Promise FastTrak TX2000

Stručný manuál

Autor: František Ryšánek <rysanek@fccps.cz>

FCC Průmyslové Systémy s.r.o.

Řadič Promise FastTrak TX2000 je IDE řadič, na kterém lze nakonfigurovat "diskové pole", anglicky RAID (Redundant Array of Inexpensive Disks) - virtuální disk skládající se z více disků.

Jedná se o levné RAID řešení, proto umí pouze RAID level 0, 1 a JBOD s tím, že pole zabere na označených discích vždy celou dostupnou kapacitu (nelze vytvářet více logických disků). Jedná se o softwarový RAID - viz též samostatný dokument o typech polí.

Většina zákazníků si pořizuje pole kvůli spolehlivosti, proto
důrazně doporučujeme RAID 1 (mirror)!
(V BIOSu řadiče je tato volba poněkud nesprávně označena anglickým slovem "security".)

S nastavením řadiče a vytvořených polí lze pracovat jednak z nástroje, který je součástí BIOSu řadiče a lze ho vyvolat klávesou Ctrl+F při startu počítače (ještě než nastartuje operační systém), nebo pomocí utility nainstalované v operačním systému Windows.

Utilita pod Windows je komfortnější a je vhodná pro opravu pole za provozu. Nástroj v BIOSu lze použít v případě, kdy systém Windows není k dispozici.

Počáteční nastavení pole

K vytvoření pole jsou potřeba alespoň dva disky, připojené k řadiči Promise FastTrak (tj. nikoli např. k IDE řadiči integrovanému na motherboardu).

Při startu PC

proběhnou úvodní hlášky BIOSu, tabulka s parametry počítače. Pak se objeví hláška "scanning IDE Devices...", kterou již vypisuje BIOS řadiče Promise FastTrak. Vzápětí se objeví výzva "stiskněte Ctrl+F".

Stiskněte Ctrl+F.
stiskněte "1" (= Auto-setup)
-> zvolte "Security" (= Mirror)
-> zvolte "Create only" (vyrobí prázdné pole bez zdlouhavého kopírování stávajících dat)

Instalace Windows2000

Windows 2000 nepoužívají pro přístup k disku BIOS a neobsahují ovladače pro řadič FastTrak TX2000. Proto Windows při standardní instalaci pole nenajdou a nahlásí, že se nemají kam nainstalovat.

Je třeba instalátoru ve vhodný okamžik podstrčit disketu s ovladači - to se dělá takto:

- vložte CDrom Windows2000 do mechaniky
- nastartujte počítač - pokud je správně nastaveno pořadí bootovacích zařízení v BIOS setupu, nastartuje z CDčka intalátor Windows2000
- jakmile se začne rozbíhat instalátor Windows2000, čekejte pozorně na výzvu ke stisku F6 (tato výzva po několika sekundách zmizí beze stopy). Stiskněte včas klávesu F6.
- po stisku F6 instalátor chvíli pracuje na disku a pak se zeptá na disketu s ovladači. Stiskněte "S", vložte disketu, stiskněte <ENTER> - a pokračujte v instalaci.

Instalace aplikace pro správu pole v systému Windows

"PAM Utilita" (PAM = Promise Array Management) se nachází na CD od výrobce. Pro správnou funkci pole (včetně obnovy po havárii) není tato utilita striktně nutná – nicméně doporučujeme ji nainstalovat, protože umožňuje mnohem lepší přehled o situaci než holý ovladač.

Po vložení CD do mechaniky v systému Windows se automaticky spustí základní obrazovka instalátoru, ve které lze spustit vlastní instalaci.

Protože jde ve skutečnosti o client-server systémek, bude instalace klienta chtít IP adresu - pokud budete utilitu používat pouze pro přístup k lokálnímu stroji, ponechte standardní nastavení, tj. 127.0.0.1, uživatel "administrator" (s malým "A"), bez hesla.

Po instalaci PAM utility je třeba počítač restartovat - utilita pole nenajde dříve než po prvním restartu.

