SOLUTION BRIEF

Trillium Multi-Core SIP: Why Throughput Matters

INTRODUCTION

WHY DOES SIP MATTER?

Session Initiation Protocol (SIP) is the foundation for most emerging communication services; it is transport layer independent and easily extensible. In fact, SIP-based applications can be deployed by all communications service providers from traditional Fixed Line and Mobile Operators to Internet Service Providers (ISPs).

SIP-based services range from required utilities to new revenue-generating applications such as Femtocell Gateways, Convergence Servers, IP Multimedia Subsystem (IMS) elements, softswitches, mobile-to-fixed gateways, Voice over IP (VoIP) elements, file sharing applications, voice or video conferencing, messaging, and gaming platforms.

WHAT DRIVES SIP DEMAND?

Demand for SIP-based services and for high SIP call processing throughput comes from many sources:

1. Dramatic increase in Smart Mobile Devices
2. Large increase in fixed and mobile bandwidth – especially 3G, WiMAX, and Femtocells
3. Extensibility of SIP itself as more services become SIP-based
4. Convergence – fixed and mobile as well as IP and circuit switched
5. Explosive subscriber growth in emerging economies
6. Increasing prevalence of mobile commerce

WHY DOES SIP THROUGHPUT AND SCALABILITY MATTER?

As these and other SIP-based services proliferate, SIP call processing throughput will determine cost and capacity. Architectural choices impact upfront and long-term costs, and the ideal solution is one that can start small – as in a limited, inexpensive field trial – but scale easily, quickly, and cost-effectively. When such trials succeed, operators will demand rapid increases in user call processing without forklift upgrades. Because market windows are small, first-mover advantage is a key element in the service provider's competitive arsenal.

WHY DO MULTI-CORE AND MULTI-THREADING MATTER?

The increase in central processing unit (CPU) cores and threads, coupled with dynamic random access memory (DRAM) density, dramatically increases the available throughput of today's entry-level servers. Unless the software running on those servers is optimized to use of all the cores, threads, and DRAM, the increased capacity will be wasted and the deployment will be inefficient.

It's all about cost: cost of the first subscriber and cost of large-scale implementations – from both a capital expenditure (CapEx) and operating expenditure (OpEx) perspective. Multi-core and multi-threaded implementations provide the ability for service providers to start small and scale fast in the most cost-effective manner.

WHY IS TRILLIUM MULTI-CORE SIP SOFTWARE IMPORTANT?

Simply put, Trillium® Multi-Core SIP software, which is optimized for multi-core and multi-threading processing environments, provides all of the characteristics which are important to service providers and telecom equipment manufacturers, including:

- Lower cost
- Higher capacity
- More predicable performance under load
- More headroom for revenue-generating SIP-based services

Global Headquarters
Continuous Computing

9450 Carroll Park Drive
San Diego, CA 92121 USA
T +1.858.882.8800
F +1.858.777.3388
info@ccpu.com

www.ccpu.com

In other words, Trillium Multi-Core SIP software delivers a "must-have" competitive advantage to service providers deploying SIP-based applications.

This Trillium Multi-Core SIP solution brief explores the details of the topics above and will:

- Analyze the factors driving the growth of SIP-based applications
- Explain the theoretical need for multi-core / multi-threaded software
- Provide performance benchmarks for Trillium Multi-Core SIP
- Show the advantage of Trillium Multi-Core SIP in a solution architecture

SIP MARKET DRIVERS

SIP-based services range from required utilities to new revenue-generating applications; a sample list includes Femtocell Gateways (FGWs), Convergence Servers (CSVRs), IP Multimedia Subsystem (IMS) elements, softswitches, mobile-to-fixed gateways, Voice over Internet Protocol (VoIP) elements, file sharing applications, voice or video conferencing, messaging, and platforms for gaming.

Several factors are working together to increase demand for these and other SIP services:

1. Proliferation of Smart Mobile Devices:

   The industry refers to the term "mobile device" rather than cell phone or smart phone – and rightly so. Technology makes it cheap and easy to add more features and media types to mobile devices, and all the devices' features and media types require servers in the network. Improved user experience on mobile devices increases the amount of SIP servers and SIP services required. There are 2.5 billion mobile devices in use globally today. Over time, more of these will become smart mobile devices, which will in turn drive demand for SIP-based services.

