A cross-architectural quantitative evaluation of mobility approaches

Vasanta Chaganti, James Kurose, Arun Venkataramani
University of Massachusetts, Amherst
Email: {vchaganti, kurose, arun}@cs.umass.edu

***Abstract*—Future Internet Architectures must support the rapid growth of traffic generated by mobile endpoints in a manner that is scalable and ensures low latency. We present a quantitative evaluation of three distinct approaches towards handling endpoint mobility: name-based forwarding, indirection and a global name service (GNS). Using a range of parameterized mobility distributions and real ISP topologies, we describe representative instantiations of each approach and evaluate their performance using four key metrics: update cost and update propagation cost in the control plane; and forwarding traffic cost and time-to-connect (TTC) in the data plane. (1) We show that by leveraging the fact that realistic endpoint mobility distributions show a high probability of being at a small subset of visited locations, name-based forwarding strategies can provide up to 60% improvement in control costs over simple best-port forwarding. (2) We show that the TTC in these name-based forwarding strategies is comparable to the TTC in the GNS. (3) Finally we show that a GNS-based approach offers the most suitable balance of total (combined data and control) cost to TTC across all approaches, all endpoint mobility distributions, and all ISP topologies considered.**

I. Introduction

Global traffic generated by mobile users accounts for 8% of the world's IP traffic today, and is projected to increase to 20% by 2021 [2]. And yet, our understanding of the relative benefits of various proposed approaches for handling mobility in a network that will be increasingly dominated by mobile consumers and producers [3] is lagging.

In today's Internet, the lack of a separation between an endpoint's identity and its network location results in higher-layer connections being broken every time an endpoint changes its network attachment point. Although there is consensus on the need for location-identifier separation [11], establishing a scalable location-independent approach to mobility that has low latency and that limits data and control cost is an important topic of on-going research.

Several location-independent 'clean-slate' architectures [8] have advocated various approaches to handle mobility, with proposals that vary from name-based forwarding [20], to in-network DHT [10], to logically centralized name-resolution services [12] and to hybrid name-based and DHT schemes [4]. Performance analyses of these mobility approaches have shown significant advantages over current Internet mobility approaches – resilience to congestion and failed links in NDN [20], significantly lower latency using a globally-distributed name service (GNS) over state-of-the-art DNS in MobilityFirst [12], and a DHT-based rendezvous mechanism that implements load-balancing with up to 12 million name-resolution servers in NetInf [4].

However, these architectural analyses have been focused primarily on a specific approach; what is still lacking is a cross-architectural comparison among these various mobility approaches. The goal of our paper is to (a) define an appropriate set of quantitative metrics that can be used to compare and evaluate mobility approaches; (b) provide insight into the factors that effect each mobility approach; and (c) present a quantitative argument for a mobility approach that provides the most suitable balance of performance trade-offs.

In this paper, we provide a cross-architectural parametrized comparison of three distinct approaches towards handling endpoint mobility: (a) name-based forwarding, (b) indirection, and (c) a global name service (GNS). We describe representative instantiations of each approach and evaluate their performance using four key metrics; two control plane metrics: update cost and update propagation costs, and two data plane metrics: forwarding traffic cost and time-to-connect (TTC). We also define a total cost (combined forwarding traffic and update propagation) that we trade off against TTC. Based on real network topologies, and parametrized workload and mobility distributions, we show the following key results:

- **Most suitable balance of performance trade-offs in a GNS:** We show that a GNS-based approach with $g = \sqrt{n}$ uniformly distributed servers, can achieve a mean TTC that is approximately 1.5 times that of best-port (that has the lowest TTC) and a mean total cost approximately 6 times that of indirection (that has the lowest total cost).
- **Control cost scalability of name-based forwarding approaches:** We show that by leveraging the fact that realistic endpoint mobility distributions show a high probability of being at a small subset of visited locations, name-based forwarding strategies can provide up to 60% improvement in control costs over simple best-port forwarding.
- **Cost trade-off of GNS vs. name-based forwarding approaches:** We show that the time-to-connect (TTC) in a GNS-based approach is comparable to name-based forwarding strategies that leverage the skewed-nature of realistic mobility distributions. We show that across a range of mobility distributions, a GNS-based approach can result in up to 20% improvement in mean total cost over name-based forwarding in the largest ISP, and we show that the mean total cost in the GNS-based approach is no worse than name-based forwarding, across all ISPs.

The remainder of this paper is structured as follows. In Section II we define representative instantiations of these three fundamental mobility approaches. Section III defines our evaluation framework, and performance metrics, and Section IV describes the network topologies, endpoint mobility and workload distributions used in our analyses. Section V presents our results and insights for specific mobility approaches, while Section VI presents our comparative analyses and insights across mobility approaches. Section VII provides discussion, and Section IX the conclusions.

II. Three approaches towards mobility

In this section we describe canonical forwarding strategies used to evaluate three broader classes of mobility approaches — (a) name-based forwarding, (b) indirection, and (c) a global name resolution service (GNS). We illustrate each approach using the simple examples in Figs. 1 and 2.

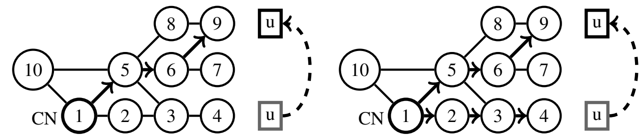
We define an endpoint mobility event as a change in that endpoint's point of attachment to the network. We define a content request as a request directed from a correspondent node (CN) to the mobile endpoint. Figs. 1 and 2 show an endpoint u changing network attachment points, and a CN at router 1 attempting to connect to the mobile endpoint u. We assume that a routing protocol maintains shortest path routes explicitly in the network. In each of the mobility approaches described, we show the forwarding actions (denoted by $\rightarrow$) taken by the first-hop router for the first packet to traverse from the CN to the mobile endpoint. Across all of our forwarding strategies every endpoint and every router has a unique and permanent network identifier.

A. *Name-based forwarding*

We describe two name-based forwarding strategies: best-port forwarding and parallel-multicast forwarding and discuss the differences between these two strategies in the costs incurred when handling a mobility event. In the case of best-port forwarding, an endpoint's mobility event will always incur a non-zero control cost, as explained below. In the parallel-multicast strategy the same mobility event may not incur any control cost.