Po restartu spustíme nainstalovanou utilitu. Ve skutečnosti se v pozadí automaticky rozběhly také

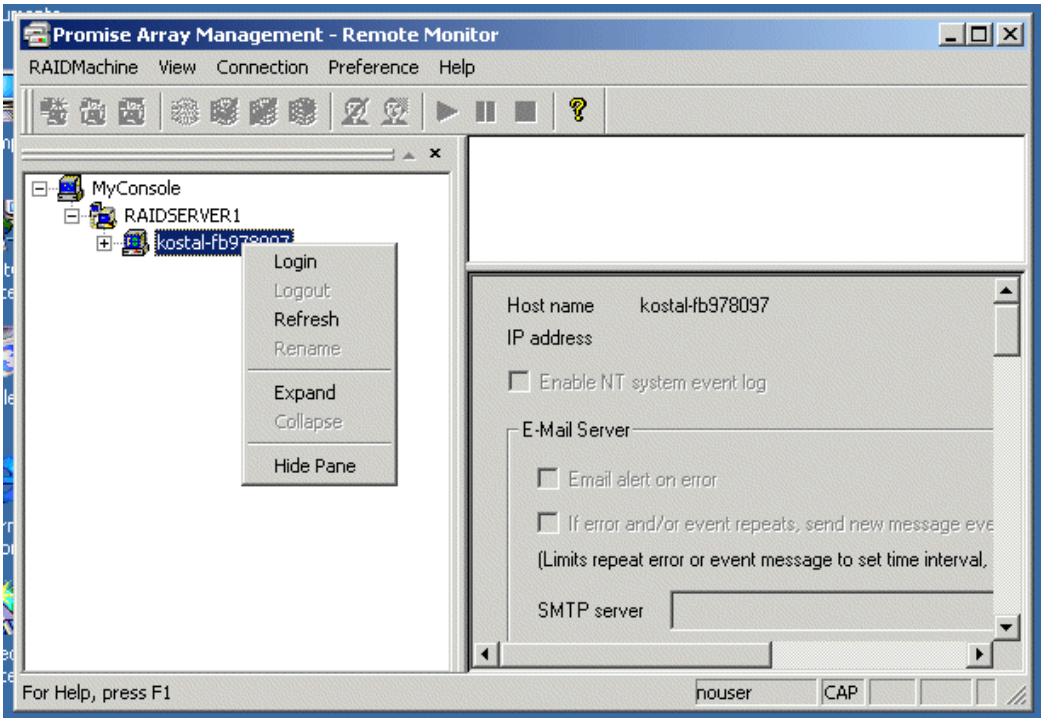
dvě nové služby.

V levém okně se stromovým prohlížečem je vidět kořen stromu zvaný “MyConsole”, dostupné "message servery", jim podřízení "message agenti".

Podle spartánské nápovědy se jedná o dvoustupňový client-server management, distribuovaný po lokální síti. Je možné nastavit jeden počítač v síti jako server, který má přehled o polích na několika podřízených strojích. Proto ten složitý strom.