2. Increase in Fixed and Mobile Bandwidth:

   Higher fixed and mobile data rates improve users' experience of new features and media types (such as mobile video), which further increases demand. This "virtuous cycle" is amplified even more when combined with the aforementioned improvements in smart mobile device features and user interfaces, leading to a rich multimedia user experience.

3. Extensibility of SIP Itself:

   One could say that SIP will be to the converged network what the HTTP protocol is the Internet. In fact, the architects of SIP leveraged HTTP deliberately in order to make SIP as universal and as interoperable as possible. Because of its broad applicability and backward compatibility, as well as being a future-proofed foundation for new services, there is an ever-increasing suite of services that rely on SIP.

4. Convergence:

   The "converged network" can refer to fixed / mobile convergence (FMC), packet-based and circuit-switched calling, or both – and perhaps more. Regardless of definition, SIP is in the middle of it all, and SIP-based applications and services will be required to support convergence network servers well into the future.

5. Emerging Economies:

   The massive growth of working and middle class populations in emerging countries is fueling strong demand for mobile devices. For example, according to a recent Unstrung market research report, India added 8.6 million mobile devices in May 2008 alone – an incredible amount by any measure. This global increase in demand for SIP-based services must be delivered cheaply and quickly, and only low-cost, well-run operators will be able to survive in such dynamic emerging economies.

6. Mobile Commerce:

   Already in widespread use in Asia for paying mass transit fares, subscribers can use smart mobile devices to purchase goods similar to how credit and debit cards are used. In Africa, mobile device minutes have become currency among villagers and townspeople. As these and other types of mobile commerce proliferate, they will require more SIP-based infrastructure, thereby perpetuating demand.

SIP MULTI-CORE PERFORMANCE: THEORY

Two oft-quoted technology maxims are applicable to the increasing demand for SIP: Moore's Law and Amdahl's Law.

Moore's Law states that circuit density doubles approximately every 18 months. Today, that means CPU core or thread density doubles every 18 months. As of this writing in mid-2008, commodity servers are based on two Intel Architecture processors with eight total processing cores. Sun Microsystems' entry-level servers feature the UltraSPARC® T2 processor with chip multithreading (CMT) and have 64 total threads (eight cores with eight threads each).

Amdahl's Law states that a high degree of parallelization is needed as the number of cores or threads increases in order to achieve an improvement in overall throughput. So, for example, even if 90% of a software code implementation is "parallelized", one half of the throughput capacity of an eight-core server is still wasted. With Sun's CMT servers – boasting 64 total threads – the need for high throughput, multi-threaded software is even greater if overall performance improvement is to be achieved (see Table 1 and Figure 1).

Cores / Threads	99% Parallel	95% Parallel	90% Parallel	75% Parallel	50% Parallel
1	1.0	1.0	1.0	1.0	1.0
2	2.0	1.9	1.8	1.6	1.3
4	3.9	3.5	3.1	2.3	1.6
8	7.5	5.9	4.7	2.9	1.8
16	13.9	9.1	6.4	3.4	1.9
32	24.4	12.5	7.8	3.7	1.9
64	39.3	15.4	8.8	3.8	2.0

Table 1. Effective throughput as a function of cores and parallelization.

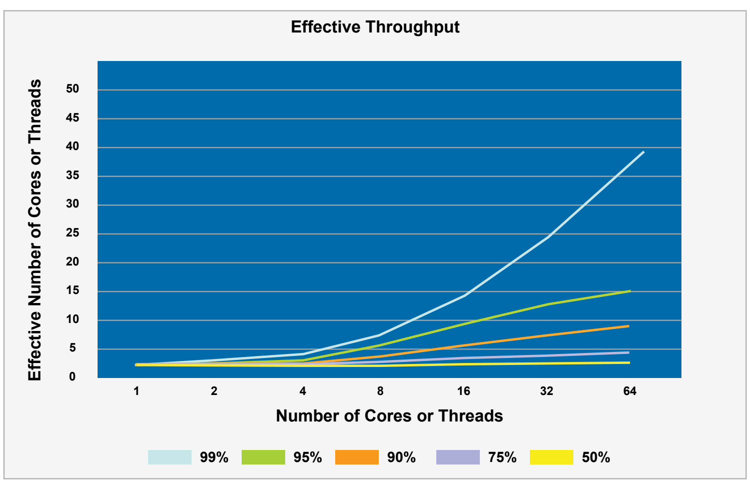

Figure 1. Very high parallelization is needed for overall improvement in throughput.

