pap252 Review Details

pap252 Review Details

Title: VNET/P: Bridging the Cloud and High Performance Computing Through Fast Overlay Networking
Authors: Bridges, Cui, Dinda, Lange, Tang, Xia

Key for the below column headings: hide

Review Categories (higher is better):

mer: **OVERALL MERIT.** In your estimation, how would you rank this paper with respect to other papers that have been submitted to HPDC. If you are unfamiliar with HPDC submissions, how would you rank this paper with respect to papers that are submitted to a range of high-performance conferences of the same caliber (e.g., SC, ICS, NSDI, USENIX ATC).
[Bottom 50% of submitted papers, Top 50% of submitted papers, but not in the top 25%, Top 25% of submitted papers, but not in the top 10%, Top 10% of submitted papers, but not in the top 5%, Top 5% of submitted papers. Implicitly recommended for the best paper award]

conf: **REVIEWER FAMILIARITY.** How familiar are you with the area.
[I know nothing or almost nothing about this topic, I am somewhat familiar with this area, but I can't claim I am an expert, or, my work in this area is out of date, I am well versed in this area, but it isn't my direct area of specialty, This is my area]

Summary of reviews of pap252s1

Reviewer	mer	conf
Reviewer 1	Top 5% of submitted papers. Implicitly recommended for the best paper award (5)	This is my area (4)
Reviewer 2	Top 10% of submitted papers, but not in the top 5% (4)	This is my area (4)
Reviewer 3	Top 10% of submitted papers, but not in the top 5% (4)	This is my area (4)
Reviewer 4	Top 5% of submitted papers. Implicitly recommended for the best paper award (5)	This is my area (4)
Reviewer 5	Top 25% of submitted papers, but not in the top 10% (3)	This is my area (4)
Reviewer 6	Bottom 50% of submitted papers (1)	I am well versed in this area, but it isn't my direct area of specialty (3)
Averages:	3.7	3.8

Commitee Comments jump

Reviewer 1 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental concepts, or increased understanding, or practical relevance?

Currently, virtualization could achieve near-native performance while it is helpful to deploy and manage a large-scale computing system. So the virtualized system is expected to be adopted to run tightly-coupled HPC applications. Although virtual networking systems have sufficiently low overhead to effectively host loosely-coupled scalable applications, their performance is insufficient for tightly-coupled applications. This paper is focused on extending VNET to tightly-coupled environments, so that VMs hosting tightly-coupled HPC applications may be seamlessly migrated between distributed cloud resources, tightly-coupled supercomputing and cluster resources. Then VNET/P is proposed and implemented based on Palacios, in order to achieve near-native throughput high-speed Ethernet, and InfiniBand, etc.

Does this paper present convincing results?

Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?

VNET/P is implemented within the VMM, Palacios. A series of experiments are conducted with TCP and UDP microbenchmarks, MPI microbenchmarks, HPCC benchmarks, etc. These benchmarks could achieve a good performance, and the NAS parallel benchmarks generally achieve near-native performance on both 1 and 10 Gbps. The experiments on real platforms demonstrate the efficiency of VNET/P. So the method proposed in this paper is reasonable and feasible, and can be used to bridge future Clouds and HPC through this fast overlay networking.

Is this paper sufficiently well executed?

Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?

This paper focuses on a real problem, proposes a feasible solution, implements it, and evaluates its function in a believable manner. It is a well-written and thoroughly enjoyable paper.

I note that, in the experiments, the number of virtual cores of all VMs is usually less than the number of physical cores, and correspondingly the number of processes is less than the number of physical cores. When the number of processes of parallel benchmarks is nearly equal to the number of physical cores, how well VNET/P performs.

In addition, there are some writing errors. Page 3, “we we demonstrated”. Page 7, “Figure 5 show”.

Reviewer 2 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental concepts, or increased understanding, or practical relevance?

This paper focuses on solving the problem of portability in Could/HPC environments. In prior work, the authors developed VNET/U, a userspace implementation of the overlay that focused on transparency in a wide range of network environments. VNET/P, the subject of this paper, has the task of maintaining the functionality and benefits of the VNET project, but achieving native performance -- both latency and throughput -- in a software solution through tight integration with the VMM.

