COPA Use Case:
Distributed Secure Joint Computation

Rushi Patel*, Pouya Haghi*, Shweta Jain‡, Andriy Kot‡, Venkata Krishnan‡,
Mayank Varia† and Martin Herbordt*
*College of Engineering, Boston University, Boston, MA
†Computing & Data Sciences, Boston University, Boston, MA
‡Intel Corporation, Hudson, MA
Email: *{ruship,haghi,herbordt}@bu.edu †varia@bu.edu ‡{shweta.jain,andriy.kot,venkata.krishnan}@intel.com

Data centers provide good environments for distributed computing as they are easily accessible and may have low-latency communication between nodes [1]; often, however, performance is limited by network bandwidth. These network bottlenecks drive the need for alternative communication resources to improve performance of large-scale applications. SmartNICs [2]–[4] have been introduced to perform the same tasks of standard NICs, but contain additional resources to allow for network function optimization with additional hardware. Adoption of SmartNICs continues to increase as a means to accelerate network functions and offload packet processing tasks away from CPU resources [5]–[13].

Intel's Configurable Network Protocol Accelerator (COPA) [14], [15] was developed as a SmartNIC with configurable FPGA resources. COPA supports use of the open-source software library, OpenFabric interface (OFI) libfabric [16], for platform-agnostic development and as a standard for networking and acceleration invocation. The COPA framework provides two options to reconfigurable accelerators, inline and lookaside, both of which are directly accessible from the libfabric API. COPA uses a unique architecture to enable high speed remote direct memory access (RDMA) between nodes at 100Gb/S line rate. So far, however, there has been no published work demonstrating or evaluating COPA with respect to a distributed application; that is our goal here.

As a candidate application we have selected Multi-Party Computation (MPC), which would greatly benefit from the features available through the COPA framework. MPC is the cryptographic process of performing calculations on confidential data between multiple data holders while maintaining a level of confidentiality, integrity, and assurance of one's own private data. This form of joint computation is especially important for industries such as healthcare and finance, as user data is typically under protection through laws and regulations. FPGA accelerated Multi-Party Computation continues to be a progressive research topic [17]–[29] as significant performance improvements can be obtained from hardware acceleration.

We argue that combining the COPA tool-set with state-of-the-art MPC algorithms can reduce the communication bottleneck for a high performance computation inside a datacenter environment. We show that utilizing the COPA system enables a method of performing low-level MPC operations with minimal CPU interaction, while enabling improved performance compared to traditional CPU and NIC implementations.

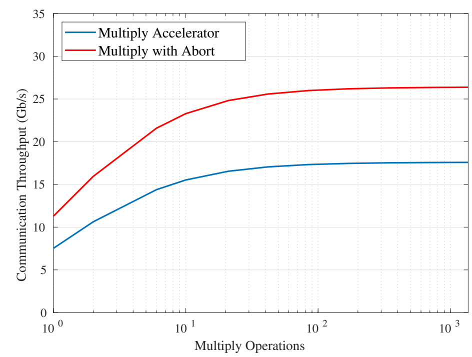

Fig. 1. Throughput comparison between base MPC accelerator and MPC with malicious security running at 275 MHz with varying batch sizes of multiply operations.

In our implementation, each party maintains ownership of a single FPGA connected to a host system using the COPA framework for communication between party members. Acceleration is performed through the use of unique commands sent from each host system directly to the FPGA lookaside accelerator through a dedicated queue. The command format allows for batch operations on a stream of data from a specified source and saves local results back to host memory while preparing the network data for transfer to each party member.

Using a single accelerator and batching multiplication operations over a stream of source data, the accelerator results are similar to past implementations saturating a traditional 10Gb/s link [28], [29]. Examining the throughput of large batches of multiplication operations, Figure 1 shows that a single accelerator needs a 17.5Gb/s connection, while the inclusion of additional hashed data values for malicious security requires larger than a 26.3Gb/s connection to avoid saturation. For COPA, these results show that we can run up to 6 parallel MPC accelerators before saturating the network. We compare our hardware implementation results against a traditional datacenter CPU and 10Gb/s network and show the potential for a 2x-10x improvement in MPC operations performed with minimal additional FPGA resources.

