Abstract—Despite the Internet’s ability to recover from link failures, the process is laborious. Scalable One-hop Source Routing (SOSR) hastens the recovery, without complex routing algorithms, by routing around failures via indirect paths created using randomly-selected intermediate end-nodes. Even with only 39 intermediaries available, SOSR effectively masks out 89% of Internet core link failures. However, SOSR is restricted in 2 ways: (1) it employs only end-node intermediaries, thus traffic utilizing the indirect paths has to detour to the intermediaries at the Internet edges en-route to their destinations and (2) although the number of intermediaries can be scaled up to increase the indirect paths available, failure-masking rate is unlikely to improve much since it is difficult to find an indirect path that can mask a given failure, even if one exists, through random intermediary selection.

To overcome these, we introduce the “AI-RON-E” prophecy—a loosely-federated infrastructure consisting of clients, “oracles” and Internet routers that are all One-hop Source Routing (OSR) aware. OSR routers can act as intermediaries to redirect traffic, avoiding the need to detour to the Internet edges, thereby shortening indirect paths formed. To increase the probability of finding a failure-masking intermediary from ineffective ones, AI-RON-E clients select intermediaries from partial views of the Internet obtained from the oracles and apply heuristics to filter out “bad” candidates from those views during the selection process.

By hypothetically analyzing around 4500 link failures in 375 paths, we can conclusively foretell, even in the absence of the yet-to-be-built AI-RON-E infrastructure, that indeed AI-RON-E can be deployed at Internet-scale to seek out indirect paths faster and masks more link failures while offering shorter hop-count indirect paths at the expense of a small cache of path information.

I. INTRODUCTION

The Internet routing’s meticulous design ensures that it can support the over-arching goal of delivering high availability [1]. However, routing convergence, which is required to mask link1 failures, is not rapid enough to render them insusceptible to downtime [2].

Nonetheless, the possibility of achieving high availability in the Internet is evident, since multiple paths have been reported to exist between a given pair of source and destination (known as path diversity) [3]. Furthermore, Resilient Overlay Network (RON) [4] and Scalable One-hop Source Routing (SOSR) [5] have proven that overlay routing can utilize the path diversity to mask link failures and link congestion, eliminating the need for rapid routing convergence.

1In this paper, paths and links have different meanings, a path consists of many links.

Even with only 39 intermediaries, by randomly selecting 4 of them to create 4 indirect paths, SOSR can mask up to 89% of Internet core link failures. Despite the effectiveness, there are two points to note: (1) the intermediaries are restricted to end-nodes, systems located on the Internet edges, which results in traffic having to detour out to the periphery en-route to their destinations and (2) the number of intermediaries are limited, which may lead to a lack of possible indirect paths for failure-masking as reported by SOSR; 31% of link failures cannot be masked even if all the 39 intermediaries are used. Even with more intermediaries, to increase path diversity, failure-masking rate is unlikely to improve since random intermediary selection may not identify a failure-masking indirect path among a huge number of ineffective ones. This notion is backed by SOSR statistics showing that for some 5% of paths, a randomly selected intermediary has only 0.25 probability of successfully masking a failure.

Based on these observations, we propose the AI-RON-E prophecy—a loosely-federated infrastructure consisting of clients, “oracles” and Internet routers that are all One-hop Source Routing (OSR) aware: clients request the oracles’ partial view of the Internet in order to make better decisions on selecting intermediaries that can provide indirect paths to mask failures. The contributions of AI-RON-E are three-fold. Firstly, AI-RON-E introduces non-edge intermediaries (OSR routers) and this translates to traffic taking only a short detour to these intermediaries en-route to their destinations resulting in shorter hop-count indirect paths compared to those of SOSR. Secondly, we propose the AI-RON-E intermediary selection algorithm that, with meager path information caching, out-performs SOSR in link failure-masking ability—both in failure-masking rate and speed in discovering failure-masking indirect paths. In addition, AI-RON-E retains low resource utilization, i.e., network probes, despite having to choose from intermediary candidates of many magnitudes higher than SOSR, with each varying in locality from near the Internet edges to deep inside the core. Lastly, we employ a hypothetical link failure masking evaluation approach, which is unlike conventional empirical link failure monitoring utilized by existing related works such as RON and SOSR that relies on actual failure occurrences for evaluation. This methodology enables us to: (1) foretell the effectiveness of AI-RON-E even

Akihiro Nakao
University of Tokyo
7-3-1 Hongo Bunkyo-ku
Tokyo 113-0033, Japan
Email: nakao@iii.u-tokyo.ac.jp

Soon Hin Khor
University of Tokyo
7-3-1 Hongo Bunkyo-ku
Tokyo 113-0033, Japan
Email: khor.soon.hin@iii.u-tokyo.ac.jp

1 Pronounced as “irony”
though its infrastructure does not yet exist and (2) minimize
link failure effects by finding alternate paths for as many links
as possible prior to failure (see Section V). Before concluding,
we also describe how we can employ source IP spoofing
to enable the existing Internet to support AI-RON-E in the short-
run as we await for it to gain adoption in a large scale.

