Auditoria de senhas em hardware paralelo com o John the Ripper

O impacto das tecnologias de processamento paralelo na quebra de senhas

Claudio André claudio.andre@correios.net.br

Motivação

- Seu computador é um dual core, quad core, hexa core?
- Você sabe como aproveitar todo este poder de processamento?
- OpenCL, OpenMP.

Motivação

- Novas GPUs oferecem centenas, algumas milhares de núcleos de processamento!
- De novo: você sabe como aproveitar todo este poder de processamento?
- Este grande poder de processamento aumenta os riscos de segurança?

Contexto

- Segurança depende de problema computacionalmente difícil;
 - RSA: dificuldade de se fatorar um número inteiro grande;
 - SHA-512: uma senha grande necessita de trilhões de tentativas para obter sucesso.

Contexto

Mas:

- Grande avanço na computação paralela e distribuída;
- GPGPU ("antigas" placas gráficas) com 2048 processadores (stream processors).

Objetivo

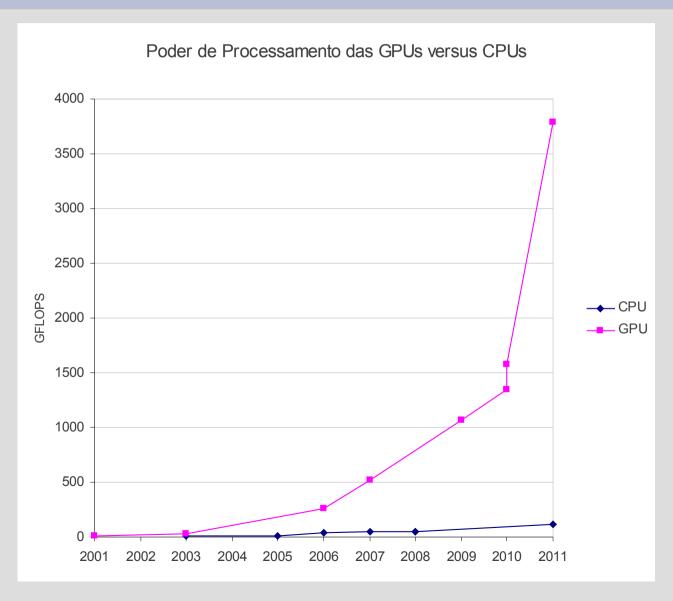
- Mostrar que é possível usar uma GPGPU para quebrar a senha do Linux (SHA-512).
- Mostrar que o investimento necessário é baixo.
- Mostrar que a relação custo x resultado é boa.

Mais que isto!

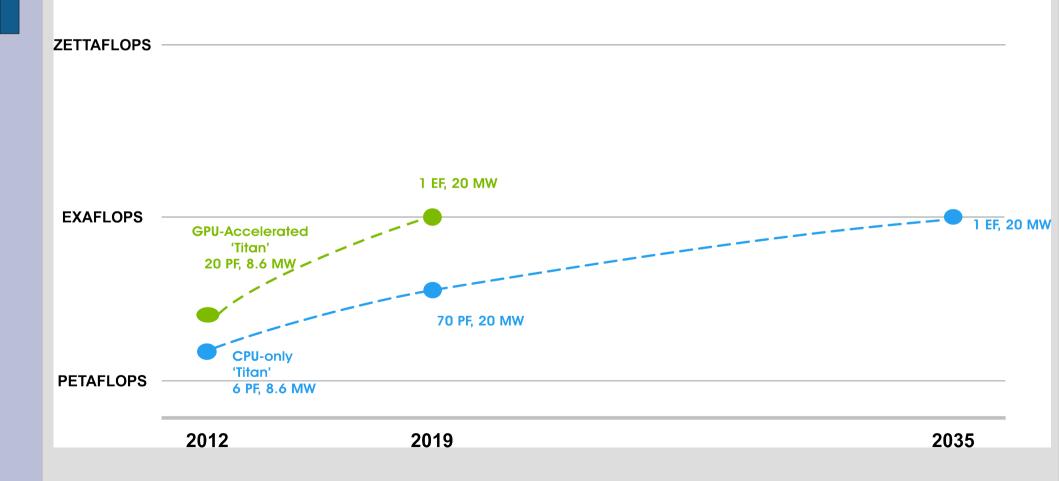
- Mostrar que uma função de hash de senha pode ser implementada na GPU.
- Qual o impacto nos atuais mecanismos de autenticação?
- A relação custo x benefício é muito boa.

