

Revolutionary Direction of Secured Supercomputer Technologies Development in Russia

Andrey S. Molyakov ^a

Institute of IT and Cybersecurity, Russian State University for the Humanities, Moscow, Russia

Keywords: J7, J10, Project "Angara", BNW-network.

Abstract: The article describes a promising project for the development of secured strategic supercomputer "Angara", which is a set of nodes of different types, united by several communication networks. The service nodes are built on conventional superscalar microprocessors. Computing nodes are built on special multicore multithread-stream microprocessors (microprocessors of the J series), are combined into modules in the form of multi-socket boards, and can work on a logically single addressable memory (globally addressable memory) formed by local memories of modules with computational nodes. There are two models of J series microprocessors: the younger one is J7, the older one is J10. Further, the characteristics and principles of operation of the younger model J7 are considered, and for the older model, only a list of fundamental differences and general characteristics is given.

1 INTRODUCTION

SCSN "Angara" is a set of nodes of different types, united by several communication networks, one of which has a unique property of transmitting large streams of short packets with high throughput. This network is necessary to implement work with globally addressable memory, hereinafter we will call it the basic working network (BNW-network). The nodes are connected to the basic working network and can be computing and service (Molyakov, 2020).

Computing nodes are built on special multicore multithread-stream microprocessors (microprocessors of the J series), are combined into modules in the form of multi-socket boards, and can work on a logically single addressable memory (globally addressable memory) formed by local memories of modules with computational nodes. There are two models of J-series microprocessors: the younger one is J7, the older one is J10. Further, the characteristics and principles of operation of the younger model J7 are considered, for the older model a list of fundamental differences and general characteristics is given.


Service nodes are built on conventional superscalar microprocessors, perform the functions of

input-output, user connection, interface with the global network, and can also perform computations if they are well localized and efficiently executed on these nodes. Computing and service nodes are connected to another network (RAS network), which is a component of the reliability, availability, and service subsystem. A detailed description of the system-wide specification can be found in publication (Mitrofanov, Slutskin and Eisymont, 2008).

2 METHODOLOGY

New architectural and technological approaches in the field of creating "new generation" supercomputers.

The J7 microprocessor is referred to as a "multicore, multithread-stream microprocessor with support for globally addressable memory operations". The term "globally addressable memory" in the name of the microprocessor should be understood as follows: in the J7 / J10 microprocessors, the virtual memory is organized so that when it is accessed, it is automatically recognized during the translation of the address which node of the system should be accessed and this call is made without user intervention. The J7

^a  <https://orcid.org/0000-0003-4560-5911>

microprocessor has two multi-threaded cores (MTcore0 and MTcore1). Four command pipelines are available in one multi-threaded core. Each of them works with 16 threaded devices, each of which can run one process (Semenov, 2010).

Several tasks can be executed simultaneously in one microprocessor core. Each task is assigned one protection domain, one of the kernel tasks is the operating system. The user's task performed in the microprocessor can be simultaneously executed in the protection domains of its different cores; information about the task's binding to the protection domains is stored in a special table of the microprocessor.

The main idea behind hiding delays, i.e. ensuring its insensitivity to these delays in terms of the developed real performance, - ensuring a high rate of execution of operations with memory and network. This requires a special organization of the processor and the applications running on it, a special organization of a communication network, a special organization of memory. All of these devices require the ability to perform a large number of operations simultaneously and high pipelining. The computational model of the application is required to be able to issue a large number of operations, which is why multithreading is needed.

Memory and network latencies can be up to ten times higher, the pace is not always able to withstand the same operation per clock cycle, especially in the communication network - the physical limitations of the bandwidth of data transmission links between nodes within the network effect. For these reasons, a larger number of threaded devices are selected, up to 64-128 per MT core, which makes it possible to have up to 512-1024 concurrent memory or network accesses in one core. Commands of memory accesses are used behind short vectors, for example, up to 8 64-bit words, which allows to further increase the number of concurrent memory accesses and reduce the overhead of organizing one memory access.