II. DESIGN GOALS

AI-RON-E aims to provide the following:

Reduced hop-count indirect paths With only end-nodes
available as intermediaries, traffic detouring to them en-route
to the destinations has to travel from one edge of the Internet
(source) to another edge (intermediary) before arriving at the
final edge (destination). This network edge-to-edge-to-edge
routing is superfluous and can be reduced.

Better failure-masking ability than SOSR AI-RON-E
strives to improve failure-masking ability from two angles—
mask more link failures and find failure-masking indirect paths
quicker compared to SOSR, i.e. a failure-masking intermediary
can be found with fewer tries.

Light-weight intermediary selection algorithm From a
large pool of candidates, the intermediary selection algorithm
has to, within a few attempts, successfully find an intermediary
that can mask a given link failure using minimal resources such
as network probing for path discovery.

Internet-scale infrastructure and utilization of Internet
path diversity We want AI-RON-E to deliver an infra-
structure that can support a large Internet community and also
enable Internet path diversity to be better utilized than it
currently is.

Incremental deployment Although the AI-RON-E prophecy
foretells a future where all Internet routers are OSR capable,
a partial deployment is definitely useful thus an incremen-
tal deployment makes sense. Moreover, AI-RON-E requires
making modifications to the Internet routers and this has tra-
ditionally faced steep resistance. Besides progressive changes,
incremental deployment enables users to opt-in with resource
investments that suits their technology-risk appetite with the
option to scale up the infrastructure, requiring minimal re-
work when the need arises. The appeal of incremental de-
ployment has been shown to work in practice, e.g., in Planet
Lab [6] and peer-to-peer networks such as BitTorrent [7].

III. ARCHITECTURE

In the AI-RON-E infrastructure (see Figure 1), any Internet
system, e.g., servers, routers, etc., may play any/all combina-
tion of the 3 roles: “oracles” that provide their partial view-
point of the Internet (similar to existing traceroute servers [8]),
intermediaries (OSR routers) that redirect traffic or/and clients
that utilize the infrastructure to mask link failures.

Once an AI-RON-E client (P) detects a failure (c-e) in its
direct path to a destination (Q), it will choose an intermediary
to bypass the failure by consulting a random oracle (ORC1)
to obtain its “oracle path”—the path the oracle takes to reach
the same destination, e.g., ORC1’s oracle path is ORC1-
k-d-f-g-h-j-Q. Nodes along the oracle path, excluding the
destination (ORC1-k-d-f-g-h-j) are candidate intermediaries
that P can choose from since they are OSR routers that
are capable of re-directing packets. For example, with d as
an intermediary, P can bypass c-e using the indirect path
constructed of client-intermediary path P-a-b-d and intermediary-destination path d-f-g-h-j-Q.

The architecture is designed with the following properties
to fulfill the design goals described in Section II and in
Section IV we show that AI-RON-E can indeed achieve them.

OSR routers further from the Internet edges offer shorter
indirect paths With only end-nodes capable of redirecting
traffic as in SOSR, in Figure 1, client P only has ORC1
available as intermediary and the indirect path formed would
have 12 hops (P-a-b-d-k-ORC1-k-d-f-g-h-j-Q). In AI-RON-E,
an OSR router that is further from the Internet edge, e.g.,
d, can be used as an intermediary thereby creating a shorter
indirect path of 8 hops (P-a-b-d-f-g-h-j-Q).

Oracle-based heuristic intermediary selection algorithm
for better failure-masking ability To formulate an interme-
diate selection algorithm that rivals the failure-masking rate
of SOSR, we borrow SOSR’s random intermediary selection
algorithm and add a twist to it—since random end-node
intermediaries can provide good failure-masking performance
by virtue of their location at the Internet edges that gives
rise to indirect path disjointedness from the direct path, our
hunch is that by selecting a random node along an end-node
intermediary-destination path as an intermediary, AI-RON-E
can deliver similar effectiveness.