Por que a GPU?

- Suporta programas de propósito geral;
- Oferece grande paralelismo de processamento.
- Quebra de senhas é um problema facilmente paralelizável.

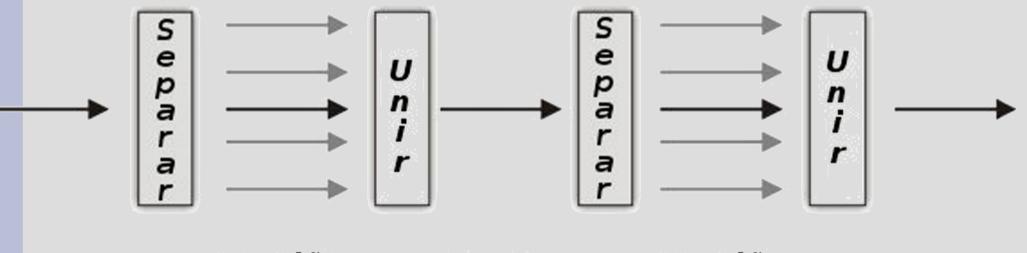


The Road to Exascale



O problema das senhas

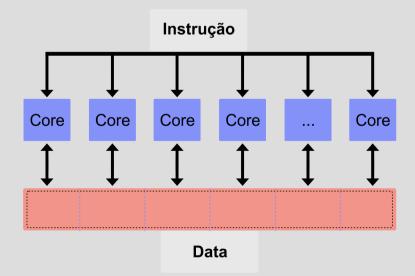
- Capacidade de memorizar a senha;
- Capacidade de gerar senhas aleatórias;
- Espaço de caracteres é limitado.
- Aumentar a entropia das senhas sem aumentar a do suporte?



Região Paralela Região Serial Região Paralela

SIMD

Uma instrução, múltiplos dados

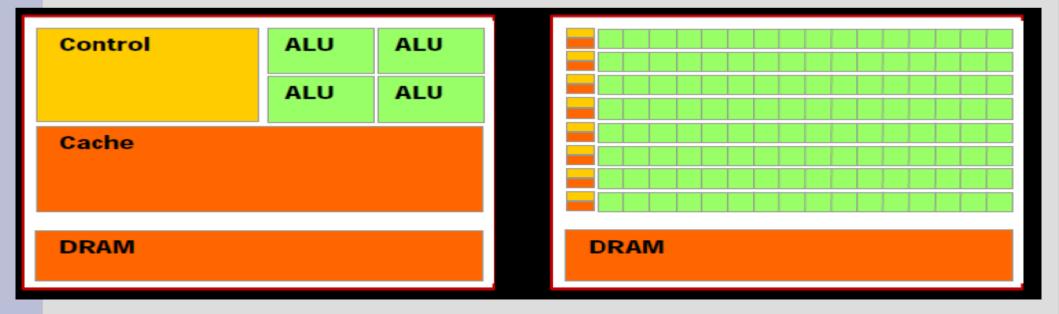


Cache, foco em desempenho de uma thread individual.

Note a área reservada para o controle.

ALU massivamente paralela.

Note que a área reservada para o controle é muito menor.



Como programar para GPU:

- OpenCL;
- CUDA;

Antes de mais nada:

- Como você, meu colega programador, acessa todo o poder de processamento da sua CPU?
- Você consegue colocar todos os núcleos pra trabalhar?