The term "streaming" is more applicable to the architecture of the J10 microprocessors and can be used in the sense of providing the ability to process data streams using data flow graph models. Two graph flow models are supported - static and dynamic graphs. These models are used to provide more parallelism and asynchrony, and static graphs are used to reduce the number of memory accesses when transferring data between nodes. A static graph node appears along with the entire graph, functions for some time, and then is deleted along with the graph. A node of a dynamic graph can appear and be destroyed during the operation of the graph. This possibility is preserved and strengthened in the

Chinese version (Molyakov, 2019). For a node of a dynamic graph, such a sequence of data arrival in an arc can be violated, so the data comes with special tags, by the coincidence of which they can find a pair for themselves in the stream of another arc. The operation may be normal if there is a pair for which the operation can be performed. Such selection of data corresponding to each other in the streams of arcs requires the use of memory with associative access, in this case, the associative address is a data tag. Such memory is implemented in software. RPC commands are actively used in the implementation of dynamic graphs.

3 RESEARCH RESULTS

New principles of access security model based on outbound command assembly and multi-domain protection.

Given the development of vulnerability search methods, the implementation of reactive information protection methods should be considered, along with preventive ones. Instead of the classical concept of "localized task" for supercomputers (SC), one should speak of parallel distributed stream structures generated and processed at different levels of the command pipeline hierarchy by processor devices connected by high-speed networks.

Object access attributes and subject privileges, connections between them are formalized as a set (conjunction) of predicates. You can track interaction and control access by a set of characteristic features (markers), represented as tuples of boolean variables.

Information security is based on controlling access to objects of management and guest operating systems; these objects can be attributed to different levels of protection. The traditional approach to access control assumes the use of access attributes (rights) in requests to these objects to perform some operations on them. If the verification of such attributes is successful, then access to the object at its security level is allowed, then the requested operation is performed on it. With this approach, it is technically possible to intercept the request and use its access rights in a surrogate request aimed at malicious impact.

We propose a new mathematical model for security. It is based on an 8-tier model (one level of protection was added) and new logic for controlling access to requests to perform actions with OS objects, which is implemented by a hypervisor with an appropriate organization, which additionally uses the

architectural capabilities of the proposed supercomputer model extended by the author.

In the course of experimental studies, new scientific results have been obtained that confirm the effectiveness and minimal loss of performance of the use of hardware virtualization technology in the form of a multi-level "sandbox" for promising SCs in comparison with the use of traditional superclusters. Since it is possible to control the execution of requests at the level of the components of the hypervisor and the transactional memory controller and it is impossible to control the operation of all equipment, during the research, the maximum level of functioning of the ISS agents was found - S8. With this configuration, when the number of levels of the hierarchy is $N = 8$, the execution of context-dependent operations becomes quasi-deterministic with a confidence level of approximately 0.9.

In classical processors with the von Neumann architecture, data and program codes are shared, which prevents the effective restriction of one object's access to the address space and data of another. They also do not implement multilayer protection mechanisms against attacks when executing system calls in a multi-level context of nested guest and control operating systems. Tagged architecture on the example of MCST Elbrus processor does not support hardware virtualization technology. However, the lack of support for hardware virtualization makes such architectural solutions highly specialized and does not support the emulation of various hardware and support for widely used hypervisors. Some features of the hardware virtualization technology also speed up the operation of virtual machines and increase the level of security. Hardware-assisted virtualization and multi-layered protection can reduce the overhead of creating an isolated runtime environment. In any operating system, when the kernel code is paired with hardware at the physical level, forbidden states appear - a zero tuple of data, which the processor prohibits accessing even programs in the zero protection ring. Only the processor module can perform active task context switching in protected mode since when shadow copying the data of the executable code, the programmer cannot get direct access to the information. This requires the implementation of a tiered query processing hierarchy.

This approach did not apply to previous generation microprocessors due to poor performance. The introduction of additional levels of privileges and levels of protection greatly slowed down the system as a whole (Molyakov, 2019). The high performance of supercomputers, on the contrary, makes it possible

to quickly analyze descriptor tables and calculate hash values of processes. They are distinguished by self-diagnostics, multi-level protection, binding each thread device to a specific domain, programming using non-functional, non-procedural languages - chains of calculations in the form of selectors, substitutions on the right, left of functional calculations, multi-level parallelization of algorithms, etc.).

When the boot program is executed, the threading device operates in a special privilege mode - `IPL_LEVEL`. Physical addressing when accessing instruction memory and data memory is used in this mode, then the virtual infrastructure manager and host OSs with hypervisor support are loaded. In this case, the `KERNEL_LEVEL` level is used for the kernel modules and the `SUPERVISOR_LEVEL` level for the equipment manager. At the final stage, the guest OS is launched at the `USER_LEVEL` level.