Does this paper present convincing results?

Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?

The paper presents a wide array of results, ranging from UDP/TCP microbenchmarks, to MPI microbenchmarks, muti-node HPCC benchmarks, and finally the NAS benchmark suite.

The results consistently show that for 1Gbps ethernet, VNET/P can deliver near-native performance at 1500 byte MTUs. For 10Gbps ethernet, VNET/P can deliver 60-70% of native throughput at 9000 byte MTUs, and 2-3 times the latencies (less than 200 microseconds). While these numbers could be better, they represent impressive performance for a software overlay routed infrastructure. Moreover, when considering real application scenarios, many applications -- less communication intensive, actually show near native performance, while the more communication intensive applications have corresponding performance of 60-70 percent of native.

Finally, the authors present a work-in-progress result - applying the same technology to work over an infiniband stack, transparently to the virtual machines.

Is this paper sufficiently well executed?

Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?

Update The biggest problem with execution seems to be a failure to adequately position this work with respect to the related work in virtualization and networking. A related work section and comparing in evaluation with relevant prior work (including VNET/U) would go a long way toward making this clear.

While I find that the paper is generally well written, executed, and evaluated, there are some concerns, mostly with the problem space. While portability is an important problem, and address flexibility is a potential solution, I fear that it is one not well matched to the concerns of an HPC environment.

For example, conventional wisdom is that clouds (and the overheads of virtualization and networking, and Ethernet in general) are unsuitable for many HPC applications due to their I/O in efficiency. On the plus side, VNET/P is comparing performance to Native speed, not virtualized speed, so 60-70% is not in addition to the cost of virtualization.

Additionally, though VNET/P is seeking to achieve native performance in high speed Ethernet networks, the VNET vision would allow applications to transparently migrate wherever. While this flexibility is potentially helpful, it has a lot of opportunity for abuse as well -- clearly migrating machines to a wide area for applications designed to run in the tightly coupled data center environment will have grave impacts. In a sense - VNET is "overkill" - the portability supports migration, yes, but a much broader degree of migration than such coupled applications might benefit from.

On a different concern, the paper consistently uses large packet sizes, and advocates using ever larger packet sizes. This is because VNET/P adds per-packet overheads, which is totally sensible. In this case, the HPC environment probably favors larger packet sizes, since large data can be packetized in large packets. However, the paper could benefit from empirical evidence that MPI-based HPC applications typically use larger packet sizes.

Despite these concerns, overall, I think the paper is good work.

Reviewer 3 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental concepts, or increased understanding, or practical relevance?

This paper describes VNET/P, an extension of previous work on VNET/U. VNET/U is a virtual network overlay which allows the owner of a guest VM to have full control over network configuration on their machine while providing location independence, and networking hardware independence. It does this by taking ethernet frames from the virtual guest NIC and encapsulating them in UDP packets. VNET/P extends this work by attempting to optimize it to achieve near native throughput and latency in high performance computing

environments, while still retaining the benefits of virtualization (separation, ease of migration, etc.).

Does this paper present convincing results?

Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?

VNET/P consists mainly of an extension to the VMM (the VNET/P Core) and a linux kernel module (the VNET/P Bridge). Much of its increased performance over VNET/U can be attributed to the fact that it is not implemented in user space, and therefore avoids kernel/user context switches that limit VNET/U's performance.

Routing and packet forwarding occur in the VNET/P core. Routing is based on mac addresses with a hash based cache system that allows for constant time lookups in the common case. One interesting aspect of VNET/P is the packet processing system. Packet dispatchers can get packets from virtual NICs in two different modes. One mode is optimized for latency and one for throughput. The latency mode (guest-driven) reacts to activity from the guest OS and handles packets as soon as the guest needs them to be. The throughput mode (VMM-driven) aggregates packets to handle multiple simultaneously and reduce per packet overhead; instead of the guest notifying the packet processor whenever it has data, the VMM polls the NIC and handles all available packets.

Is this paper sufficiently well executed?

Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?