Best-port forwarding. In best-port forwarding, each router maintains a mapping to each endpoint's identifier through a single best port. When an endpoint moves, the control cost incurred in this strategy, is the cost of broadcasting the endpoint's identifier from its new location. Routers receiving the endpoint's updated location, only forward the update if their forwarding interface to the endpoint has changed. Fig. 1(a) shows the path taken by the first packet from the source router 1 to the endpoint. As each router always has an updated mapping of the mobile endpoint, there is no path stretch incurred in this strategy over the shortest path provided by the routing protocol.

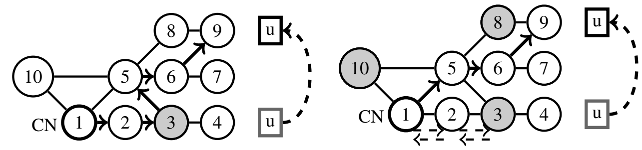
Parallel-multicast forwarding. This name-based forwarding strategy leverages the skewed nature of popularity distribution of mobility locations, i.e., the fact that current network mobility distributions show that an endpoint is at a few



(a) Best-port forwarding: The CN at router 1, forwards only on one interface to the router 9, the endpoint's current location.

(b) Parallel-multicast forwarding: The CN at router 1, forwards on all interfaces to popular locations 4 and 9 of the endpoint. Subsequent routers on-path use global network information to route to the popular location if they are on the shortest path to that location.

Fig. 1: Name-based forwarding: best-port and parallel-multicast



(a) Indirection using triangle routing with an indirection server at 3. The CN at router 1, forwards the packet to an indirection server co-located with router 3. The indirection server forwards the packet to router 9, the endpoint's current location.

(b) Global name-service with servers at $3, 8$ and, 10. The CN at router 1, corresponds with a GNS server co-located with router 3. The GNS server responds with the current location of the endpoint. Router 1 forwards packets to the endpoint's current location.

Fig. 2: GNS and Indirection

dominant or popular network locations with a high probability, and has a small probability of being at a large number of "unpopular" locations [19].

In this strategy, an endpoint announces a set of popular locations when it first arrives to the network and the forwarding information base (FIB) at each router maintains interfaces to this set of popular locations. The FIB *does not* maintain any per-flow state. Upon subsequent mobility events, the endpoint only announces its new location when this location is not one of the popular locations, i.e., is an "unpopular" location. Therefore, the FIB also maintains an interface to its last known unpopular location. At any point in time, the endpoint can be reached either via the last known unpopular location or at one of the popular locations.

For a source router to reach an endpoint, the source router first tries all popular locations in parallel, and then the last known unpopular location. To forward the first packet to the set of popular locations, the source router forwards the packet along each interface to a popular location. Subsequent on-path routers (second-hop onwards) use global network information to decide to which popular location to route the packet. If a router finds that it is on the shortest path to a location, then the router continues forwarding the packet towards that location. Each router that receives the packet performs this look-up to avoid the packet traversing multiple paths to reach the same destination. If m popular locations are downstream from an nth-hop router, rather than forwarding m copies of the packet, the nth router only forwards one packet along the link to the $(n+1)$st router. The $(n+1)$st router may then replicate the packet if there are multiple interfaces from this router to the m popular locations. Fig. 1(b) shows router 1 forwarding the first packet on two outgoing ports, corresponding to two popular locations of the endpoint.

B. Indirection

We instantiate an indirection approach using an indirection server that is responsible for maintaining up-to-date information about an endpoint's mobility. The endpoint updates the indirection server upon every mobility event with its new location. A CN requests content from the endpoint by first routing the request to the indirection server. The indirection server then routes the request to the endpoint. Fig. 2(a) shows a CN at router 1 attempting to contact the endpoint u. Router 1 first forwards packets to the indirection server that then forwards packets to the endpoint. We assume that the indirection servers are placed statically.

C. Globally-distributed name resolution service

We instantiate a global name service (GNS) approach similar to proposals in NetInf [4] and MobilityFirst [17]. In this strategy, whenever an endpoint moves, it sends its new location to the closest GNS server. The endpoint's new location is then replicated across GNS servers. In our results section, we detail heuristics for selecting the number of GNS servers in various ISP topologies.

Fig. 2(b) shows a CN at router 1 contacting a GNS server at router 3 to resolve the current location of the endpoint. Once the endpoint location is resolved, the source router sends packets directly addressed to the location of the endpoint. The GNS servers are assumed to be distributed uniformly in the network, and each CN maintains a mapping to its closest GNS server to minimize latency when requesting content from a mobile endpoint. We assume that the GNS servers are placed statically.

III. Quantitative evaluation

In this section, we define the four performance metrics that we use in the control and data plane trade-offs of our representative forwarding strategies.

A. Evaluation Framework

We define three network-centric cost metrics; *update cost* and *update propagation cost* in the control plane, and *forwarding traffic cost* in the data plane. We trade off these cost metrics against *time-to-connect* (TTC) for each of the forwarding strategies defined in Section II. Our evaluation of the control and data plane trade-offs are based on (a) the network topology, (b) forwarding plane strategy (c) content request workload, and (d) endpoint mobility patterns.

Let a network topology with routers be represented by the set $\mathcal{N} = \{1, \ldots, r, \ldots, n\}$, $\forall r \in \mathcal{N}$ and $|\mathcal{N}| = n$. Let an edge between neighboring routers $i \in \mathcal{N}$ and $j \in \mathcal{N}$ be represented by (i, j). Let the set of endpoints in the network be represented by $\mathcal{U}$. Each endpoint has a unique and permanent network identifier that we refer to as u. The endpoint's identifier can belong to either a flat name-space [3] or a hierarchical name-space [20]. The RIB at each router maintains end-to-end paths, and the FIB maintains a mapping from an endpoint's unique identifier to a set of outgoing ports.

We assume that endpoint mobility and content requests are discrete events and we represent an endpoint's expected mobility rate in network by $\mu(u)$ and the expected content request rate by $\lambda(u)$. When an endpoint announces its new location upon a mobility event, we assume that the update is propagated instantaneously, i.e., we do not consider convergence time of updates in our performance metrics. We also assume (a) connected topology without disruptions or node failures, (b) sufficient bandwidth to support each forwarding strategy.