The only "stumbling block" in multicore multiprocessor systems is the problem of efficient implementation of the on-chip network and working with memory. The multicore microprocessors used in computational nodes have become a necessary measure to ensure the growth of their peak performance, and it is caused by the termination of the direct influence of Moore's law on the growth of processor cores performance. Multicore has given rise to the problem of efficient implementation of the on-chip network and exacerbated the problems of working with memory. Possible solutions include the use of multi-threaded and streaming architectures in processor cores.

The multithreading organization allows multiple threads of instructions to be executed concurrently, which makes it possible to increase the multitude of executable instructions and increase the flow of concurrent memory operations. The streaming architecture assumes the use of the decision fields of elementary processors in the form of static graphs of data streams. This makes it possible to reduce the total number of memory accesses since, in the decisive field, data is transferred from one fast resource to another without accessing memory (Molyakov, 2019).

The author proposes a method for reconfiguring the runtime environment, taking into account the requirements of mobility and ensuring the specific performance characteristics of the program for the safe expansion of the functionality of the system or application.

Only hardware support for sharded stacks can reduce compiler complexity and runtime overhead. In any operating system, when the kernel code is paired

with hardware at the physical level, forbidden states appear - a zero tuple of data, which the processor prohibits accessing even programs in the zero protection ring.

Only the processor module can perform active task context switching in protected mode since when shadow copying the data of the executable code, the programmer cannot get direct access to the information. These collisions are resolved by a new approach - marker scanning, which uses generative tables (Molyakov, 2016).

In addition to coding the address separation of memory protection rings, the OS of different classes implements a strongly typed interface for interfacing with the processor's hardware core and managing the context of binary code execution, taking into account the compilation and assembly profile - the use of system object markers.

The command processing pipeline is as follows. After fetching and issuing by any threading device a command to access the memory, the command enters the LSU functional block for executing memory access commands. The executive memory address prepared in the LSU is then transmitted to the MMU, in which the virtual address is translated into a physical address or a global virtual address.

If it is necessary to access the memory of a remote node, which is automatically determined in the MMU, a hardware-generated system short message packet with a memory access command is transmitted through the MSU network message control unit, the on-chip network, the network interface with the communication network.

When moving through the network, some packets can deviate from the path prescribed by the routing table, thus bypassing all kinds of "traffic jams" that they can independently detect. After such a rejection, the packet, under certain conditions, can return to a guaranteed deadlock-free movement over the network.

Globally addressable memory provides not only additional programming convenience, which, as experts expect, should affect the productivity of parallel programming by about 10 times. An increase in the efficiency of parallel programs is also expected, since two-way interaction models of the "send-receive" type, as a rule, by long messages, are replaced by one-way interactions using short messages. The reality of achieving greater efficiency with such a transition to a new memory model and organization of computations has already been proven in many experiments (Gorbunov, and Eismont, 2010).

The J7 / J10 microprocessors also use a special apparatus of tag bits and bits for controlling the execution of memory accesses, which are available both directly in memory cells and pointers to them. The programs can use physical and virtual addressing. Controlling the issuance of physical or virtual addresses when accessing memory is performed by setting a special bit in the thread state word. Physical addressing is allowed only in privileged kernel and boot modes. The memory of commands and data is addressed according to different schemes. Only data addressing is discussed below.

4 PRACTICAL RESULTS OF RESEARCH ON MODELING STANDS. PROTOTYPES OF CALCULATORS

General characteristics of the J7 microprocessor:

- There are two multi-threaded cores with 64 threaded devices in each and the ability to work in each core with four program and data protection domains;
- each thread device can execute software threads of the same type, and each such thread has access to only 32 64-bit general-purpose registers, 32 64-bit registers for storing floating-point numbers, 8 one-bit feature registers, 8 registers for storing control transfer addresses, a set of system registers and tables available only at the system software or hardware level;
- general registers and registers for storing floating-point numbers are implemented as single-level register memory, registers are used only for temporary storage of data;
- a set of commands contains about 160 commands, processed data - 64-bit signed and unsigned integers, 64-bit and 32-bit floating-point numbers;
- The organization of virtual memory is segment-page, three-level, work with shared memory distributed over many computing nodes (address distribution is cyclic and block-cyclic) is provided, its volume can reach 8 PB ($8 * 10^{15}$) on a separate task, and in one computing node - up to 256 GB, memory cells are equipped with tag bits to synchronize calls to these cells;
- the number of concurrently executed memory access operations - 1024;