To compensate for the possible reduction of path disjoint-
edness resulting from selection of an intermediary further
away from the Internet edges, AI-RON-E has heuristics to
remove candidates that cannot contribute to failure-masking;
candidates in an oracle path that also appear in the direct path
(source-destination) are filtered out. For example, in Figure 1,
ORC1-k-d-f-g-h-j are in ORC1’s oracle path but we exclude
h and j as candidates. Since they lie along the direct path
(P-a-c-e-h-j-Q), if they are used as intermediaries, P will most
likely reach them via the direct path that includes the assumed
broken link c-e resulting in bypass failure.

Fig. 1: AI-RON-E architecture consists of OSR routers (non end-node
intermediaries), oracles and clients embedded with logic to consult oracles
to aid indirect path construction.
This avoids using adjacent candidates that may have similar client-intermediary paths resulting in “fate-sharing”; both candidates succeeding or failing. For example, in Figure 1, P reaches adjacent nodes f and g through the paths P-a-c-e and P-a-c-e-f respectively, resulting in client-intermediary paths of both the candidates failing to mask the failure at c-e. Through “bad” candidates filtering, AI-RON-E naturally seeks out intermediaries that has a higher chance of masking failure quicker than otherwise. However, note that filtering may result in some oracle paths producing only 1 or no candidate intermediary.

**Caching oracle paths to retain light-weightedness** We use SOSR’s light-weightedness as a benchmark for AI-RON-E. SOSR sends N packets and waits for at most N round-trip-times (RTT), before failing or succeeding in masking failure, where N is the number of intermediaries selected. A higher N instinctively provides a better masking rate, but it also incurs more network overhead. With the law of diminishing returns applying, for SOSR, N=4 is ideal [5].

Oracle paths form the basis for AI-RON-E to select better intermediaries but are expensive in terms of network probes required to obtain them. Thus, AI-RON-E retains SOSR’s light-weightedness by trading-off some storage, i.e., requiring each AI-RON-E client to maintain previously acquired oracle paths in a cache of size 0.5N to 1.5N. These cached oracle paths do not expire. Clients always re-use their own cached oracle paths to find candidate intermediaries to bypass any broken links encountered. Expense oracle consultation is only triggered when N intermediaries cannot be provided after exhausting all cache entries. The newly obtained oracle path will replace the least used entry in the cache. As shown in Section IV-B2, despite re-using the same few cached oracle paths, there are enough intermediaries along those paths to produce diverse indirect paths to mask link failures of paths to various destinations.

**Huge pool of widely distributed OSR routers and random oracle/intermediary selection algorithm delivers scalability and helps path diversity utilization** The huge pool of OSR routers enable AI-RON-E to support a large user community. Moreover, since the pool comprises widely-dispersed OSR routers located at different depths from the Internet edges, AI-RON-E clients that obtain random oracle paths and then select random intermediaries within the oracle paths, will most likely utilize different intermediaries resulting in traffic being distributed evenly over the Internet, thereby utilizing path diversity better and reduce congested “hotspots”.

**Loose-coupling enables incremental deployment** Incremental deployment offers incremental benefits; with each OSR router introduced, it creates an extra possibility for AI-RON-E clients (1) to mask link failures or divert traffic from current congested “hotspots” and (2) better utilize Internet path diversity. Incremental deployment is possible by making AI-RON-E components—client, oracle and intermediary—loosely-coupled, i.e., a client can use any AI-RON-E node as an oracle or intermediary making bootstrapping into the AI-RON-E infrastructure extremely effortless and relatively decentralized since existing tools, e.g., traceroute from remote systems [9] or from the client itself can be used to discover routers and brief message exchanges between the client and discovered routers can determine whether the routers can act as AI-RON-E oracles/intermediaries. Moreover, destinations do not need to be AI-RON-E-aware. Thus, new AI-RON-E components can be “dropped” in at any time and any place to scale up the existing infrastructure.

**IV. Evaluation**

This section evaluates AI-RON-E’s link failure masking approach. Using our hypothetical link failure evaluation methodology, we can assess more link failures than a conventional empirical link failure monitoring approach can within a given period of time but with the assumption that only a single link within a path fails at a time. Through it, we can foretell the effectiveness of AI-RON-E even though the infrastructure has yet to exist. More specifically, we want to find out if AI-RON-E can achieve its design goals: (1) Does AI-RON-E provide shorter hop-count indirect paths than SOSR? (2) Can AI-RON-E mask link failures as well and as fast as SOSR even though the intermediaries have to be selected from a larger pool of candidates that are mostly further from the Internet edges? and (3) How light-weight is AI-RON-E’s resource utilization, specifically, cache storage?