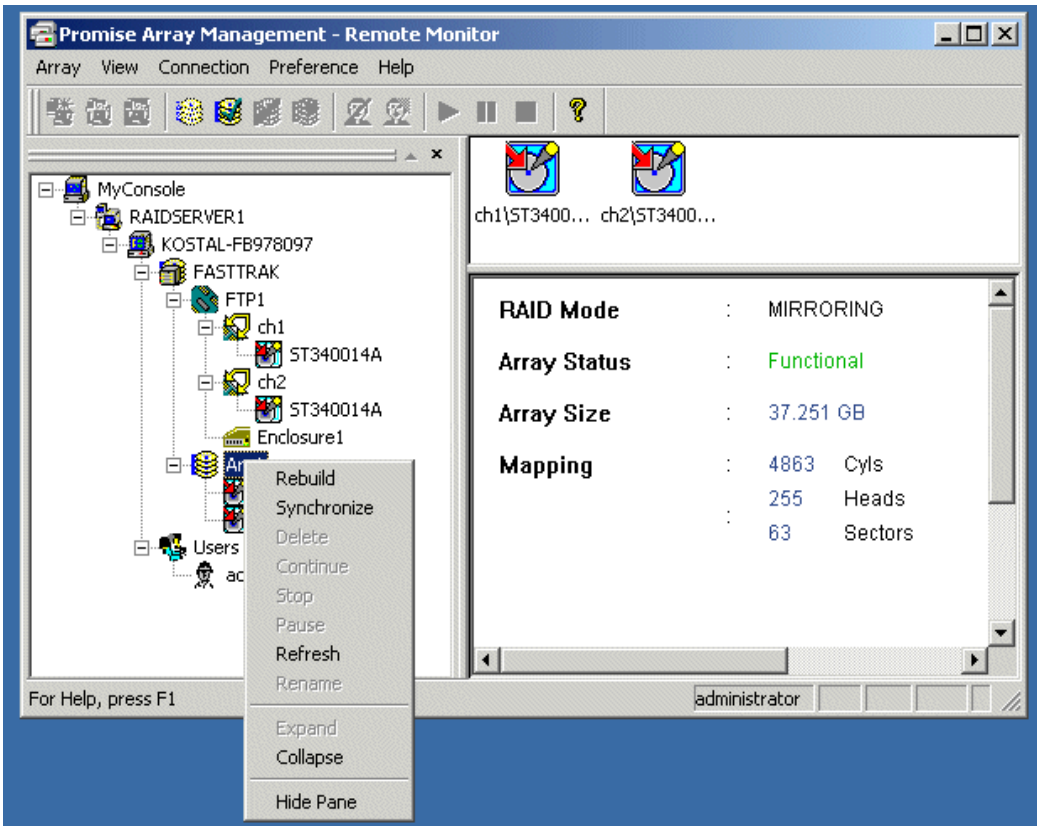
V stromovém prohlížeči nalevo se tedy lokální "message agent" objeví jako jméno "tohoto počítače". Označte ho.

Nyní je třeba se nalogovat. To lze provést z kontextového menu (pravým myšítkem na označený stroj) nebo z menu ServerMachine, které se automaticky objevilo úplně vlevo na horní liště.



Standardní login je "administrator" (malé "A") bez hesla.

Po nalogování se objeví další ikonky (větévky stromu) - je možno prohlížet uživatele, řadiče, kanály, pole a fyzické disky.



Příkaz pro obnovení pole po havárii se jmenuje "rebuild" a objevíte ho v kontextovém menu po označení konkrétního pole (ikonka s několika žlutými plotnami a modrým výložníkem hlav).

Jak se pozná havárie pole

Když něco není v pořádku, počítač pravidelně pípá v intervalu asi 1 s. Toto ovšem platí pouze v případě, že běží PAM utilita pro windows a je přilogovaná k lokálnímu "message serveru".

Konkrétní příčinu lze zjistit v nástroji dostupném při startu přes BIOS, nebo v PAM utilitě pod Windows.

Bohužel není možno "identifikovat" havarovaný disk záměrným blikáním LEDdiodou, jako to umějí některé SCSI řadiče. Dostupná je pouze informace na kterém kanálu řadiče se havarovaný disk nachází a zda se jedná o master nebo slave.

Pokud v disku úplně odejde elektronika nebo je fyzicky odpojen, pod Windows mu naskočí status "critical", ale v nástroji dostupném přes BIOS se již vůbec neobjeví. Čili pokud hledáme takto mrtvý disk v BIOSu, je to ten, který podle kanálu a pořadí v seznamu schází.

Obnova pole po havárii disku

Obnovit lze pouze pole typu RAID 1 (mirror).

Vždy je třeba nejprve vytáhnout havarovaný disk a nahradit ho novým. Pokud jsou ve stroji výměnné šuplíky na IDE disky a každý disk je sám na svém kanálu, funguje hot-swap – při výměně disku není třeba vypínat počítač, ani restartovat Windows. Při obnovitelné poruše pole by měl počítač nastartovat a běžet i s jedním funkčním diskem.