- The multi-threaded core contains four fixed-point addition, subtraction and multiplication units (FXU), two integer dividers (DIV), two floating-point units (FPU), and one memory (LSU), the level of parallelization of commands is 4 commands, but from different thread devices, all operations are performed on scalar data, only operations from commands selected from the executed threads fall on functional devices,
 - peak performance of one J-7 microprocessor operating at a frequency of 500 MHz - about 4 Gflops for 64-bit numbers;
 - the number of computational nodes in a supercomputer built on a J-7 microprocessor is up to 32 thousand nodes, i.e. the peak performance of such a supercomputer will be 256 teraflops;
 - high real performance at the level of 50-60% should be achieved due to the following basic architectural properties of the microprocessor: tolerance to delays in performing operations with memory and communication network due to its multithreading using threads of the same type;
 - the ability to localize not only data but also computations, the use of computation models over distributed shared memory.
- General characteristics of the J10 microprocessor:
- the presence of up to 8 multi-threaded cores, each of which contains 128 or 256 threaded devices and provides 16 protection domains;
 - each thread device can execute software threads that differ in the type of computations performed (light, medium and heavy computational models);
 - the number of general and floating-point registers used by one thread device can vary depending on the type of thread being executed from 16 to 64;
 - register memory of general registers and floating-point registers is implemented as two-level cacheable register files of large size, which provides greater efficiency in the use of crystal area and power consumption;
 - general and floating-point registers can be used not only for the temporary storage of data but also as input / output ports for ultra-fast inter-thread interactions;
 - a set of commands contains about 300 commands, processed data - 32-bit and 64-bit signed and unsigned integers, 32-bit and 64-bit floating-point numbers, bit matrices up to 64×64 ;
 - the organization of virtual memory is the same as in J-7;
 - the number of concurrently executed commands of memory access and inter-thread interactions through register input / output ports - up to 16384;
 - in multi-threaded cores there are blocks of functional devices FXU and FPU operating in asynchronous and synchronous mode. When working in synchronous mode, FXU and FPU can synchronously perform SIMD operations on short vectors: 8 simultaneous operations on 32-bit numbers, 4 operations on 64-bit numbers. Operations on the FXU and FPU can come from commands from executable threads, as well as directly from the function block for receiving / issuing system messages. The peak performance of a single J-10 microprocessor with eight cores and operating at 1 GHz is about 64 Gflops for 64-bit numbers. The multi-threaded core additionally introduces specialized devices for performing operations on bit matrices and a device for cryptographic algorithms and modular arithmetic;
 - the number of computational nodes in a supercomputer built on a J-10 microprocessor is up to 32 thousand. The peak performance is thus 2 Pflops for 64-bit numbers;
 - high performance up to 70-90% should be achieved due to the following basic architectural properties: processor tolerance to delays in performing operations with memory and communication network due to multithreading with threads of different types of Spatio-temporal localization of memory accesses; localization of data near computational processes and computations near data; the use of computational models of programs over distributed shared memory; a sharp reduction in the number of memory accesses and granularity of parallelization required for the execution of tasks due to the use of stream models of computations with static graphs supported by the hardware in the J-10; a sharp increase in thread parallelism due to the use of hardware and firmware supported in the J-10 stream models of computations with dynamic graphs.

5 CONCLUSION

The most advanced work in Russia is the work of the "public level" class associated with the development

of the MVS-express network, which uses the PCI-express interface and PLX communication chips. There are already two installations where this approach is used - K100 (Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences) and St. Petersburg State Polytechnic University. Improvement of hardware and software MVS-express continues, and techniques for effective parallel programming of applied tasks are also being developed. Similar works have been started at JSC "NITSEVT", but they are focused on using the HyperTransport interface. Also, NITSEVT is working on a "moderate level" class of GAS / PGAS implementation - an N-torus network router is being implemented and it is planned to be strengthened with multi-threaded cores (Eisymont, 2010). Work is underway to create a VLSI router, which is currently implemented as a prototype on an FPGA.

