest 1996]. As we demonstrate in this paper, this static approach is difficult to manage and scale and has limited fault tolerance.

Numerous fault-tolerant, clustered-computing products have been described in the past (e.g., [Kronenberg et al. 1986; Vogels et al. 1998; IBM 1998; Sun Microsystems 1999]). These clusters are often designed specifically for database fail-over, have limited scalability, and require proprietary hardware or software. Unlike these systems, Porcupine's goal is to scale to hundreds or thousands of nodes using standard off-the-shelf hardware and software.

Fox et al. [Fox et al. 1997] describe an infrastructure for building scalable network services based on cluster computing. They introduce a data semantics called BASE (Basically Available, Soft-state, Eventual consistency) that offers advantages for web-search and document-filtering applications. Our work shares many of their goals: building scalable Internet services with a semantics weaker than traditional databases. As in Fox's work, we observe that ACID semantics [Gray and Reuter 1993] may be too strong for our target applications and define a data model that is equal to the non-transactional model used by the system's clients. However, unlike BASE, our semantics support write-intensive applications requiring persistent data. Our services are also distributed and replicated uniformly across all nodes for greater scalability, rather than statically partitioned by function.

A large body of work exists on the general topic of load sharing, but this work has been targeted mainly at systems with long-running, CPU-bound tasks. For example, Eager et al. [Eager et al. 1986] show that effective load sharing can be accomplished with simple adaptive algorithms that use random probes to determine load. In [Dahlin 1999; Mitzenmacher 1998], the authors propose a class of load distribution algorithms using a random spread of nodes and a selection from the spread using cached load information. Their results show that a spread of two is optimal for a wide variety of situations in a homogeneous cluster. In the context of clusters and the Web, several commercial products automatically distribute requests to cluster nodes, typically using a form of round-robin or load-based dispatching [Cisco Systems 1999; Foundry Networks 1999; Resonate, Inc 1998; Platform Computing

1999]. In [Pai et al. 1998], the authors describe a "locality-aware request distribution" mechanism for cluster-based Web services. A front-end node analyzes the request content and attempts to direct requests so as to optimize the use of buffer cache in back-end nodes, while also balancing load. Porcupine uses load information, in part, to distribute incoming mail traffic to cluster nodes. However, unlike previous load-balancing studies that assumed complete independence of incoming tasks, we also balance the write traffic, taking message affinity into consideration.

Transparent automatic reconfiguration has been studied in the context of disks and networks. AutoRAID [Wilkes et al. 1995] is a disk array that moves data among disks automatically in response to failures and usage pattern changes. Autonet [Rodeheffer and Schroeder 1991] is a local area networking system that automatically reconfigures in response to router failures.

Porcupine uses replicated user maps to partition the user management task among nodes. This technique, called hash routing, has attracted wide attention recently, e.g., for web serving [Pai et al. 1998; Valloppillil and Ross 1998; Karger et al. 1997] and for operating system function distribution [Anderson et al. 1995; Feeley et al. 1995; Snaman and Thiel 1987]. Porcupine is the first system that combines the group membership protocol with hash routing to let each node determine the exact change in the hash map.

The replication mechanism used in Porcupine can be viewed as a variation of optimistic replication schemes, in which timestamped updates are pushed to peer nodes to support multi-master replication [Agrawal et al. 1997; Wuu and Bernstein 1984]. Porcupine's total object update property allows it to use a single timestamp per object, instead of timestamp matrices, to order updates. In addition, since updates are idempotent, Porcupine can retire updates more aggressively. These differences make Porcupine's approach to replication simpler and more efficient at scale.

Several file systems have scalability and fault tolerance goals that are similar to Porcupine's [Anderson et al. 1995; Birrell et al. 1993; Lee and Thekkath 1996; Liskov et al. 1991; Thekkath et al. 1997]. Unlike these systems, Porcupine uses the semantics of the various data structures it maintains to exploit their special properties in order to increase performance or decrease complexity.

9. CONCLUSIONS

We have described the architecture, implementation, and performance of the Porcupine scalable mail server. We have shown that Porcupine meets its three primary goals:

Manageability. Porcupine automatically adapts to changes in configuration and workload. Porcupine masks heterogeneity, providing for seamless system growth over time using latest-technology components.

Availability. Porcupine continues to deliver service to its clients, even in the presence of failures. System software detects and recovers automatically from failures and integrates recovering nodes.