OpenCL:

 Open Computing Language. Proposta de linguagem padrão e aberta para programação paralela em ambientes heterogêneos (The open standard for parallel programming of heterogeneous systems).

CUDA:

Compute Unified Device Architecture.
 Arquitetura de dispositivo unificado de computação. Proposta de arquitetura de computação paralela criada pela NVIDIA

Resultados na CPU:

#	Tempo de Processamento Sequencial	Tempo com Três Núcleos	Ganho	Tempo com Seis Núcleos	Ganho
1	00:13	00:05	2,60	00:03	4,33
2	06:55	02:26	2,84	01:24	4,94
3	24:53	08:47	2,83	05:09	4,83
4	24:03	08:28	2,84	04:54	4,91
5	05:00:33	01:46:17	2,83	01:00:53	4,94
6	08:35:27	03:02:41	2,82	01:45:58	4,86
7	01:25:38	30:18	2,83	17:28	4,90
8	31:12:57	11:00:34	2,84	06:19:24	4,94

Resultados na CPU:

#	Tempo de Processamento Sequencial	Constante de Penalização Três Núcleos	Tempo Esperado ao Paralelizar	Constante de Penalização Seis Núcleos	Tempo Esperado ao Paralelizar
1	00:13	0,87	Sequencial 0,9 * núcleos	0,72	Sequencial 0,7 * núcleos
2	06:55	0,95	Sequencial 0,9 * núcleos	0,82	Sequencial 0,8 * núcleos
3	24:53	0,94	Sequencial 0,9 * núcleos	0,81	Sequencial 0,8 * núcleos
4	24:03	0,95	Sequencial 0,9 * núcleos	0,82	Sequencial 0,8 * núcleos
5	05:00:33	0,94	Sequencial 0,9 * núcleos	0,82	Sequencial 0,8 * núcleos
6	08:35:27	0,94	Sequencial 0,9 * núcleos	0,81	Sequencial 0,8 * núcleos
7	01:25:38	0,94	Sequencial 0,9 * núcleos	0,82	Sequencial 0,8 * núcleos
8	31:12:57	0,95	Sequencial 0,9 * núcleos	0,82	Sequencial 0,8 * núcleos

Resultados na GPU:

#	Tempo Máquina I	Tempo Máquina II	Tempo GPU	Ganho Máquina I	Ganho Máquina II
1	00:00:03	00:10	00:00:17	0,18	0,59
2	00:01:24	05:05	00:07:43	0,18	0,66
3	00:05:09	18:18	00:04:04	1,27	4,50
4	00:04:54	10:50	00:04:50	1,01	2,24
5	01:00:53	03:14:42	00:41:08	1,48	4,73
6	01:45:58	06:19:15	01:10:29	1,50	5,38
7	00:17:28	01:03:27	00:12:43	1,37	4,99
8	06:19:24	-	04:12:10	1,50	-

Mas:

• Eu vi todo tipo de esquisitice do compilador;

Aprendeu isto!

$$\lim_{x \to 8} \frac{1}{x-8} = \infty$$

Mas ainda não faz o trabalho direito!

$$\lim_{x \to 5} \frac{1}{x-5} = \omega$$

Portanto, planeje antes de tentar resolver seus problemas na GPU.

Descubra como você vai paralelizar seu problema!

A camisa que me serve não tem o mesmo caimento em você!