Při výměně disku u tohoto řadiče nezáleží na tom, zda byl náhradní disk "factory clean", nebo zda již byl součástí nějakého pole (a nebyl náležitě smazán).

Pod Windows by měla obnova pole nastartovat automaticky po výměně vadného disku. Lhostejno, zda byl disk vyměněn bez restartu (hot swap) nebo s restartem (při vypnutém počítači).

Automatickou obnovu spustí samotný driver řadiče – není ani třeba mít nainstalovánu PAM aplikaci a její dvě rezidentní služby.

Pokud se obnova pole nespustí automaticky, lze ji spustit ručně jedním z následujících dvou způsobů:

1) obnova pole v nástroji dostupném z BIOSu

Pokud nevíme, který disk zhavaroval, musíme nabootovat nebo restartovat ještě s vadným diskem namontovaným. Po úvodních hláškách BIOSu se objeví obsáhlé varování, které říká dohromady pouze tolik, že "něco někde je špatně". (Pokud nyní stiskneme Escape, systém by se měl pokusit o start s jedním diskem.)

Stiskem Ctrl+F se dostaneme do menu RAID řadiče. Volbou "2" (display drive assignments) si zobrazíme disky a zjistíme, co si o nich řadič myslí. Jeden z disků bude označen jako vadný, nebo bude v seznamu chybět. Podle čísla kanálu a pozice na kabelu je třeba zjistit, o který disk se jedná, vypnout počítač a vyměnit jej. Pokud máme kontrolku ke každému disku zvlášť, třeba na výměnných šuplíkách, lze vadný disk za běhu poznat i na pohled.

Při prvním startu po výměně vadného disku řadič opět zobrazí obsáhlou chybovou hlášku. Stiskem Ctrl+F se dostaneme do menu RAID řadiče. Před opravou pole ještě zkontrolujeme v menu "2" (drive assignments), zda jsou vidět oba disky - vyměněný disk by měl být označen jako "free".

Nyní zvolte možnost "5" - rebuild (obnova pole). Budete dotázáni, který disk se má použít jako náhradní - dostanete na vybranou pouze z disků, které dosud nejsou přiřazeny do nějakého pole.

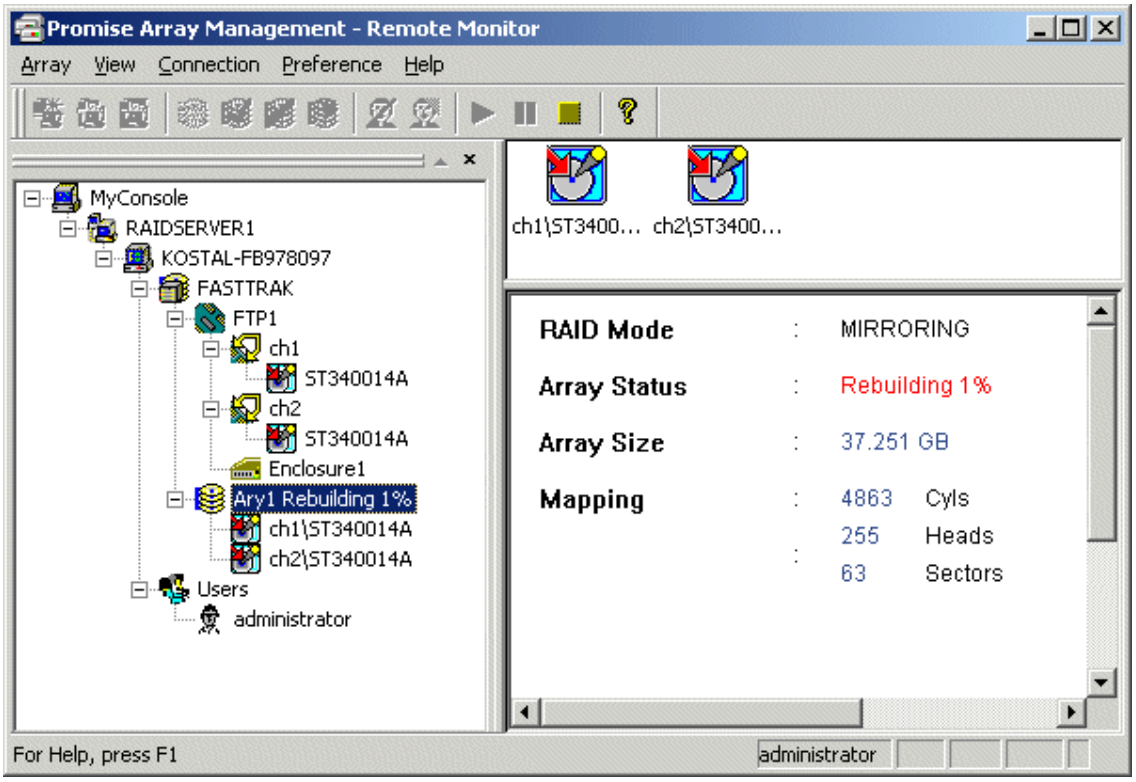
Vyberte vyměněný disk a potvrďte klávesou "enter".