Overall, VNET/P provides very near native throughput over 1Gbps and around 75% performance on 10G. Latency, on the other hand, tends to be around double the native in most cases, which could be considered disastrous for some HPC applications. The performance increases over VNET/U are huge, but at the cost of increased difficulty of deployment due to needing to modify the VMM and have a kernel module. In the "Future Work" section, the possibility of "injecting VNET/P directly into the guest" is discussed. This would theoretically further reduce overhead by allowing the guest to bypass the VMM and have hardware access. This of course would create additional complication since one must modify arbitrary guest OSes and has possible security concerns, but could be quite useful if it is able to ameliorate the latency issues that currently exist.

The ability to seamlessly migrate distributed cloud applications into a tightly coupled HPC environment is certainly a worthy pursuit, and the authors have described several interesting performance enhancements including multi-mode packet processing, and hash based mac address lookups (although one might call into question the scalability of this last). While more work remains, this paper represents significant forward progress from VNET/U in the field of virtual network overlays.

Reviewer 4 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental concepts, or increased understanding, or practical relevance?

This paper presents the design, implementation, and evaluation of VNET/P, an in-VMM, overlay-based virtual networking system (as part of the Palacios hypervisor platform). Overlay-based virtual networking is a powerful mechanism to realize virtual distributed/parallel computing, with strong isolation, portability, and recoverability. However, a challenge long faced by overlay-based virtual networking is the performance degradation - in terms of network throughput and latency - incurred by layer-2 virtualization and overlay network overhead. Yet its non-performance benefits (i.e., isolation, migration, and checkpointing) out-weight its performance overhead in many application scenarios, leading to real-world adoption in those application domains (e.g., security testbeds, virtual organizations).

This paper makes an important, timely contribution to the performance optimization of overlay-based virtual networking. It shows that, by moving the core of the network virtualization functions into the VMM, it is possible for such a layer-2 virtual network to approach native networking performance (namely, without any virtualization or overlaying), paving the way for the adoption of such virtual networks for high performance parallel and distributed applications. Furthermore, VNET/P-based virtual networks are expected to become "global citizens" that dynamically reside in (and seamlessly migrate between) cloud, HPC, and even desktop-grid platforms running a wide variety of parallel/distributed applications.

The VNET/P system is elegant, efficient, and practical. It is readily deployable in a wide range of computing platforms (from PCs to supercomputers) and agnostic to underlying physical networks (from Ethernet to InfiniBand). Moreover, its design can be easily instantiated in many other hypervisors such as Xen and KVM).

The paper provides useful insights into the virtualization of layer-2 networks as well as first-hand experience in optimizing their performance. Such insights and experience are of value to researchers working on network performance optimization for cloud, grid, and HPC environments.

Does this paper present convincing results?

> *Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?*

This paper presents solid evaluation results using micro-benchmarks and parallel computing benchmarks. The results are very encouraging as they indicate close-to-native performance at both network transport and application levels. These results will help mitigate the concerns of HPC users when deploying their applications in VNET/P-based virtual networked environments. The results also indicate that VNET/P is suitable for deployment in datacenters with high throughput and low-latency datacenter networks.

The evaluation is done quite thoroughly and follows sound methodology. Moreover, the authors are honest in also reporting the less-than-optimal results (e.g., those from the InfiniBand experiments) and suggest possible approaches to improving them.

Is this paper sufficiently well executed?

> *Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?*

This is a well-written paper with good structure and easy-to-follow content. Overall it is well executed with helpful motivation and background behind the proposed system, well-justified design, and convincing evaluation results. It is also commendable that the VNET/P system is fully implemented and made publicly available for adoption and independent evaluation.

This is one of the first efforts towards closing the performance gap for overlay-based virtual networks. Follow-up efforts by other researchers are likely. Porting of VNET/P design to other hypervisors may be of immediate interest to the open-source community and industry.

To further improve the paper, the authors are encouraged to do the following during revision:

(1) Consider a better way to motivate and introduce the VNET/P system. For example, in the introduction, they could use the performance deficiencies and inflexibilty of VNET/U as a starting point to motivate the development of VNET/P, followed by the introduction of kernel space portable routing and throughput-enhancing push and pull-based latency reduction mechanisms as key new features in VNET/P's design

(2) Present a stronger scientific comparison with the state-of-the-art (including the authors' own prior work) and clearly highlight the new technical contribution of VNET/P in comparison with relevant research and commercial efforts.