We should also mention the work of the "moderate level" class of the GAS / PGAS implementation, carried out by the staff of the Ailamazyan Institute of Applied Problems of the Russian Academy of Sciences and the RSK SKIF company in the SKIF-Aurora project. This work is in many respects similar to the implementation of FPGA-based routers carried out at JSC "NITSEVT", in which the main developers were former employees of this organization. Work of this type is also carried out by the T-Platforms Group via the Extoall network, and its version of the microprocessor is being developed, but there is no information about these developments (Gorbunov, Elizarov and Eisymont, 2011).

The results of the SCS "Angara" project, which has been carried out for more than 7 years, became the basis for Chinese work on the strategic supercomputer ST-2 in 2009 for the global information system of China's military intelligence. The TSMC factory has now produced prototypes of a 12-core mass multi-threaded microprocessor using 45 nm technology, which is a Chinese modified version of the Russian J7 microprocessor project for the Angara SCSN. The microprocessor was improved in some characteristics: computational capabilities were enhanced by introducing SIMD operations on short vectors and GPU elements such as synchronously executed threads in addition to asynchronous threads from J7.

Work in this direction in China is being carried out at NUDT, the National University of China's Defense Technologies. They have serious prospects for creating a supercomputer with an exascale performance level not only for building information systems, but also for solving scientific and technical

problems with a high level of real productivity, i.e. not with peak, but real performance in exaflops. Based on the existing domestic schemes for the implementation of the multi-threaded core J7 / J10, it is necessary to design the architecture and microarchitecture of this core, taking into account the Chinese experience of revision and American work, to carry out improvements and acceptance testing on fragments of special tasks of interest from different departments with the aim of subsequent introduction into industrial operation.

REFERENCES

- Molyakov, A.S. (2020). Based on Reconfiguring the Supercomputers Runtime Environment New Security Methods. *Advances in Science Technology and Engineering Systems Journal*, 5(3): 291-298.
- Molyakov, A.S. (2020). Main Scientific and Technological Problems in the Field of Architectural Solutions for Supercomputers. *Computer and Information Science*, 13(3). DOI: 10.5539/cis.v13n3p89
- Semenov, A.S. Development and research of the architecture of globally addressable memory of a multithread-streaming supercomputer. Dissertation for the degree of candidate of technical sciences. Specialty 05.13.15 - Computing machines, complexes and computer networks. Moscow 2010, 224 pages.
- Molyakov, A.S. (2019). China Net: Military and Special Supercomputer Centers. *Journal of Electrical and Electronic Engineering*, 7(4): 95-100. – DOI: 10.11648/j.jee.20190704.12
- Molyakov, A.S. (2019). Age of Great Chinese Dragon: Supercomputer Centers and High Performance Computing. *Journal of Electrical and Electronic Engineering*, 7(4): 87-94. – DOI: 10.11648/j.jee.20190704.11
- Molyakov, A.S. (2019). New Multilevel Architecture of Secured Supercomputers. *Current Trends in Computer Sciences & Applications*, 1(3). DOI:10.32474/CTCSA.2019.01.000112.
- Molyakov, A.S. (2019). Supercomputer and new generation operating systems. *Information security: yesterday, today, tomorrow. International scientific and practical conference: collection of articles of the Russian State Humanitarian University*. – Moscow, pages 196 – 200.
- Molyakov, A. S. (2016). A Prototype Computer with Non-von Neumann Architecture Based on Strategic Domestic J7 Microprocessor. *Automatic Control and Computer Sciences*, 50(8): 682 -686.
- Molyakov, A. S. (2016). Token Scanning as a New Scientific Approach in the Creation of Protected Systems: A New Generation OS MICROTEK. *Automatic Control and Computer Sciences*, 50(8): 687-692.
- Gorbunov, V. and Eisymont, L. (2010). Exascale Barrier: Problems and Solutions. *Open Systems*, 6: 12-15.

- Eisymont, L. (2010). DARPA UHPC - the road to exaflops. *Open Systems*, 12. <http://www.osp.ru/>
- Gorbunov, V., Elizarov, G. and Eisymont, L. (2011). HPC: regional news. *Open Systems*, 2: 12-16.
- Mitrofanov, V., Slutskin, A. and Eisymont, L. (2008). Supercomputer technology for strategic missions. *ELECTRONICS: NTB*, 7: 66-79.