In other words, as CPU core (or thread) count increases – which is certainly the trend for the foreseeable future – multi-core / multi-threaded software becomes ever more necessary.

Amdahl's Law may also be viewed in terms of an efficiency calculation. In other words, how much actual performance is delivered as a percent of the theoretical maximum CPU performance? Obviously higher efficiency is better, and the table and chart below show how important parallelization (i.e., using multi-core / multi-threaded software) is to achieving top efficiency (see Table 2 and Figure 2).

Cores / Threads	99% Parallel	95% Parallel	90% Parallel	75% Parallel	50% Parallel
1	100%	100%	100%	100%	100%
2	99%	95%	91%	80%	67%
4	97%	87%	77%	57%	40%
8	93%	74%	59%	36%	22%
16	87%	57%	40%	21%	12%
32	76%	39%	24%	11%	6%
64	61%	24%	14%	6%	3%

Table 2. Efficiency as a function of cores and amount of parallelization.

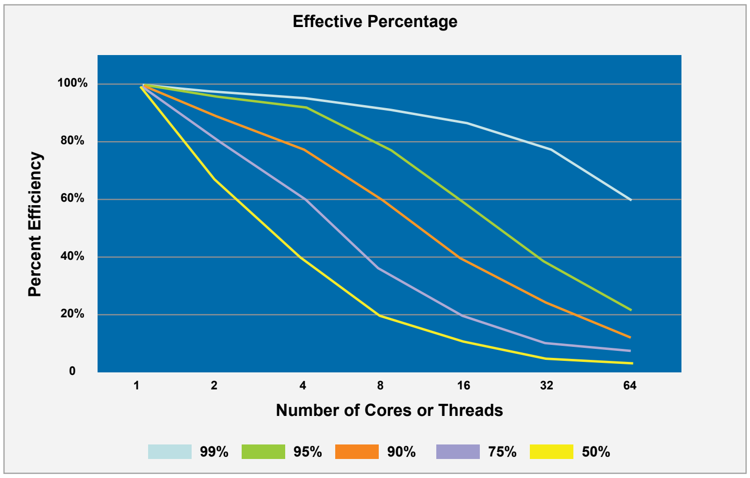

Figure 2. Very high parallelization is needed for overall improvement in efficiency.

The point is not just about CPU core / thread utilization or efficiency as a theoretical construct of optimization; there are practical advantages as well. All of the ancillary costs of a deployment – DRAM, disks, rack space, power, and cooling – are wasted unless the software takes advantage of its multi-core / multi-threaded processing environment. With massive SIP-based service deployments coming online, these cost realities will have a profound impact on service providers' profitability and competitiveness.

SIP MULTI-CORE PERFORMANCE: BENCHMARKS

Trillium Multi-Core SIP software from Continuous Computing is optimized for the multi-core / multi-threading processing environment. Trillium Multi-Core SIP provides lower cost, higher capacity, more predicable performance under load, as well as more headroom for revenue-generating SIP-based services. In short, it provides all of the characteristics important to service providers and telecom equipment manufacturers.

To illustrate the breakthrough scalability of Trillium Multi-Core SIP, Continuous Computing tested the software on a Sun Netra™ T5220 server with the UltraSPARC T2 processor featuring Sun's CMT architecture with eight cores and 64 threads (see Figure 3 and www.sun.com/servers/netra/t5220).

Figure 3. Sun Netra T5220 server with the UltraSPARC T2 processor.

The results were profound: Trillium Multi-Core SIP was able to efficiently utilize all available 64 processing threads to achieve unmatched performance – 6,000 calls per second, an improvement of at least 2X when compared with Trillium SIP on any previous generation multi-core or multi-threaded processor tested. This is a significant improvement.

And, because the performance of Trillium Multi-Core SIP scales linearly with increased load, service providers not only achieve phenomenal cost control, but they can also guarantee Service Level Agreements (SLAs) to their end-user customers. This is a key advantage for operators to maximize revenue, improve end-user satisfaction, and reduce subscriber churn.