B. Update cost

For a router $r \in \mathcal{N}$, let the Forwarding Interest Base (FIB) at time t be represented as $\mathrm{FIB}(r, t)$. Let $\mathrm{FIB}(r, u, t)$ represent the set of outgoing interfaces or ports that u resolves to based on longest prefix match. Recall that we have defined a mobility event as a change in network attachment points of an endpoint. Let $M(u)$ be the set of times at which endpoint $u \in U$ changes its point of attachment to the network. Let $t, t - \tau \in M(u)$ be the time stamps of two discrete consecutive mobility events of endpoint u. We now define the update cost $\mathrm{UC}(r, u, t)$ as the cost incurred by router r at time t, when endpoint u's mobility event causes a change in the router's forwarding behavior. The update cost at router r is given by

$$\mathrm{UC}(r,u,t) = \begin{cases} 1, & \mathrm{FIB}(r,u,t) \neq \mathrm{FIB}(r,u,t-\tau) \\ 0, & \mathrm{FIB}(r,u,t) = \mathrm{FIB}(r,u,t-\tau). \end{cases} \quad (1)$$

The update cost in the network represents the total number of routers in the network that incur an update due to the mobility of endpoint u, and is given as

$$\mathrm{UC}(u,t) = \sum_{r} \mathrm{UC}(r,u,t). \quad (2)$$

Then the expected update cost per mobility event over all endpoints and all endpoint mobility events is,

$$\mathrm{E}[\mathrm{UC}] = \frac{1}{|U|} \sum_{u \in U} \left[\frac{1}{|M(u)|} \sum_{t \in M(u)} \mathrm{UC}(u,t) \right]. \quad (3)$$

C. Update propagation cost

An update propagation cost is the cost of propagating the update from the new location of the endpoint to network routers such that the endpoint is reachable from any router in the network. The manner in which these updates flow depends on the mobility approach; in the GNS the updates are propagated through the GNS servers, in indirection the updates go to a single indirection server, and in name-based forwarding the update passes through all the routers that change their forwarding behavior. The update propagation cost incurred in propagating an update between any two neighboring routers i and j due to an endpoint u's mobility event is given by,

$$\mathrm{UPC}(u,i,j,t) = \begin{cases} 1, \text{ if the update traverses } (i,j) \text{ at time } t \\ 0, \text{ if the update does not traverse } (i,j) \\ \quad \text{at time } t, \forall i, j \in \mathcal{N}. \end{cases} \quad (4)$$

The update propagation cost in the network is the total number of edges traversed in the network to maintain reachability to the new location of the endpoint from any router in the network and is given by,

$$\text{UPC}(u,t) = \sum_{(i,j)} \text{UPC}(i,j,t). \tag{5}$$

The expected update propagation cost in the network, per mobility event over all endpoints and all endpoint mobility events is,

$$\text{E}[\text{UPC}] = \frac{1}{|U|} \sum_{u \in U} \left[\frac{1}{|M(u)|} \sum_{t \in T_m(u)} \text{UPC}(u,t) \right]. \tag{6}$$

D. Forwarding traffic cost

Forwarding traffic cost is defined as the total number of edges traversed along all paths *by the first packet* sent from a CN at router r to an endpoint u. The forwarding traffic cost incurred between any two neighboring routers i and j, due to a request from a CN at router r to a mobile endpoint u is,

$$\text{FT}(i,j,t) = \begin{cases} 1, \text{ if the first packet traverses } (i,j) \text{ at time } t \\ 0, \text{ if the first packet does not traverse} \\ \quad (i,j) \text{ at time } t, \ \forall i,j \in \mathcal{N}. \end{cases} \tag{7}$$

Let the paths traversed by the first packet from router r to the mobile endpoint u be given by $p(r,u)$. Then, forwarding traffic cost between a CN at router r and a mobile endpoint u is the total number of links over all paths traversed by the first packet from router r to the mobile endpoint given by,

$$\text{FT}(r,u,t) = |p(r,u)|. \tag{8}$$

We note that depending on the forwarding strategy used, there might be more than one path traversed from r to u.

Let $D(u)$ be the set of times at which endpoint $u \in U$ receives a request. Then the expected forwarding traffic cost per request over all endpoints and all endpoint request events is,

$$\text{E}[\text{FT}] = \frac{1}{|U|} \sum_{u \in U} \left[\frac{1}{|D(u)|} \sum_{t \in D(u)} \text{FT}(r,u,t) \right]. \tag{9}$$

E. Time-to-connect

We define time-to-connect (TTC) as the delay from when an intent to communicate was established by a CN at router r to the time when it has successfully *sent the first packet* of data to the mobile endpoint u and is given by,

$$\text{TTC}(r,u,t) = w(r,u) \tag{10}$$

where $w(r,u)$ is the latency along the shortest path to the endpoint. The expected TTC per request over all endpoints and all endpoint request events is,

$$\text{E}[\text{TTC}] = \frac{1}{|U|} \sum_{u \in U} \left[\frac{1}{|D(u)|} \sum_{t \in D(u)} \text{TTC}(r,u,t) \right]. \tag{11}$$

Note on the data plane costs. We note that both the forwarding traffic cost and the TTC are measured based on the first packet of data. The subsequent packets between the CN and endpoint follow the shortest data path as defined by the respective forwarding strategies. We believe that the data plane costs on the first packet are of great significance — particularly in the case of real-time, transactional or short-lived flows [13].

F. Total cost: combined data and control cost

We define an expected total cost $\text{E}[\text{C}]$ as the cost incurred due to the expected forwarding traffic cost and expected update propagation cost,

$$\text{E}[\text{C}] = \text{E}[\text{FT}] + \text{E}[\text{UPC}]. \tag{12}$$

IV. Parametrized model

In this section we describe the network topology used from publicly available datasets, the routing policy, and the parametrized endpoint mobility distribution and workload distribution. We also present the parameters used as inputs to our simulation analyses.