Reviewer 5 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental concepts, or increased understanding, or practical relevance?

The paper addresses the issue of how to achieve high performance network communication between VM's - within a cluster or supercomputer. The authors present the design of VNET/P, a new virtual machine networking interface implementation that includes elements of scheduling, convoying/bursting to reduce delivery latency and throughput in distinct communication scenarios.

Does this paper present convincing results?

Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?

The paper presents the design of VNET/P, and provides a set of measurements which compare the performance of the VNET/P implementation to a native NIC, for 1G and 10G networks with several different protocols, MTU's and with some simple communication patterns. These results show document the gap in performance (in most cases not large, but not zero) between the native performance and the VNET/P performance and the authors claim the different is "negligible". What criteria is used for this?

The main weakness of the paper is that while it describes the design of VNET/P adequately and the performance comparison vs. native is reasonably comprehensive at the simple communication level, there are several significant shortcomings. Perhaps the most glaring example of this is the lack of even a related work section. More specifically...

1. the description of the design of VNET/P is confusing and unclear in several places, and is not articulated to explain the difference between the VNET/P approach and others. What is the new intellectual contribution in organization?

2. the performance evaluation makes no comparison to other network virtualization approaches, much less a factored comparison which would show how the design choices in VNET/P contribute to the performance, and that that performance is better than others. There isn't even a comparison to the author's previous work on VNET/U, which is the code base for this work! It would greatly improve the ability to extract portable learnings to include such studies.

3. there's no performance evaluation to compare to network performance on other virtualization approaches -- kernel based or even hardware supported -- which might have advantages in performance, but nonetheless are useful points of comparison bot h to evaluate the contribution, but also to show other opportunitiies or weaknesses of this work.

Is this paper sufficiently well executed?

Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?

The paper focuses on what the authors did and describes that well. However, it needs to be improved by making the case for the scientific contribution -- and framing measurements and results and discussion in a fashion that make it understandable and portable.

Reviewer 6 Comments top

Does this paper address an important issue?

Does this paper address an important issue? If the issue or problem addressed were (or now is) completely understood or solved, how important would that be, in terms of either fundamental

concepts, or increased understanding, or practical relevance?

This is an implementation paper that shows the importance of reducing virtual networking overheads in clouds. Though the paper makes the claim that this represents a necessary step for the migration of HPC workloads into amazon-like clouds, the paper ignores a wide body of research that shows that there is a long way for either HPC or MapReduce style of programming to move to clouds in a serious way. Similarly the techniques suggested have been well covered in VEE and related conferences, so in summary the paper covers little new ground.

Does this paper present convincing results?

Does this paper present convincing results? Do the results provide worthwhile insight into the topic addressed? Are the results likely to be widely used by others? Does the work open up new areas, present new ideas, and/or serve as a foundation for new work?

The results of the paper are reasonable and plausible given the details of the implementation, though I would have liked if the authors had focused on depth rather than breadth. I would have liked to understand the contribution of the proposed mechanisms to latency and bandwidth in the 1 and 10 Gig cases. In particular, since today's CPUs handle 1 gig easily, a detailed introspection of 10 Gig might have benefited the paper. Overall, I would characterize the work as a confirmation of various VMM related activity in networking and storage.

Is this paper sufficiently well executed?

Is this paper sufficiently well executed? Are there flaws (e.g., technical mistakes, important uncited related work, poor assumptions, insufficient scope of evaluation, unsubstantiated conclusions, poor writing) in the paper? Are the flaws fundamental or superficial? That is, are the results likely to be true despite the flaws, or do the flaws fundamentally impact the results in the paper?

The paper suffers from the fact that we have to go to the second half of Section 4.2 to find out the real contribution of the paper. Before that, we are left guessing as to how the paper achieves near-native bandwidth and latency - there is a lot of teasing in the introduction, but no clear answer.

In the performance section, there are a lot of absolute claims - i.e. we achieve 516 MBps b/w - why is this good enough as claimed? As has been said before, a few detailed analyses would have helped the paper a lot.

There is also comparison to VNET/U - why is this being chosen as a control (not explained) and what is the significance of being better than this. Are there no other systems to compare with? I find this hard to believe.

Committee comments to authors: top

None