**A. Evaluation Methodology**

We first describe the general architecture setup followed by a description of our hypothetical link failure evaluation approach and finally how we can simulate SOSR and AI-RON-E on this same setup for comparison purposes.

We use Planet Lab (PL) nodes in about 100 distinct locations as traceroute servers (oracles). When consulted, each can provide the traceroute from itself to a given destination. Next, we select 15 PL nodes (Asia - 3, US - 3, East Europe - 3, West Europe - 3, South America - 2 and Canada - 1) as traffic source and each source performs traceroute to the same set of 25 randomly-selected distinct PL locations to obtain direct paths. In our hypothetical link failure evaluation approach, for each link along each direct path, we check if there is an indirect path that can bypass it in the event of failure.

To simulate SOSR, any one of the 100 PL traceroute servers can be used as intermediaries and the indirect path is constructed by concatenating the source-PL path (obtained by traceroute from source) with the PL-destination path (obtained by consulting the PL traceroute server).

For AI-RON-E, the 100 PL nodes act as oracles and intermediaries can be any node along a randomly selected oracle’s oracle path, excluding bad candidates that are heuristically filtered out (see Section III). An indirect path can be constructed by concatenating the source-intermediary path (obtained by traceroute from source) with the intermediary-destination path (the sub-path of the oracle path obtained by consulting the selected oracle/PL node, starting from the intermediary).

For both SOSR and AI-RON-E, we conclude that an indirect path can bypass a failed link if it does not appear in the indirect path.
B. Results

In total, we evaluate approximately 15 x 25 = 375 paths comprising about 4500 possible link failures, which is more than 3 times the amount of link failures analyzed by SOSR for the paths destined to servers. From hereon, “SOSR-N” and “AI-RON-E-N” are used to refer to SOSR and AI-RON-E algorithms with N intermediaries selected, respectively.

1) Hop-count of indirect paths: As shown in Figure 2, AI-RON-E indirect paths always have shorter hop-counts than SOSR ones regardless of the number of intermediaries selected for use.

2) Speed and ability of link failure masking capability: Figure 3 shows which intermediary, a, out of N selected ones, successfully masks a failed link. Note that the plots for SOSR-4 and AI-RON-E-4 are discontinued after a=4 while for SOSR-12 and AI-RON-E-12, they end after a=12 since the algorithms themselves limit the number of intermediaries (N) being selected. These end-points, a=4 and a=12, also represents the number of links that can be masked by each algorithm variations after trying all N intermediaries. For N=4, AI-RON-E masks 69% of failures compared to SOSR’s 60% while for N=12, AI-RON-E masks 74% of failures versus SOSR’s 69%. In both cases, AI-RON-E masks more link failures than its SOSR counterpart.

Figure 4 also shows the higher failure-masking rate of AI-RON-E but from a different angle; it shows the location of links that can be masked by indirect paths. Link locations, l, are normalized within the path, i.e., l vary from 0 to 1 with l~0 indicating links near the source of a path and l~1 indicating links near the destination. As expected, links nearer the Internet core l~0.5 has better path diversity resulting in higher failure-masking rate, i.e. close to 100% for l=0.5 to 0.7.

Besides showing the masking rate of the algorithms, Figure 3 also shows that AI-RON-E can seek out indirect paths faster than SOSR between 6-12% of the time. For instance, AI-RON-E-4 can find 52% of failure-masking indirect paths by just using the 1st intermediary while SOSR-4 can only manage to find only 40%. The filtering of bad candidates by AI-RON-E obviously aids finding an intermediary that can mask a given failure quicker.

3) Intermediary selection algorithm resource consumption: Figure 5 allays the fear one may have about the size of the oracle path cache required by each AI-RON-E client. Since we select at most 2 candidates from each oracle path, for AI-RON-E-N, the minimum number of oracle paths needed to
find $N$ intermediaries is $N/2$. This explains why the CDFs for AI-RON-E-4, 8 and 12 start at 2, 4, and 6 respectively on the x-axis. However, after filtering out bad candidates, some oracle paths are left with no candidates. Under those circumstances, more than $N/2$ oracles paths will be required to produce the number of intermediaries specified by the algorithm.

As $N$ increases, the number of cached oracle paths, thus the cache size, required to provide those intermediaries increases. However, the plots show that to provide $N$ required intermediaries to 70-80% of links, the upper-limit of the cache size is bounded (“knee” of the plots); a meager cache size of 0.5$N$ to 1.5$N$ is sufficient. For the remaining 20-30% of links, it is extremely challenging to provide the required number of intermediaries because those links have low path diversity thus more resources, i.e., cache storage or network probes has to be sacrificed to find candidate intermediaries.