Details of the Trillium Multi-Core SIP benchmark test setup and the results are as follows:

- Powerful Sun Netra T5220 server running the eight-core, 64-thread UltraSPARC T2 processor and Solaris™ 10 operating system
- High-performance Trillium Multi-Core SIP call stateful server software
- SIPp performance tester acting as both a User Agent Client (UAC) and User Agent Server (UAS), caller and callee respectively

Calls per Second (CPS)	Messages per Second (Sent + Received)	TCP CPU Utilization (mpstat)	UDP CPU Utilization (mpstat)
500	6,000	6	6
1,000	12,000	11	12
1,500	18,000	17	18
2,000	24,000	23	24
2,500	30,000	29	30
3,000	36,000	35	36
3,500	42,000	42	42
4,000	48,000	49	48
4,500	54,000	56	55
5,000	60,000	63	62
5,500	66,000	70	69
6,000	72,000	78	76

Table 3. Benchmark test results showing 6,000 CPS with Trillium Multi-Core SIP.

Again, the highlight of the benchmark test is that the end-to-end SIP network in a single server achieved 6,000 calls per second, a phenomenal performance statistic representing today's state of the art.

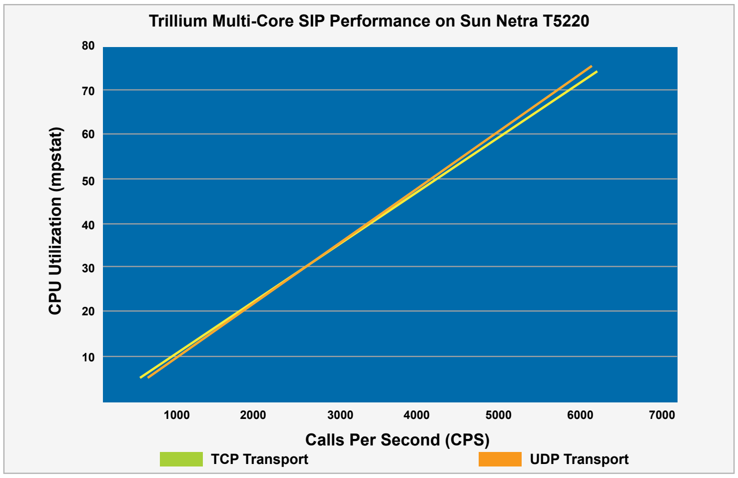

Figure 4. Trillium Multi-Core SIP performance on the Sun Netra T5220; performance scales linearly with load.

SIP MULTI-CORE SOLUTION ARCHITECTURE ADVANTAGES

Fixed / mobile convergence (FMC) servers, also known as Convergence Servers (CSRV), rely on SIP to convert from IP mobility to legacy transport protocols.

Emerging technologies such as Femtocells (i.e., home base stations) rely on CSRVs to create seamless service provisioning to subscribers as they migrate from one air interface to another – such as from a 3G macrocell network to an indoor Femtocell, or from one Femtocell to another.

Femtocells today are in the trial stage; numerous operators worldwide are rolling them out in targeted cities to test various architecture options as well as different end-user pricing plans. Market research firms project that Femtocell usage will grow rapidly due to the very low cost of Femtocell transceivers – but first, service providers want to start small and make refinements before scaling up for mass deployment.

As shown in Figure 5 below, Femtocell applications such as user-defined features and preferences, VLR (Visiting Location Register) access, mobility, and E911 can co-exist with macrocell network SIP call processing and even legacy transport protocols like SS7 and SIGTRAN. This seamless coexistence allows for cost-effective trials because in order to scale up to large deployments, the same architecture "just works" if it is based on software optimized for the multi-core environment – such as Trillium Multi-Core SIP.

Lastly, as described previously in the benchmark testing section, Trillium Multi-Core SIP on Sun CMT scales linearly, which allows for SIP protocol processing and applications to share the server. Such co-existence not only saves in CapEx (i.e., fewer servers) but it also saves in OpEx by reducing costs due to complexity, power, cooling, floor space, and the need for equipment spares.