A. Routing topology and policy

We evaluate the four canonical forwarding strategies described in Section II, using the following ISP topologies: Sprintlink, Ebone, Tiscali, Exodus and Abovenet, with their associated link-weights and link-latencies from the RocketFuel ISP topologies dataset [15]. We use OSPF as the network routing policy.

We assign IP prefixes to routers in an ISP, by dividing the 32-bit IP address range into equal-sized address spaces and choose a /27 address space for each router in the network with uniform random probability. A /27 address space allows 32 endpoints to attach to each router in the network. When an endpoint enters the network, the endpoint receives a hierarchical IP prefix belonging to the subnet of the router to which it is attached. Each endpoint is reachable via its IP prefix and maintains the IP prefix for the duration of the endpoint's presence in the network. The network location to which the endpoint arrives is chosen from a uniform distribution.

B. Endpoint mobility and workload

Each endpoint in the network has discrete arrival, mobility and content request events. For the purpose of this simulation, we set the expected mobility rate across all endpoints to μ and the expected content request rate to λ, such that $\mu(u_1) = \mu(u_2)\ldots\mu(u_z) = \mu$ and $\lambda(u_1) = \lambda(u_2)\ldots\lambda(u_z) = \lambda$, $\forall\{u_1,u_2,\ldots u_z\} \in \mathcal{U}$.

Endpoint mobility distribution. We use a power-law distribution to describe endpoint mobility, as it has the versatility to express a range of distributions including the zipfian distribution and to leverage the fact that current network mobility distributions are heavy tailed, i.e., an endpoint is at a few dominant or popular network locations with a high probability and at a large number of "unpopular" locations with a very small probability [19].

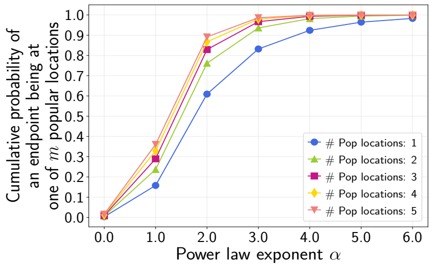


Fig. 3: Cumulative probability distribution of an endpoint being at one of m popular locations for varying α in the Sprint ISP.

The probability with which an endpoint's next visited location is chosen is governed by a power-law distribution with an exponent α. The visited locations for each endpoint are given by the totally ordered set $\mathcal{L}(u)$, $|\mathcal{L}(u)| = n$ where $\mathcal{L}(u)$ is chosen from a uniform random shuffling over the set of routers in the network. The popular locations for each endpoint u are the first m locations in $\mathcal{L}(u)$ that correspond to the highest visit probabilities in the power-law distribution.

Fig. 3 shows the cumulative power-law distribution for the Sprint ISP where, $n = 315$, the power-law exponent $\alpha \in [0, 6]$, and the number of popular locations, $m \in [1, 5]$. For an $\alpha = 0$, the endpoint visits every router in the network with equal probability (uniform probability distribution). As $\alpha \to 6$, the power-law distribution tends to a point (delta) distribution with the endpoint always being at the most popular location. For $\alpha > 2.0$ and $m > 3$, the cumulative probability of being at a popular location is greater than 0.8 and consequently the probability of being at an "unpopular" location is less than 0.2. *We want to show the effect of endpoint mobility between popular and unpopular locations, and therefore we focus on the mobility distributions with* $\alpha \leq 2.0$.

Workload distribution. Forwarding traffic for an endpoint is generated as a content request originating from a CN. The network location of the CN is chosen from a uniform distribution and the request events are generated with exponential inter-arrival times with mean $\lambda(u)$.

C. Simulation parameters

We have developed a discrete-event simulator in Python that models the four forwarding strategies, and collects the performance metrics defined in Section II. Each run of the simulation takes as inputs the parameters in Table I. To compare the performance across forwarding strategies, the mobility and workload traces of all the endpoints are written to text files, and the same traces are used for all forwarding strategies. We set the number of endpoints in the system $|U| = 200$, which we found to approximate uniformly distributed arrival locations across all ISPs.

V. Results per mobility approach

In this section, we evaluate particular aspects of trade-offs among the control and data plane metrics for each of the four canonical forwarding strategies that are relevant to a particular mobility approach. We first discuss the effect of the mobility distribution on performance metrics in name-based forwarding in Section V-A. Next, we study the effect of the number of GNS servers on the trade-off between TTC and update cost in Section V-B. Finally, in Section V-C, we discuss the trade-off between TTC and update cost in indirection and the impact of indirection on the data plane performance in long-lived packet flows.

TABLE I: Simulation parameters

# of endpoints u	200
Simulation time interval T	20000
Mobility distribution	Power-law distribution
Power-law exponent α	[0, 6]
Number of popular locations m	$[1, 5]$
Avg. mobility & content request rate $\mu(u)$, $\lambda(u)$	1, 1
Content request distribution	Uniform distribution
# GNS servers	$\sqrt{n}$
# Indirection servers	1
Unique GNS, Indirection server placements	400

A. Name-based forwarding

Fig. 4 shows the mean performance metrics in best-port and parallel-multicast forwarding for a range of mobility distributions with $\alpha \in [0, 6]$. The data plane metrics (TTC and forwarding traffic cost) are shown in Figs. 4(a) and 4(b), and the control plane metrics (update cost and update propagation cost) are shown in Figs. 4(c) and 4(d). Each datapoint corresponds to one of the mean metrics defined in Eqs. (3), (6), (9) and (11).

Best-port forwarding. In best-port forwarding (shown in the blue-triangle in Fig. 4), the performance metrics do not vary with a changing endpoint mobility distribution. Best-port incurs the lowest TTC and forwarding traffic cost, as the packet follows the shortest path from the source to destination, irrespective of the endpoint's mobility or workload distribution. However, best-port also has the highest control costs among all the forwarding strategies as the endpoint sends an update upon every mobility event irrespective of the mobility distribution. This is seen in the Sprint ISP in Fig. 4(c) where endpoint mobility results in an average of 150 of the 315 routers in the network receiving an update, affecting approximately half of the network.