Performance. Porcupine's single-node performance is competitive with other systems, and its throughput scales linearly with the number of nodes. Our experiments show that the system can find and exploit added resources for its benefit.

Porcupine achieves these goals by combining three key architectural techniques based on the principle of functional homogeneity: automatic reconfiguration, dynamic transaction scheduling, and replication. In the future, we hope to construct, deploy and evaluate configurations larger and more powerful than the ones described in this paper.

ACKNOWLEDGEMENTS

We thank Eric Hoffman, Bertil Folliot, David Becker, and other members of the Porcupine project for the valuable discussions and comments on the Porcupine design. We also thank the anonymous reviewers for helping us improve the paper.

REFERENCES

AGRAWAL, D., ABBADI, A. E., AND STEIKE, R. C. 1997. Epidemic algorithms in replicated databases. In *16th ACM Symp. on Princ. of Database Systems*. ACM, Tucson, AZ, 161–172.

ANDERSON, T., DAHLIN, M., NEEFE, J., PATTERSON, D., ROSELLI, D., AND WANG, R. 1995. Serverless network file systems. In *15th Symposium on Operating Systems Principles*. ACM, Copper Mountain, CO.

BIRRELL, A. D., HISGEN, A., JERIAN, C., MANN, T., AND SWART, G. 1993. The Echo distributed file system. Tech. Rep. 111, Compaq Systems Research Center. September.

BRISCO, T. P. 1995. RFC1794: DNS support for load balancing. http://www.cis.ohio-state.edu/htbin/rfc/rfc1794-.html.

CHANKHUNTHOD, A., DANZIG, P., NEERDAELS, C., SCHWARTZ, M., AND WORRELL, K. 1996. A hierarchical internet object cache. In *Winter USENIX Technical Conference*.

CHEN, P. M., LEE, E. K., GIBSON, G. A., KATZ, R. H., AND PATTERSON, D. A. 1994. RAID: High-performance, reliable secondary storage. *ACM Computing Surveys 26,* 2 (June), 145–185.

CHRISTENSON, N., BOSSERMAN, T., AND BECKEMEYER, D. 1997. A highly scalable electronic mail service using open systems. In *Symposium on Internet Technologies and Systems*. USENIX, Monterey, CA.

CHRISTIAN, F. AND SCHMUCK, F. 1995. Agreeing on processor group membership in asynchronous distributed systems. Tech. Rep. CSE95-428, UC San Diego.

CISCO SYSTEMS. 1999. Local director. http://www.cisco.com/warp/public/751/lodir/index.html.

CRISPIN, M. 1996. RFC2060: Internet message access protocol version 4 rev 1. http://www.cis.ohio-state-.edu/htbin/rfc/rfc2060.html.

DAHLIN, M. 1999. Interpreting stale load information. In *The 19th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, Austin, TX.

DEROEST, J. 1996. Clusters help allocate computing resources. http://www.washington.edu/tech_home/-windows/issue18/clusters.html.

EAGER, D. L., LAZOWSKA, E. D., AND ZAHORJAN, J. 1986. Adaptive load sharing in homogeneous distributed systems. *IEEE Trans. on Software Engineering 12,* 5 (May), 662–675.

FEELEY, M. M., MORGAN, W. E., PIGHIN, F. H., KARLIN, A. R., LEVY, H. M., AND THEKKATH, C. A. 1995. Implementing global memory management in a workstation cluster. In *15th Symposium on Operating Systems Principles*. ACM, Copper Mountain, CO, 130–146.

FOUNDRY NETWORKS. 1999. ServerIron Switch. http://www.foundrynet.com/serverironfspec.html.

FOX, A., GRIBBLE, S. D., CHAWATHE, Y., BREWER, E. A., AND GAUTHIER, P. 1997. Cluster-based scalable network services. In *16th Symposium on Operating Systems Principles*. ACM, St. Malo, France, 78–91.

GRAY, J. AND REUTER, A. 1993. *Transaction Processing: Concepts and Techniques*. Morgan-Kaufmann.

IBM. 1998. *High Availability Cluster Multi-Processing for AIX*. Available at http://www.rs6000.ibm-.com/doc_link/en_US/a_doc_lib/aixgen/hacmp_index.html.

KARGER, D., LEHMAN, E., LEIGHTON, T., PANIGRAHY, R., LEVINE, M., AND LEWIN, D. 1997. Consistent hashing and random trees: distributed caching protocols for relieving hot spots on the World Wide Web. In *Symposium on Theory of Computing*. ACM, El Paso, TX, 654–663.