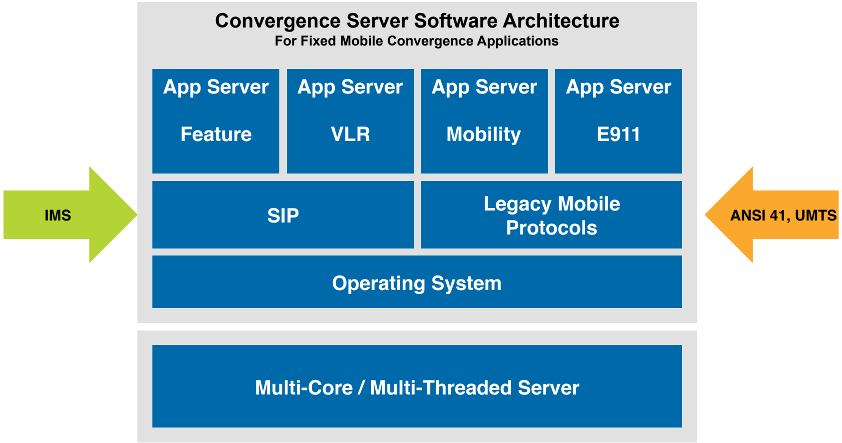

Figure 5. Software architecture of a typical Convergence Server; SIP is fundamental to its operation.

CONCLUSIONS

SIP is the foundation for most emerging communication services, and SIP-based applications are projected to grow exponentially. There are multiple market drivers for these services and for high SIP call processing throughput in general, such as a large increase in smart mobile devices, a rapid increase in fixed and mobile bandwidth, explosive subscriber growth, and the coming mass adoption of mobile commerce.

The processing reality of today and the foreseeable future is all about multi-core and multi-threading. Both theory and actual test results show that in order to take advantage of the multi-core processing environment, software needs to be architected accordingly. And following Moore's Law – with core and thread counts doubling every 18 months for commodity servers – and Amdahl's Law – where parallelization is needed in the multi-threaded environment in order to maximize overall throughput – operators that don't take advantage of applications optimized for multi-core systems will quickly lose out to competitors due to cost, feature, and performance limitations.

Because Trillium Multi-Core SIP software is optimized for the multi-core / multi-threading processing environment – providing lower cost, higher capacity, more predicable performance under load, and more headroom for revenue-generating SIP-based services – it is the ideal solution for operators looking to start small and scale fast while maximizing efficiency, utilization, and performance.

The bottom line is that Trillium Multi-Core SIP software delivers a "must-have" competitive advantage to operators deploying SIP-based applications by delivering all of the characteristics important to service providers and telecom equipment manufacturers.

ABOUT CONTINUOUS COMPUTING

Continuous Computing® provides integrated systems and services that enable telecom equipment manufacturers to rapidly deploy Next Generation Networks (NGN). More than 150 customers worldwide benefit from the company's unique blend of customized professional services, Trillium® protocol software, AdvancedTCA and CompactPCI systems, and BladeCenter hardware. Continuous Computing helps customers reduce platform lifecycle costs, optimize data delivery, and accelerate deployments of NGN, 3G Wireless, and IP Multimedia Subsystem (IMS) infrastructure. The company is ISO-9001 and CMMI certified and is based in San Diego with development centers in China and India. For more information, visit **www.ccpu.com**

Continuous Computing, the Continuous Computing logo, Create | Deploy | Converge, Flex21, FlexChassis, FlexCompute, FlexCore, FlexDSP, FlexPacket, FlexStore, FlexSwitch, FlexTCA, Network Service-Ready Platform, Quick!Start, TAPA, Trillium, Trillium+*plus*, the Trillium logo, upBeat, upDisk and upSuite are trademarks or registered trademarks of Continuous Computing Corporation.

Sun, Solaris and Netra are trademarks or registered trademarks of Sun Microsystems, Inc. or its subsidiaries in the United States and other countries.

Other names and brands may be claimed as the property of others.

Continuous Computing®

©2008 Continuous Computing Corporation.

Continuous Computing, the Continuous Computing logo, Create | Deploy | Converge, Flex21, FlexChassis, FlexCompute, FlexCore, FlexDSP, FlexPacket, FlexStore, FlexSwitch, FlexTCA, Network Service-Ready Platform, Quick!Start, TAPA, Trillium, Trillium+*plus* and the Trillium logo are trademarks or registered trademarks of Continuous Computing Corporation.

Other names and brands may be claimed as the property of others. MC00180-01 0708/TM/BP/EC

Global Headquarters

9450 Carroll Park Drive
San Diego, CA 92121 USA
T +1.858.882.8800
F +1.858.777.3388
info@ccpu.com

www.ccpu.com