> Insight 1: *Best-port offers the lowest TTC across all forwarding strategies, at the expense of a significantly higher update cost. Nearly half of the routers in the Sprint ISP are updated upon every mobility event, irrespective of the mobility distribution.*

Parallel-multicast forwarding. Fig. 4 also shows the data and control plane metrics in the Sprint ISP for parallel-multicast with up to five popular locations ($m = 5$). As the probability of being at a popular location increases (increase in α), the data and control plane costs, and TTC decrease. Recall that a router always forwards the first packet along all paths to known popular locations of an endpoint; if the endpoint is not found at a popular location, then the router forwards the first packet along the path to the last updated unpopular location. Therefore, initially, for small values of α, where the endpoint mobility distribution is closer to a uniform probability distribution, the endpoint has a higher probability

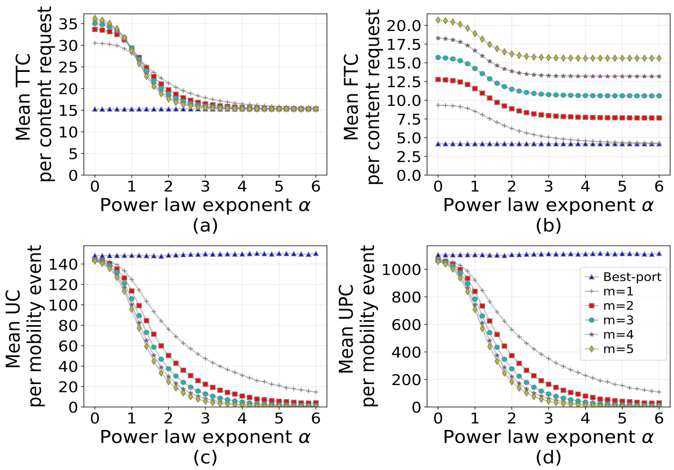


Fig. 4: Mean TTC in (a) and mean forwarding traffic cost (FTC) in (b) are shown per content request for the first packet of data. In the control plane, the mean update cost (UC) in (c) and mean update propagation cost (UPC) in (d) are shown per mobility event. The legend shows plot lines for best-port and parallel-multicast with m popular locations. The performance metrics are shown against the power-law exponent α of the endpoint's mobility distribution.

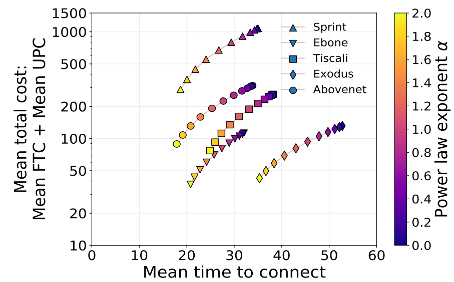


Fig. 5: Mean total cost vs. mean TTC: The mean total cost is the sum of the mean forwarding traffic cost (FTC) and the mean update propagation cost (UPC) is shown against the mean time-to-connect (TTC) in five ISPs for parallel-multicast forwarding with (m = 3) popular locations and endpoint mobility distribution with a power-law exponent range $\alpha \in [0, 2]$.

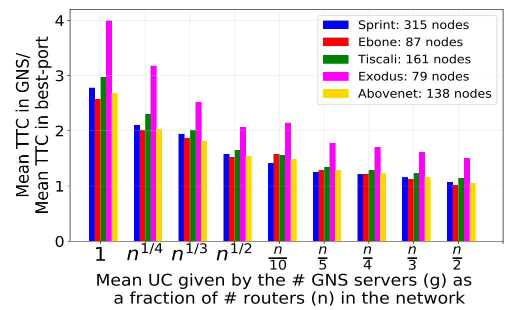


Fig. 6: Fraction of the mean TTC in the GNS to the mean TTC in best-port vs. mean update cost (UC) shown for for five ISP topologies.. The mean update cost in the GNS is equal to the # GNS servers g in the network.

of being at an unpopular location, and incurs higher forwarding traffic cost, and larger TTC. As α increases, the probability of being at a popular location also increases, and the forwarding traffic cost, and TTC decrease, as shown in Figs. 4(a) and 4(b). Figs. 4(c) and 4(d), show that the control costs in parallel-multicast also decrease as α increases since there is no update (update propagation) cost incurred when the endpoint is at a popular location.

In Fig. 5 we show the trade-off in the parallel-multicast strategy with $m = 3$ popular locations with the mean total cost (12), plotted against the mean TTC. The color bar shows the endpoint mobility distribution with a power-law exponent range $\alpha \in [0, 2]$. Fig. 5 shows that the mean total cost and mean TTC decrease with increasing α and the mean total cost falls by approximately 60% from $\alpha = 0$ (where the cost is equal to best-port) to $\alpha = 2$, across all the ISPs. As discussed in Section IV-B, we only show results for $\alpha \in [0, 2]$, as the endpoint tends to stay at the most popular location with a probability $> 70\%$ for $\alpha > 2$. We found similar results showing a decrease in total cost with decrease in TTC, for $m \in [1, 5]$ and $\alpha \in [0, 6]$.

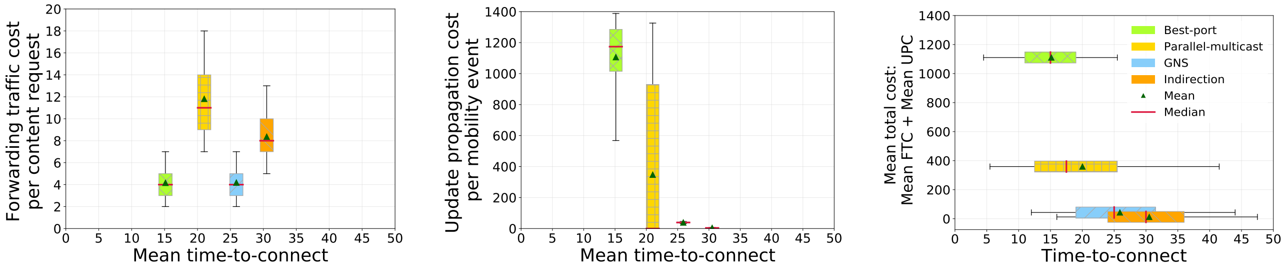
Insight 2: *Parallel-multicast offers up to 60% reduction over best-port in mean total cost by leveraging the fact that an endpoint has a high probability of being at a small subset of locations but trades off approximately $m\times$ increase in forwarding traffic cost on the first packet of data, where m is the number of popular locations.*

B. GNS: Optimizing the number of servers for update cost vs. TTC trade-off

In our instantiation of the GNS-based approach, we have assumed that the GNS server placements are uniformly distributed in the network, and that every GNS server receives an update upon endpoint mobility. We ran 400 simulation runs with different randomly generated GNS server placements.

