Extended Abstract: Trustworthy System Security through 3-D Integrated Hardware

Ted Huffmire
Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
tdhuffmi@nps.edu

Jonathan Valamehr
Dept. of Electrical and Computer Engineering
UC Santa Barbara
Santa Barbara, CA 93106
valamehr@ece.ucsb.edu

Timothy Sherwood
Dept. of Computer Science
UC Santa Barbara
Santa Barbara, CA 93106
sherwood@cs.ucsb.edu

Ryan Kastner
Dept. of Computer Science and Engineering
UC San Diego
La Jolla, CA 92093
kastner@ucsd.edu

Timothy Levin
Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
televin@nps.edu

Thuy D. Nguyen
Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
tdnguyen@nps.edu

Cynthia Irvine
Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
irvine@nps.edu

While hardware resources in the form of both transistors and full microprocessor cores are now abundant, economic factors prevent specialized hardware mechanisms required for secure processing from being integrated into commodity parts. We are exploring a novel way in which commodity hardware can be augmented *after fabrication* to enhance secure operation for only those systems that require it. Our methods will be applicable to a wide range of security problems, including the detection and isolation of hardware subversion and Trojan horses, cache-based side channels in chip multi-processors (CMPs), embedded systems security, and hardware intrusion detection and prevention.

Utilizing off-the-shelf components to build trustworthy systems results in a constant battle with the underlying machine to provide separation, isolation, and protection. This problem is exacerbated by the movement to multi-core processors since security functionality (e.g., strong security primitives) is rarely considered a priority at the platform ISA or micro-architecture levels and since features exploitable by adversaries (e.g., resource sharing) are included for performance at the expense of security. Without a significant shift in the way computing systems are constructed (from the software down to the circuits), unacceptable amounts of time and resources will be spent attempting to contain the vulnerabilities introduced by each new processor performance feature. ***To address these problems, we are pursuing a radical transformation in the way trustworthy systems are developed and deployed, one that allows direct hardware support for fine grain control of the underlying hardware system, yet that can still leverage the performance and cost benefits provided by the latest commodity parts through the augmentation of those parts with a 3-D Integration approach.***

Hardware manufacturers are reluctant to make hardware support for trustworthy systems a priority. Incorporating strong security enhancements requires significant resources, and integrating these mechanisms into a complex design presents many practical and theoretical problems, driving up the costs and prolonging the release schedule – all of which is unacceptable in the extremely cost sensitive desktop market. Trustworthy computing systems are caught between the competing pressures to provide complete and precise security policy enforcement and the need to leverage the performance and resources associated with the latest commodity products and parts.

We intend to disentangle the security mechanisms from the design, consolidating them onto a security overlay, *a separate layer* of circuitry, called a *control plane*, that is stacked on top of a commodity integrated circuit. The security mechanisms that reside in the control plane can then be connected to the underlying chip, called the *computation plane*, with any number of die-stacking technologies, yet can be left unattached to enable the

manufacturer to continue to sell the un-enhanced product at a lower cost. Attaching multiple planes together in 3-D stacks is a new yet already marketed technology [1], which is being explored by many major microprocessor manufacturers [2], [3], [4]. 3-D devices contain multiple active layers[1], which are interconnected using techniques such as through-silicon vias (TSVs), also referred to as *posts*. A similar 3-D Integration approach has already been used to implement profiling functions in the control plane [5], [6].

This work has the potential to cut across all levels of the system stack, from the application software, through the operating system, the computer architecture, down to the level of circuits and packaging. Various problems need to be solved at each layer for this approach to be successful. For example, the control plane's ability to enforce policies depends on circuit-level capabilities for monitoring and restricting activity if needed. The challenge is that these restrictions can disable some functionality in the computation plane, but the computation plane must be fully functional in its absence. Another challenge is how to integrate different technology nodes (e.g., a 130nm and 45nm fabrication process for the control and computation planes, respectively).

The ability to enhance commodity hardware with application-specific security functionality is applicable to a wide range of hardware security issues:

- CMPs suffer from cache-based side channels because cores can observe the cache evictions of other cores. To address this problem, we will intercept and manage traffic destined for the cache-bus, forcing signals on the computation plane to take a detour to the control plane, where a cache manager will mediate access to the cache.
- To isolate untrusted software running on the multiple cores in the computation plane, we will use posts from the control plane to disable specific wires in the computation plane, allowing us to cut those connections that violate isolation requirements. To provide controlled sharing of resources, we will use the control plane to mediate the interaction of cores and other system resources in the computation plane. Using disabling posts paired with bypass posts, we will force signals on the computation plane to be rerouted to the control plane, where the legality of the requested action can be checked against a security policy. The control plane can also be used for isolated execution of sensitive code such as proprietary programs or crypto processing, as well as for storage of high integrity code and data.
- The control plane can be used to monitor and log the interactions of cores in the computation plane. A 3-D approach provides protection of both the configurable hardware audit mechanism and protected storage of the audit records. Control plane functions can be used to detect hardware intrusions, subversions, Trojans, and information leakage, including covert channels, side channels, and direct channels in the computation plane.
- The control plane can contain an object reuse mechanism that clears the sensitive state of a core after one task finishes but before the next task begins. Logic in the control plane can sanitize both the intra- and inter-core shared resources. The control plane can also be used to configure and initialize computation plane cores.
- The control plane can be used to implement a variety of high-integrity data tagging schemes, such as security classification tags that mark outgoing datagrams with the sensitivity label of the sending core, provenance tags that indicate the source of inter-core data transfers, and persistent memory tags that indicate which core most recently modified a given memory segment.

The current trends of building trustworthy systems atop increasingly complex and less well understood hardware make such systems increasingly costly to deploy and maintain. This research introduces a fundamentally new method by which security mechanisms can be incorporated into hardware and thus has the potential to significantly shift the economics of trustworthy systems. With our approach, trustworthy systems can be built quickly using unmodified commodity processors equipped with a customized control plane that provides configurable, application-specific security functionality.

[1]The active layer is the silicon layer where transistors reside, and metal layers are fabricated above that to connect the transistors together.

References

[1] Ziptronix, "3D integration for mixed signal applications," Morisville, NC, 2002. [Online]. Available: http://www.ziptronix.com/images/pdf/analog_applications.pdf

[2] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir, "Design and Management of 3D Chip Multiprocessors Using Network-in-Memory," *Proceedings of the 33rd annual International Symposium on Computer Architecture (ISCA)*, pp. 130–141, July 2006.

[3] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, D. McCauley, P. Morrow, D. W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. Shen, and C. Webb, "Die Stacking (3D) Microarchitecture," *Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 469–479, December 2006.

[4] K. Puttaswamy and G. H. Loh, "Thermal analysis of a 3D die-stacked high-performance microprocessor," *Proceedings of the 16th ACM Great Lakes symposium on VLSI*, pp. 19–24, May 2006.

[5] S. Mysore, B. Agrawal, S. Lin, N. Srivastava, K. Banerjee, and T. Sherwood, "Introspective 3-D chips," in *Proceedings of the 12th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, San Jose, CA, October 2006.

[6] ——, "3-D integration for introspection," *IEEE Micro*, vol. 27, no. 1, January 2007.