REFERENCES

[1] A. Caulfield, E. Chung, A. Putnam, H. Angepat, J. Fowers, M. Haselman, S. Heil, M. Humphrey, P. Kaur, J.-Y. Kim, D. Lo, T. Massengill, K. Ovtcharov, M. Papamichael, L. Woods, S. Lanka, D. Chiou, and D. Burger, "A cloud-scale acceleration architecture," in *49th IEEE/ACM Int. Symp. Microarchitecture*, 2016, pp. 1–13.

[2] N. Zilberman, Y. Audzevich, G. A. Covington, and A. W. Moore, "Netfpga sume: Toward 100 gbps as research commodity," *IEEE micro*, vol. 34, no. 5, pp. 32–41, 2014.

[3] NVIDIA. Bluefield™ smartnic ethernet. [Online]. Available: https://www.mellanox.com/products/BlueField-SmartNIC-Ethernet

[4] ——. NVIDIA Mellanox Innova-2 Flex Open Programmable SmartNIC. [Online]. Available: https://www.nvidia.com/en-us/networking/ethernet/innova-2-flex/

[5] Y. Le, H. Chang, S. Mukherjee, L. Wang, A. Akella, M. M. Swift, and T. V. Lakshman, "Uno: Uniflying host and smart nic offload for flexible packet processing," in *Proceedings of the 2017 Symposium on Cloud Computing*, ser. SoCC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 506–519. [Online]. Available: https://doi.org/10.1145/3127479.3132252

[6] W. Schonbein, R. E. Grant, M. G. F. Dosanjh, and D. Arnold, "Inca: In-network compute assistance," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3295500.3356153

[7] H. Eran, L. Zeno, M. Tork, G. Malka, and M. Silberstein, "NICA: An infrastructure for inline acceleration of network applications," in *2019 USENIX Annual Technical Conference (USENIX ATC 19)*. Renton, WA: USENIX Association, Jul. 2019, pp. 345–362. [Online]. Available: https://www.usenix.org/conference/atc19/presentation/eran

[8] M. Tork, L. Maudlej, and M. Silberstein, *Lynx: A SmartNIC-Driven Accelerator-Centric Architecture for Network Servers*. New York, NY, USA: Association for Computing Machinery, 2020, p. 117–131. [Online]. Available: https://doi.org/10.1145/3373376.3378528

[9] S. Grant, A. Yelam, M. Bland, and A. C. Snoeren, "Smartnic performance isolation with fairnic: Programmable networking for the cloud," in *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*, ser. SIGCOMM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 681–693. [Online]. Available: https://doi.org/10.1145/3387514.3405895

[10] S. Miano, R. Doriguzzi-Corin, F. Risso, D. Siracusa, and R. Sommese, "Introducing smartnics in server-based data plane processing: The ddos mitigation use case," *IEEE Access*, vol. 7, pp. 107 161–107 170, 2019.

[11] H. Shahzad, A. Sanaullah, and M. Herbordt, "Survey and Future Trends for FPGA Cloud Architectures," in *IEEE High Performance Extreme Computing Conference*, 2021, dOI: 10.1109/HPEC49654.2021.9622807.

[12] M. Zink, D. Irwin, E. Cecchet, H. Saplakoglu, O. Krieger, M. Herbordt, M. Daitzman, P. Desnoyers, M. Leeser, and S. Handagala, "The Open Cloud Testbed (OCT): A Platform for Research into new Cloud Technologies," in *IEEE International Conference on Cloud Networking (IEEE CloudNet)*, 2021, doi: 10.1109/CloudNet53349.2021.9657109.

[13] C. Bobda, J. Mandebi, P. Chow, M. Ewais, N. Tarafdar, J. Vega, K. Eguro, D. Koch, S. Handagala, M. Leeser, M. Herbordt, H. Shahzad, P. Hofstee, B. Ringlein, J. Szefer, A. Sanaullah, and R. Tessier, "The Future of FPGA Acceleration in Datacenters and the Cloud," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 15, no. 3, pp. 1–42, 2022, doi: 10.1145/3506713.