To find the most suitable balance of TTC to update cost, in Fig. 6 we show the update cost on the x-axis, which in the GNS is equal to the number of servers that undergo an update due to endpoint mobility, and can be expressed as $g = \mathrm{E}[\mathrm{UC}](\mathrm{GNS})$. On the y-axis we show the fraction of the mean TTC in GNS over the mean TTC in best-port. TTC in the GNS is given by the round trip time to the closest GNS server from the CN, in addition to the time to reach the mobile endpoint from the CN, as shown Fig. 2(b). We chose to compare the TTC in GNS to the TTC in best-port as we have shown in Insight 1 that best-port has the lowest TTC of all the forwarding strategies.

Fig. 6 shows that the biggest reduction of TTC in the GNS is when g is a small fraction of the number of nodes in the network, and that there is decreasing marginal utility in increasing g. For example, when the number of GNS servers is increased from $g = 1$ to $\sqrt{n}$ servers, the mean TTC in the GNS reduces from an approximately three-fold increase over the mean TTC in best-port to within 1.5 times the mean TTC in best-port. However, a large increase in GNS servers from $g = \sqrt{n}$ to $n/4$ results in a very small decrease in mean TTC — from 1.5 times to 1.2 times that of mean TTC in best-port. In our performance comparison across mobility approaches, we choose $g = \sqrt{n}$ which results in a mean TTC of approximately 25 across all the ISPs and is approximately 1.5 times the mean TTC in best-port.



(a) Forwarding traffic cost (data plane) vs. mean time-to-connect (TTC).

(b) Update cost (control plane) vs. mean time-to-connect (TTC).

(c) Mean total cost: mean forwarding traffic cost (FTC) + mean update propagation cost (UPC) vs. Time-to-connect (TTC).

Fig. 7: Cost comparison across forwarding strategies in the Sprint ISP: Box plots for best-port, parallel-multicast, GNS and, indirection are shown with the mean and median values. The ends of the box show the inter-quartile range and the whiskers represent the 5th and 95th percentiles.

Insight 3: *In a GNS-based strategy, much of the reduction in TTC comes from a small number of GNS servers, and there is decreasing marginal utility of adding more GNS servers. We show that across five different ISP topologies, $g = \sqrt{n}$ achieves a mean TTC of approximately* 1.5 *times the mean TTC in best-port.*

C. Indirection: Effect of length of packet flows

In indirection there is only one indirection server that is responsible for routing data from the CN to the endpoint, and therefore indirection has the smallest number of servers that receive an update of any forwarding strategy, with $\mathrm{E}[\mathrm{UC}](\text{Indirection}) = 1$. However indirection suffers from an increase in forwarding traffic cost and TTC over best-port since the packets are sent from the CN to the indirection server, and then from the indirection server to the endpoint. We note that unlike the other forwarding strategies that only incur forwarding traffic cost and TTC on the first packet, indirection incurs an increased data plane cost in terms of links traversed and latency on *every subsequent packet* as well.

Insight 4: *Indirection has the lowest control cost but unlike the other forwarding strategies, indirection would incur the highest data plane cost in terms of links traversed and latency to the endpoint for flows greater than one packet irrespective of the endpoint mobility distribution, network topology, and server placement.*

VI. Results: Cost comparison across mobility approaches

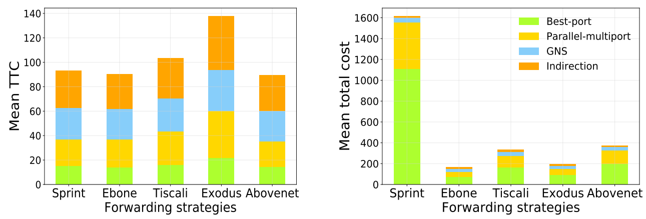
In this section, we compare the costs of each of the four forwarding strategies — best-port, parallel-multicast, GNS, and indirection. Our goal is to identify which of the four forwarding strategies can handle frequent endpoint mobility, while achieving a small forwarding traffic cost, a small update propagation cost, and a low TTC. In Figs. 7 - 8 we use the following parameters to evaluate the forwarding strategies — for the parallel-multicast strategy we choose an $\alpha = 1.8$ and $m = 3$, for which we found the smallest total cost among all values of α and m in Section V-A. For the GNS, based on our discussion of update cost vs. TTC trade-off in Section V-B, we chose the number of servers to be $g = \sqrt{n}$.

Fig. 7 shows the cost comparison across forwarding strategies in the Sprint ISP. The box plots show the mean and median values, the inter-quartile range, and the whiskers represent the 5th and 95th percentiles. Fig. 7(a) shows that the mean forwarding traffic cost is the highest in the parallel-multicast strategy, as the forwarding traffic cost increases with the number of popular locations as shown in Insight 2. Fig. 7(b) shows that the mean update propagation cost is the highest for best-port as we have shown in Insight 1. The inter-quartile range in parallel-multicast is large due to the following reasons; when the endpoint is at a popular location (for $\alpha = 1.8$, $m = 3$ the probability of being at a popular location is approximately 0.8) the update propagation cost is zero, however when the endpoint is not at a popular location, there is a significantly higher update propagation cost.

Fig. 7 shows that best-port forwarding and indirection represent the two extremes of the data and control plane trade-off — best-port has the lowest TTC (shown in Fig. 7(c)) and the highest update cost (shown in Fig. 7(b)), while indirection has the highest TTC and the lowest update cost. Fig. 7(c) shows that the TTC in both parallel-multicast and GNS is approximately 1.5 times that of best-port, and both these strategies offer a better TTC than that in indirection (a two-fold increase). After indirection, GNS has the next best total cost, approximately 6 times that of indirection, while both the name-based forwarding strategies, parallel-multicast and best-port, have a significantly higher total cost (approximately 35 – 88 times the cost of indirection).