V. DISCUSSION

A. Minimizing link failure effects

The hypothetical approach to analyzing link failures without them actually occurring gives rise to AI-RON-E’s ability to minimize link failure effects prior to link failures, i.e., once a direct path is established, AI-RON-E immediately attempts to find indirect paths to bypass each link along the direct paths, in the background in preparation for link failures. Moreover, these indirect paths may also be used concurrently to enable multi-paths and increase transmitted bandwidth [12], [13]. However, minimizing link failure effect only makes sense for long-term connections.

B. Deployment issues and workaround

Using PL nodes in the existing Internet as non-edge intermediaries may not suffice since most are located at the Internet edges rendering them unsuitable. Therefore, we have to exploit a peculiarity in IP routing—source IP spoofing, to “coax” existing routers to function as non-edge intermediaries. Upon encountering a link failure, through oracle consultation, an AI-RON-E client selects an “non-edge intermediary” (this can be any conventional router) to mask the failure. To coax the conventional router to behave as an intermediary, the client crafts a packet by spoofing the source IP header with the intended destination IP and insert into the destination IP header, the IP of the router (intermediary) with TCP port number set to 0. When the packet reaches the router, it gets rejected with “ICMP port unreachable” and the router attempts to send it back to its “originator” which in this case, is the intended destination. By stuffing the data that we want to transmit into the IP options headers such as the Record Route, we ensure that the data is fully preserved and delivered by the triggered ICMP packet to the destination. We are aware that IP filters reject spoofed IP packets so this mechanism may not work for all nodes. However, it is a viable way to evaluate AI-RON-E without modifying existing routers.

VI. RELATED WORK

RON was the first work that sought to route around link failures through intermediaries. Despite its effectiveness, RON requires formation of a full-mesh among the intermediaries and constant network probing to seek out intermediaries that can mask link failures. Full-mesh maintenance and network probings are costly in terms of both network and system resources thus restricting the size of RON networks to a few tens. To overcome the costly network probings, SOSR randomly selects an intermediary without any heuristics, while AI-RON-E requires only a partial view of Internet to aid in intermediary selection.

Surprisingly, with just random intermediary selection, SOSR can effectively mask 89% of core link failures. AI-RON-E attempts to out-perform SOSR by introducing non-edge intermediaries and rely on a more intelligent intermediary selection algorithm to identify candidates that have a high chance of masking a given failure.

Cha et al. [14] propose a way to find optimal static locations for a given number of intermediaries within an Autonomous System (AS). Each possible intermediary location is given a rating based on how well it can mask the link failures for each source-destination pair within the AS. The set of intermediary locations that can minimize the cost function, which is the sum of the ratings of all the intermediary locations in the set, will be declared the optimal static intermediary locations for the AS.

This methodology requires a centralized system to find optimal intermediary placements using “difficult-to-obtain” comprehensive network topologies and minimizing the cost function is resource-intensive, thus it is probably not feasible for topologies as large as the Internet. Moreover, all the clients within the AS will contend for the same set of intermediaries to perform indirection. In contrast, AI-RON-E does not require a centralized omniscient system but instead empowers clients themselves to discover indirect paths by just obtaining partial views from oracles. By delegating indirect path discovery to the distributed clients themselves, AI-RON-E can function at Internet scale. In addition, AI-RON-E clients are not restricted to a set of intermediaries but have the entire routing infrastructure available to them as intermediaries with each client available.
probably using a different intermediary.

VII. CONCLUSION

We described the AI-RON-E prophecy that involves equipping every router with OSR capability and clients with logic to select failure-masking intermediaries through oracle consultation. The proximity of intermediaries to the Internet core offers shorter hop-count indirect paths of up to 30% while their sheer number enables AI-RON-E to exploit path diversity to find failure-masking indirect paths. AI-RON-E can improve SOSR failure masking rate by 6-8% and seeks out the indirect paths faster 6-12% of the time. All this come at the expense of a slightly more involved algorithm, i.e., filtering of bad candidates and the maintenance of an oracle path cache with 0.5N to 1.5N entries at each client.

VIII. ACKNOWLEDGMENTS

The author, Soon Hin Khor would like to express his gratitude to ISC\textsuperscript{2} for providing a grant to support this research. Both Soon Hin Khor and Akihiro Nakao are also grateful for the grant provided by Japan Science and Technology Agency under the CREST project.

REFERENCES