Spustí se obnova. Obnovu lze přerušit pouze tvrdým resetem - systém v průběhu obnovy nereaguje na klávesnici.

2) obnova pole pomocí PAM utility pod Windows

Pokud provedete výměnu disku za běhu, počkejte pár desítek vteřin, než si řadič osahá nový disk. Měl by ho najít (havarovaný disk zmizí ze stromu a objeví se nový).

Vzápětí by se měla automaticky rozběhnout obnova pole, což je indikováno červenou popiskou "rebuilding" u ikonky pole.



Pokud by se obnova pole nerozběhla automaticky, lze ji spustit ručně z kontextového menu na ikonce pole - příkaz se jmenuje "rebuild".

Poznámka: Pokud proces obnovy přerušíte restartem počítače, pole po restartu pokračuje s obnovou tam, kde skončilo. To platí pro přerušení a pokračování obnovy v BIOSu i ve Windows, a to i ve vzájemné kombinaci.

Obnova pole po havárii řadiče

Stačí vyměnit řadič za nový a připojit disky na původní kanály/pozice (nepoplést pořadí).

Prakticky to znamená nejlépe předem zkontrolovat nový řadič (spuštěný bez připojených kabelů), zda má povoleny všechny (oba) potřebné kanály apod. Konfigurace pole je uložena na discích a měla by se automaticky načíst a spustit.

Podpora řadiče pod Linuxem

Pod Linuxem existuje několik možností, jak řadič provozovat. Každá má své výhody a nevýhody. Linux tento řadič zná pod názvem PDC20271. Informace v tomto materiálu se týkají linuxového jádra ve verzi okolo 2.4.20 – řekněme 2.4.18 až 2.4.22.

1) open-source driver řadiče + ataraid

PDC20271 AKA FastTrak TX2000 je v zásadě standardní IDE řadič (UATA/133), který má v Linuxu svůj driver.

Do verze 2.4.20 vanilla se tento driver jmenoval drivers/ide/pdc202xx.c a vlastně v něm řadič PDC20271 nebyl podporován. Pouze RedHat 8.0 s upraveným jádrem 2.4.18-14 již tento nový řadič podporuje. Od verze 2.4.21 vanilla (v –ac patchích o něco dříve) se driver rozdělil na pci/pdc202xx_old.c a pci/pdc202xx_new.c – čip PDC20271 je podporován větví "new". Tyto drivery se v 'make menuconfig' nacházejí ve větvi "ATA/IDE/MFM/RLL" atd. Pro tento čip není zapotřebí povolovat "FastTrak feature" (je irelevantní, nemá vliv).

Funkce RAIDu vytváří modul ataraid, podpora pro PDC20271 je obsažena v podřízeném modulu pdcraid.c. V 'make menuconfig' tomu odpovídají dva body v posledně jmenované větvi týkající se IDE. Tyto dva body se jmenují "Support for IDE RAID controllers" a jemu podřízený "Support for Promise RAID controllers".

Driver pdcraid.c v době psaní tohoto článku obsahoval bug v detekci pole – v některých případech špatně si spočítal LBA offset režijního bloku dat na fyzických discích a následně nenašel pole.