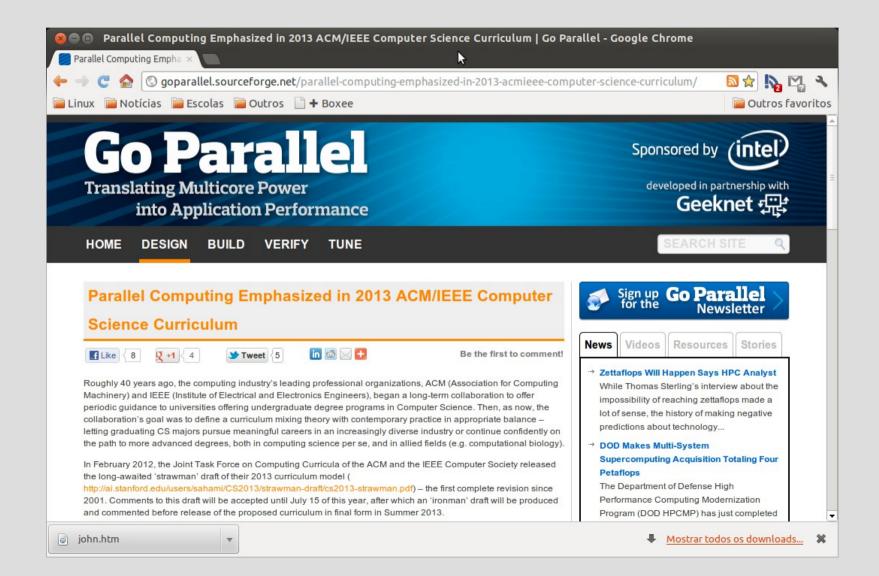


Contudo, o processamento paralelo é o futuro!! A nova onda chegará!!



Você tem alternativas à computação paralela?





Quanto mais rápido?

- Radeon HD 6770
 - 4 vezes mais que um notebook Turion™ II Mobile P560 2,50 GHz
 - 1,5 vezes mais que o AMD Phenom™ II X6 1075T
 - 7 vezes mais que o citado Turion™ II Mobile P560 (1 core)
 - 5 vezes mais rápida que o AMD X6 1075T (1 core)
 - R\$ 350,00 (eu paguei menos que isto)
 - Placa "modesta"
 - Você pode fazer 4 vezes o que seu bom e velho "dual core" faz
 - Se tiver um slot livre, usando o seu bom e velho "dual core"

Quanto mais rápido?

GTX 570

- 5,6 vezes mais que o AMD FX™-8120 Eight-Core Processor
- 5,1 vezes mais que o AMD Phenom™ II X6 1075T
- 31,4 vezes mais que o citado FX™-8120 (1 core)
- 26,2 vezes mais rápida que o AMD X6 1075T (1 core)

Radeon HD 7970

- 4,3 mais que o AMD FX™-8120
- 3,9 mais que o AMD Phenom™ II X6 1075T
- 23 vezes mais que o citado FX™-8120 (1 core)
- 20 vezes mais rápida que o AMD X6 1075T (1 core)

Quanto mais rápido?

- GTX 570
 - Placa intermediária (NÃO É TOPO DE LINHA)
 - Você precisa de 5 computadores top de linha para fazer o mesmo
 - E de um programador que saiba tirar tudo deste computador
 - Caso contrário, seriam 26 computadores top de linha
 - Na Amazon a placa custa US\$ 300,00
 - No Brasil, R\$ 1.000,00
 - Cada computador turbinado destes custaria uns R\$ 2.000,00
 - De R\$ 10.000,00 você faz o mesmo trabalho com R\$ 1.000,00

Conclusões:

- Pequeno investimento gerou ganho de desempenho adequado;
- Porém, esperava mais considerando as especificações da GPU AMD; SI ou NI;
- OpenCL ainda é tecnologia em evolução;

Conclusões:

- Os riscos de segurança ficaram maiores;
- GPUs seguem evoluindo;
- Programação paralela está em nosso futuro.
- Programadores precisam conhecer esta nova tecnologia.

Conclusões:

Cursos? Onde?

Jamais se sinta "seguro":

- Hacker desenvolve método para burlar chip mais seguro (2010);
- Certificado digital da Microsoft foi usado para assinar o malware Flame (2012);
- Crackers vazam 6,4 milhões de senhas do LinkedIn (2012).

Obrigado

Mais Informações

http://www.openwall.com/john/

http://openwall.info/wiki/john/OpenCL-SHA-512

www.claudioandre.drivehq.com/outros/john.htm

Claudio André claudio.andre@correios.net.br