KRONENBERG, N. P., LEVY, H. M., AND STRECKER, W. D. 1986. VAXclusters: A closely-coupled distributed system. *ACM Trans. on Computer Systems 2,* 4, 130–146.

LAMPORT, L. 1978. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM 21,* 7 (July), 558–565.

LEE, E. K. AND THEKKATH, C. 1996. Petal: Distributed virtual disks. In *7th International Conf. on Architectural Support for Prog. Lang. and Operating Systems*. ACM, Cambridge, MA, 84–92.

LISKOV, B., GHEMAWAT, S., GRUBER, R., JOHNSON, P., SHRIRA, L., AND WILLIAMS, M. 1991. Replication in the Harp file system. In *13th Symposium on Operating Systems Principles*. ACM, Pacific Grove, CA, 226–238.

LISKOV, B., SHRIRA, L., AND WROCLAWSKI, J. 1991. Efficient at-most-once messages based on synchronized clocks. *ACM Trans. on Computer Systems 9,* 2, 125–142.

MILLS, D. L. 1992. RFC1305: Network time protocol (version 3). http://www.cis.ohio-state-.edu/htbin/rfc/rfc1305.html.

MILLS, D. L. 1994. Improved algorithms for synchronizing computer network clocks. In *SIGCOMM*. ACM, London, UK, 317–327.

MITZENMACHER, M. 1998. How useful is old information? Tech. Rep. 98-002, Compaq Systems Research Center. Feb.

MYERS, J. G. AND ROSE, M. T. 1996. RFC1939: Post office protocol version 3. http://www.cis.ohio-state-.edu/htbin/rfc/rfc1939.html.

PAI, V. S., ARON, M., BANGA, G., SVENDSEN, M., DRUSCHEL, P., ZWAENEPOEL, W., AND NAHUM, E. 1998. Locality-aware request distribution in cluster-based network servers. In *8th International Conf. on Architectural Support for Prog. Lang. and Operating Systems*. ACM, San Jose, CA, 206–216.

PLATFORM COMPUTING. 1999. LSF. http://www.platform.com.

POSTEL, J. 1982. RFC821: Simple mail transfer protocol. http://www.cis.ohio-state.edu/htbin/rfc/rfc821.html.

RESONATE, INC. 1998. Central Dispatch. http://www.resonate.com/products/central_dispatch/.

RODEHEFFER, T. AND SCHROEDER, M. D. 1991. Automatic reconfiguration in Autonet. In *13th Symposium on Operating Systems Principles*. ACM, Pacific Grove, CA, 183–187.

SCHROEDER, M. D., BIRRELL, A. D., AND NEEDHAM, R. M. 1984. Experience with Grapevine: The growth of a distributed system. *ACM Transactions on Computer Systems 2,* 1 (February), 3–23.

SNAMAN, W. E. AND THIEL, D. W. 1987. The VAX/VMS distributed lock manager. *Digital Technical Journal* 5.

SUN MICROSYSTEMS. 1999. *Sun Cluster Architecture*. Available at http://www.sun.com/clusters/wp-clusters-arch.pdf.

THEKKATH, C., MANN, T., AND LEE, E. 1997. Frangipani: A scalable distributed file system. In *16th Symposium on Operating Systems Principles*. ACM, St. Malo, France, 224–237.

TS'O, T. 1999. Ext2 home page. http://web.mit.edu/tytso/www/linux/ext2.html.

VALLOPPILLIL, V. AND ROSS, K. W. 1998. Cache array routing protocol v1.0. Internet draft. http://www.ircache.net/Cache/ICP/carp.txt.

VOGELS, W., DUMITRIU, D., BIRMAN, K., GAMACHE, R., MASSA, M., SHORT, R., VERT, J., BARRERA, J., AND GRAY, J. 1998. The design and architecture of the Microsoft cluster service. In *28th International Symposium on Fault-Tolerant Computing*. IEEE, Munich, Germany, 422–431.

WILKES, J., GOLDING, R., STAELIN, C., AND SULLIVAN, T. 1995. The HP AutoRAID hierarchical storage system. In *15th Symp. on Operating Systems Principles*. ACM, Copper Mountain, CO, 96–108.

WUU, G. T. J. AND BERNSTEIN, A. J. 1984. Efficient solutions to the replicated log and dictionary problems. In *Proceedings of the 3rd Symposium on Principles of Distributed Computing*. ACM, Vancouver, Canada, 233–242.