Autorovi se podařilo tuto chybu opravit, jak dokládá patch umístěný na http://sweb.cz/Frantisek.Rysanek/pdcraid/

Výsledný celek přesto rozhodně nelze pokládat za použitelné RAIDové řešení. Spíše by se dalo říci, že ta věc se sotva drží nad vodou.

Z pole lze číst a bootovat, lze na něj zapisovat, ale to je tak všechno. Protože neexistuje správcovská utilita, nelze kontrolovat stav pole nebo je udržovat: měnit přiřazení disků, pole rušit, vytvářet a především obnovovat. Pouze v /proc filesystému je několik málo hodně kusých položek.

Absence správcovské utility ale vlastně ani nevadí, protože pád disku beztak způsobí okamžitou havárii operačního systému – modul ataraid se opakovaně pokouší o zápis na IDE zařízení, až se z toho systém zblázní a utone ve vodopádu chybových hlášek z podkladového IDE ovladače.

Z narušeného pole pak nejde nabootovat – natožpak za běhu obnovovat. Ovladač ataraid patrně funkce pro obnovu pole vůbec neobsahuje. Teprve po obnově pole pomocí BIOSové utility lze opět nastartovat linux a po fsck namontovat diskové oddíly.

Při výpadku disku tedy patrně nedojde ke ztrátě dat, ale to je tak všechno. Veškerá údržba se musí odehrávat out-of-band mimo Linux a tento funguje pouze v případě, že je pole v bezvadném stavu.

2) semi-closed-source driver od výrobce

Firma Promise dodává pro tento svůj diskový řadič také proprietární ovladač. Zlí jazykové tvrdí, že jde o port nějaké starší verze z Windows, už podle toho, že ovladač emuluje SCSI zařízení, což je jediný rozumně jednoduchý způsob, jak přidat do Windows virtuální diskové zařízení.

Ovladač je tvořen closed-source (binary-only) knihovnou a open-source wrapperem, který se zkompiluje tak, aby odpovídal uživatelem používanému jádru. V původní podobě, tak jak ho lze stáhnout z webu Promise, lze ovladač zkompilovat výhradně jako modul v samostatném adresáři. Tento způsob kompilace je patrně ústupkem vývojářům kernelu, kteří binary-only ovladačům po zásluze spílají. Kromě toho udržování patchů pro modifikaci vanilla stromů neustále nových verzí jádra je možná vnímáno jako obtížnější a konfliktnější než údržba samostatného modulu.

Potažmo, pokud chceme z pole bootovat, budeme muset použít initial ramdisk (initrd).

Modul ovšem lze také ručně monolitizovat – v tomto případě není initrd zapotřebí. Na webu na adrese http://sweb.cz/Frantisek.Rysanek/pdcraid/ lze stáhnout balíček obsahující původní driver, patch a skript, které provedou monolitizaci. Ovladač pak lze nalézt a přidat v “make menuconfig” ve větvi SCSI -> SCSI hardware drivers.

Pokud se rozhodneme pro tento proprietární modul, je třeba vypnout IDE driver pdc202xx (a pochopitelně také pdcraid). Kromě toho je potřeba kernelovým parametrem při bootu vyřadit detekci IDE kanálů pro všechny řadiče vyšší než standardní, které jsou patrně na motherboardu – tj např. v lilo.conf je třeba přidat **`append="ide2=0 ide3=0 ide4=0 ... ide9=0"`**.

Příklad souboru lilo.conf lze nalézt na výše zmíněném webu.

Výsledek je opět značně pochybný. Výpadek disku ve většině případů pošle systém ke dnu. Správcovská utilita pod Linux neexistuje. Ve wrapperu jsou vidět nějaká potenciálně relevantní IOCTL, ale to je všechno. Havarované pole nešlo v BIOSu ani opravit (rebuild), je třeba ho zrušit a vytvořit znovu s kopírováním zdravého disku na vyměněný (ekvivalent obnovy). Z havarovaného pole nejde nabootovat Linux (jádro se natáhne, ale nepřimontuje rootovský svazek), obnova pole podle všeho také ani pod Linuxem nenastartuje automaticky. V newsech byly stížnosti na tuhnutí systému při kopírování delších souborů.