We have shown above that in the Sprint ISP, best-port forwarding and indirection represent the two extremes of the data and control plane trade-off and have very small costs in either the data or the control plane but not in both. The GNS however, has a mean TTC which is approximately 1.5 times that of best-port (which has the smallest TTC), while the mean total cost is approximately 6 times that of indirection (which has the smallest total cost) and therefore is cost-effective in both the control and the data plane.

Fig. 8 shows the mean total cost and the mean TTC for five



(a) Stacked bar plot of the mean TTC in four forwarding strategies.

(b) Stacked bar plot of the mean total cost in four forwarding strategies.

Fig. 8: Stacked bar plot showing the mean TTC and the mean total cost in five ISP topologies with inputs, GNS with $g = \sqrt{n}$, end-point mobility distribution with $\alpha = 1.8, m = 3$ popular locations.

different ISPs across the four forwarding strategies. Similar to the Sprint topology, Fig. 8(a) shows that best-port has the lowest TTC and indirection has the highest TTC across all the forwarding strategies, for all the ISPs. We show in Fig. 8(a) that the TTC in GNS and parallel-multicast are comparable across all five ISPs given an endpoint mobility distribution with $\alpha = 1.8$, GNS with $g = \sqrt{n}$ servers and parallel-multicast with $m = 3$, popular locations. Fig. 8(b) shows that after indirection, the smallest mean total cost is for the GNS and we found that across ISPs the mean total cost in the GNS is approximately 6 times that of indirection. In Ebone and Exodus, the smallest ISPs (with the number of nodes in the network $n = 79$ and 87 respectively), the total cost in parallel-multicast is comparable to the total cost in the GNS. But across all the ISPs (Sprint with $n = 315$ nodes to Exodus with $n = 79$ nodes, the GNS has the lowest total cost after indirection, and TTC comparable to parallel-multicast.

Insight 5: *Across all ISP topologies, a GNS-based strategy with $g = \sqrt{n}$ uniformly distributed servers, achieves a mean TTC that is approximately* 1.5 *times that of best-port (that has the smallest TTC) and a mean total cost that is approximately* 6 *times that of indirection (that has the smallest total cost). We show that a GNS-based strategy has the most suitable balance of data and control plane costs across all the forwarding strategies.*

VII. Discussion

In this section, we discuss effects of varying workload and mobility, and comment on the insights shown in Sections V and VI.

Effect of endpoint mobility distribution on parallel-multicast. We have used a power-law distribution for a range of exponents to vary the endpoint distribution from a uniform distribution, to a zipfian distribution, and in the limit to a point-delta distribution. The power-law distribution however cannot model all classes of distributions, for example, it cannot model a step-distribution. Our results for the parallel-multicast strategy that are shown in Section V and VI however, are not dependent on the particular mobility distribution but on the cumulative probability of being at a popular location m (refer Fig. 3).

Ratio of mobility to request events. We have used a 1:1 ratio of content request to mobility events in our simulations. The mean data and control plane metrics as defined in Section III, are averaged across the total number of demand and mobility events respectively, and are therefore not dependent on the relative rate of content requests to mobility events. Individually analyzing the dependence of aggregate cost metrics in the data (control) on content request (mobility events) is beyond the scope of this paper and would be an interesting area for future research.

TTC reduction and limits on the GNS. We have assumed a uniform content request distribution and uniform server placements in the GNS. A natural question would be if our conclusions for the GNS were to change with different workload distributions and different server placements. A demand-aware server placement in a GNS-based approach has been shown to only reduce update cost and TTC over uniform placement and workload distributions [12]. However, irrespective of the number of GNS servers, or server placements, the GNS still incurs an irreducible TTC due to look-up on the first packet of data, and therefore even in the limit where a GNS server is colocated with every router, the TTC in GNS would still induce a small processing delay compared to the TTC in best-port.

Based on our results, we next explain three different scenarios where each of the three mobility approaches would excel. If there are a very small number of packets in a connection, and TTC is not of concern, then the best mobility approach would be indirection, since it has a very low control overhead for a small number of flows. If on the other hand, TTC were the most important metric of interest, and control bandwidth is expendable, then best-port would be the best mobility approach, and in either of these two extreme cases, a GNS would not be the best fit. However, in the scenario where a small inflation to the TTC is acceptable but control bandwidth is expensive, a GNS-based approach would offer the most suitable balance of control costs and TTC; our position is that this latter scenario is more representative of real-world concerns.

Our results show quantitatively that proposals for indirection based schemes [7], [9], [22] suffer from long delays with the mean TTC showing a two-fold increase over best-port as seen in Fig. 8. We show that even with parallel-multicast, a name-based forwarding strategy [20], the total cost can be up to 20% higher than the GNS and we show that the total cost in the GNS-based approach is no worse than parallel-multicast across all ISPs. We find that a GNS-based approach can provide the most suitable balance of costs across a range of mobility distributions and ISPs for any one of the location-independent architectures [16], [17], [20]. Recent work in MobilityFirst [12] and NDN [21], has also advocated for a name-resolution service to handle endpoint mobility, which we have validated based on a quantitative evaluation using real-world ISP topologies and endpoint mobility distributions.

VIII. Related work

Our goal is to quantitatively evaluate distinct architectural approaches towards handling mobility. We refer to Wroclawski [18] who distinguishes between architecture, and its instantiation, as follows; "[Architecture is a set of] high level design principles that guide the technical development of a system [which is] a realized instantiation that meets the design principles of the architecture". To compare the three distinct architectural approaches to handling mobility; name-based forwarding, name-resolution, and indirection; we have instantiated representative forwarding strategies. While quantitative evaluations of FIA projects have shown improvement over state-of-the-art mobility handling approaches [4], [12], [14], [20], our goal is to perform a cross-architectural comparison using representative instantiations and *not* to improve upon specific instantiations of each of these mobility approaches.