[14] V. Krishnan, O. Serres, and M. Blocksome, "COnfigurable Network Protocol Accelerator (COPA)," in *2020 IEEE Symposium on High-Performance Interconnects (HOTI)*, 2020.

[15] ——, "Configurable network protocol accelerator (copa)," *IEEE Micro*, vol. 41, no. 1, pp. 8–14, 2020.

[16] P. Grun, S. Hefty, S. Sur, D. Goodell, R. D. Russell, H. Pritchard, and J. M. Squyres, "A brief introduction to the openfabrics interfaces - a new network api for maximizing high performance application efficiency," in *2015 IEEE 23rd Annual Symposium on High-Performance Interconnects*, 2015, pp. 34–39.

[17] K. Järvinen, V. Kolesnikov, A. R. Sadeghi, and T. Schneider, "Embedded SFE: Offloading server and network using hardware tokens," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6052 LNCS, pp. 207–221, 2010.

[18] ——, "Garbled circuits for leakage-resilience: Hardware implementation and evaluation of one-time programs," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6225 LNCS, pp. 383–397, 2010.

[19] T. K. Frederiksen, T. P. Jakobsen, and J. B. Nielsen, "Faster maliciously secure two-party computation using the GPU," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8642, no. grant 61061130540, pp. 358–379, 2014.

[20] E. M. Songhori, S. Zeitouni, G. Dessouky, T. Schneider, A. R. Sadeghi, and F. Koushanfar, "GarbledCPU: A MIPS processor for secure computation in hardware," *Proceedings - Design Automation Conference*, vol. 05-09-June, 2016.

[21] S. U. Hussain, B. D. Rouhani, M. Ghasemzadeh, and F. Koushanfar, "MAXelerator: FPGA Accelerator for Privacy Preserving Multiply-Accumulate (MAC) on Cloud Servers," *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, pp. 1–6, 2018.

[22] E. M. Songhori, M. S. Riazi, S. U. Hussain, A. R. Sadeghi, and F. Koushanfar, "ARM2GC: Succinct garbled processor for secure computation," *Proceedings - Design Automation Conference*, 2019.

[23] S. U. Hussain and F. Koushanfar, "FASE: FPGA acceleration of secure function evaluation," *Proceedings - 27th IEEE International Symposium on Field-Programmable Custom Computing Machines, FCCM 2019*, pp. 280–288, 2019.

[24] X. Fang, S. Ioannidis, and M. Leeser, "Secure function evaluation using an FPGA overlay architecture," *FPGA 2017 - Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, pp. 257–266, 2017.

[25] ——, "SIFO: Secure computational infrastructure using FPGA overlays," *International Journal of Reconfigurable Computing*, vol. 2019, 2019.

[26] K. Huang, M. Gungor, X. Fang, S. Ioannidis, and M. Leeser, "Garbled circuits in the cloud using FPGA enabled nodes," *2019 IEEE High Performance Extreme Computing Conference, HPEC 2019*, pp. 1–6, 2019.

[27] M. Leeser, M. Gungor, K. Huang, and S. Ioannidis, "Accelerating large garbled circuits on an FPGA-enabled cloud," *Proceedings of H2RC 2019: 5th International Workshop on Heterogeneous High-Performance Reconfigurable Computing - Held in conjunction with SC 2019: The International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 19–25, 2019.

[28] P.-F. Wolfe, R. Patel, R. Munafo, M. Varia, and M. Herbordt, "Secret Sharing MPC on FPGAs in the Datacenter," in *IEEE Conference on Field Programmable Logic and Applications*, 2020.

[29] R. Patel, P.-F. Wolfe, R. Munafo, M. Varia, and M. Herbordt, "Arithmetic and Boolean Secret Sharing MPC on FPGAs in the Data Center," in *IEEE High Performance Extreme Computing Conference*, 2020, doi: TBD.