Previous efforts in quantitative cross-architectural evaluations have focussed on the trade-offs of edge-based caching over pervasive caches [5], and on the power-efficiency trade-off in larger packet sizes that maintain state versus smaller packets that include routing table lookups [1]. Both these works do not consider the impact of mobility on architecture. The quantitative comparison of mobility approaches in [6] shows that pure name-based forwarding may not be suitable for highly mobile content but does not present a comparison across mobility approaches and importantly does not analyze the impact of cost trade-offs on TTC — arguably a metric of fundamental importance in any Internet architecture [13].

IX. Conclusion

Mobility is pervasive in the Internet, and Future Internet Architecture efforts must choose mobility approaches that can handle frequent endpoint mobility while achieving a low update cost, a low forwarding traffic cost, and a small TTC. Our work provides a quantitative comparison of three distinct mobility approaches — (a) name-based forwarding, (b) indirection, and (c) a global name service (GNS) using forwarding traffic cost and TTC in the data plane and update cost and update propagation cost in the control plane. We have shown that by leveraging the fact that an endpoint has a high probability of being at a small subset of locations, there exist name-based forwarding strategies that can reduce control costs compared to simple best-port forwarding by up to 60%. We have shown that there exist name-based forwarding strategies that can achieve TTC comparable to the TTC in GNS-based strategies. Finally, we have shown that in GNS-based strategies the TTC is 1.5 times the TTC in best-port forwarding, for a scalable data and control plane cost. We find that a GNS-based approach can provide the most suitable balance of costs across a range of mobility distributions and ISPs for any one of the location-independent architectures [16], [17], [20].

Future Work. Our work is an important step in providing a quantitative comparison of mobility approaches across location-independent architectures. In future work, we plan to investigate the effects of different routing policies including BGP, different workload distributions, and the effect of caching on endpoint mobility.

References

[1] C. Chen, D. Barrera, and A. Perrig. Modeling Data-Plane Power Consumption of Future Internet Architectures. In *Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on*, pages 149–158. IEEE, 2016.

[2] CISCO. Top 5 surprises from the 2017. Mobile VNI Study. http://blogs.cisco.com/sp/top-5-surprises-from-the-2017-mobile-vni-study, February 2017.

[3] D. D. Clark, K. Sollins, J. Wroclawski, and T. Faber. Addressing reality: an architectural response to real-world demands on the evolving internet. In *Proceedings of the ACM SIGCOMM Workshop on Future Directions in Network Architecture*, FDNA '03, pages 247–257. ACM, 2003.

[4] C. Dannewitz, M. D'Ambrosio, and V. Vercellone. Hierarchical dht-based name resolution for information-centric networks. *Computer Communications*, 36(7):736–749, 2013.

[5] S. K. Fayazbakhsh, Y. Lin, A. Tootoonchian, A. Ghodsi, T. Koponen, B. Maggs, K. Ng, V. Sekar, and S. Shenker. Less pain, most of the gain: Incrementally deployable icn. *SIGCOMM Comput. Commun. Rev.*, 43(4):147–158, Aug. 2013.

[6] Z. Gao, A. Venkataramani, J. F. Kurose, and S. Heimlicher. Towards a quantitative comparison of location-independent network architectures. In *Proceedings of the 2014 ACM Conference on SIGCOMM*, pages 259–270, 2014.

[7] J. Lee, S. Cho, and D. Kim. Device mobility management in content-centric networking. *IEEE Communications Magazine*, 50(12), 2012.

[8] NSF Future Internet Project. `http://www.nets-fia.net/`, 2010.

[9] C. E. Perkins and D. B. Johnson. Route optimization for Mobile IP. *Cluster Computing*, 1(2):161–176, 1998.

[10] D. Raychaudhuri, K. Nagaraja, and A. Venkataramani. MobilityFirst: A Robust and Trustworthy Mobility-Centric Architecture for the Future Internet. *ACM SIGMOBILE Mobile Computing and Communications Review*, 16(3):2–13, 2012.

[11] J. Saltzer. On the naming and binding of network destinations. *Internet Engineering Task Force (IETF)*, 1993.

[12] A. Sharma, X. Tie, H. Uppal, A. Venkataramani, D. Westbrook, and A. Yadav. A global name service for a highly mobile internetwork. In *Proceedings of the 2014 ACM Conference on SIGCOMM*, number 4, pages 247–258, 2014.

[13] A. Singla, B. Chandrasekaran, P. Godfrey, and B. Maggs. The internet at the speed of light. In *Proceedings of the 13th ACM Workshop on Hot Topics in Networks*, page 1. ACM, 2014.

[14] V. A. Siris. Popularity-aware intra-domain mobility management. In *Proceedings of the Workshop on Mobility in the Evolving Internet Architecture*, MobiArch '17, pages 13–18, New York, NY, USA, 2017. ACM.

[15] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with Rocketfuel. In *Proceedings of the 2002 ACM Conference on SIGCOMM*, pages 133–145, 2002.

[16] D. Trossen and G. Parisis. Designing and realizing an information-centric internet. *IEEE Communications Magazine*, 50(7), 2012.

[17] A. Venkataramani, J. F. Kurose, D. Raychaudhuri, K. Nagaraja, M. Mao, and S. Banerjee. MobilityFirst: A Mobility-Centric and Trustworthy Internet Architecture. *SIGCOMM Comput. Commun. Rev.*, 44(3):74–80, July 2014.

[18] J. Wroclawski. All Hat No Answers: Some Issues Related to the Evaluation of Architecture. Spring, 2013 NSF FIA PI meeting, Salt Lake City,, 2013.

[19] S. Yang, J. Kurose, S. Heimlicher, and A. Venkataramani. Measurement and modeling of user transitioning among networks. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 828–836. IEEE, 2015.

[20] L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, P. Crowley, C. Papadopoulos, L. Wang, B. Zhang, et al. Named Data Networking. *SIGCOMM Comput. Commun. Rev.*, 44(3):66–73, 2014.

[21] Y. Zhang, A. Afanasyev, J. Burke, and L. Zhang. A Survey of Mobility Support in Named Data Networking. In *Computer Communications Workshops (INFOCOM WKSHPS), 2016 IEEE Conference on*, pages 83–88. IEEE, 2016.

[22] Y. Zhang, H. Zhang, and L. Zhang. Kite: A mobility support scheme for NDN. In *Proceedings of the 1st international conference on Information-centric networking*, pages 179–180. ACM, 2014.