Kromě toho SCSI emulace spotřebovává nezanedbatelný procesorový čas. Celkový výkon proprietárních ovladačů zaostává například za nativním softwarovým RAIDem, který je standardní součástí kernelu.

Takže tudy cesta také nevede.

3) univerzální softwarový RAID pod Linuxem

Funguje stejně dobře jako na kterémkoli jiném IDE řadiči. Viz samostatný dokument o RAIDu pod Linuxem. Stručně: při výpadku disku takřka neodvratně spadne systém, ale z degradovaného pole lze okamžitě nabootovat a pod Linuxem spustit na pozadí obnovu.

Linuxový sw RAID nevidí pole vytvořená BIOSem nebo proprietárními ovladači Promise – tyto ovladače tedy nelze použít k manipulaci s obsahem disku vytvořeným pod Windows.

Zařízení vytvářená linuxovým RAIDem (/dev/md0 atd) nelze rozdělit fdiskem – filesystém se vytváří rovnou na RAIDovém zařízení.

Linux – shrnutí

Jistý Eduard Bloch se v diskusní skupině news://linux.debian.user.german svého času vyjádřil takto: *"Jetzt hasst du gelernt, warum viele Promise-Controller Schrott ist."*

Pořizovat tento řadič kvůli Linuxu jsou vyhozené peníze. Jeho použití v Linuxu lze pochopit v případech, kdy na daném stroji jsou primárním operačním systémem Windows. Případně jako součást Linuxového záchranného CDčka na vyprošťování dat z NTFS a FAT oddílů na polích vytvořených pomocí tohoto řadiče. To je tak všechno, čeho jsou Linuxové ovladače schopny.

Relativně dobře funguje nativní linuxový SW RAID – ale ten funguje stejně dobře na onboard UDMA řadičích. Není třeba vyhazovat hříšné peníze za SW RAID pro Windows, který firma Promise přibaluje ke svému vcelku obyčejnému UDMA řadiči.

Balík "Promise FastTrak TX2000" představuje vlastně poměrně tuctový stand-alone UATA/133 řadič, přibalený k poměrně kvalitní implementaci softwarového RAIDu pro Windows. Proprietární linuxový port tohoto softwaru, uvolněný jako open-source wrapper kolem binary-only knihovny, zdaleka nedosahuje kvalit originálu z prostředí MS Windows.

Firma Promise nedávno uvolnila pod GPL zdrojový kód "ovladačů" pro své řadiče SATA RAID – těžko říci, zda to v budoucnu zlepší podporu i pro starší řadiče Promise.

Současný stav je výsledkem historického vývoje. Firma Promise odstartovala jako výrobce přídavných UDMA řadičů v dobách, kdy se objevily první disky podporující UDMA a mezi lidmi bylo ještě spousta motherboardů, jejichž integrované řadiče nepodporovaly UDMA nebo velké disky. V té době se řadiče Promise balily například k diskům Western Digital a firma Promise byla pod Linuxem velkým pojmem. Když se později objevily UDMA řadiče integrované v south-bridgi na všech motherboardech, přišly hubené časy a firma Promise si našla na trhu jinou skulinu, kterou nyní vydatně přiživuje řemeslně solidním marketingem.

Doporučený postup pro čištění použitých disků

Bližší vysvětlení důvodů apod. viz samostatný obecný dokument o RAIDu.

Je třeba vyčistit MBR+partition table. Dále je vhodné vyčistit také “RAID superblock”, který se nachází na začátku poslední stopy disku. Pozor, nejde nutně o poslední stopu dle CHS, tj. *ofset != C*H*S - S*. Některé moderní disky s LBA přístupem hlásí celkovou kapacitu (počet sektorů) větší než součin CHS. V tom případě zřejmě platí, že *ofset = Kapacita – S*.

V linuxu lze parametry CHS a LBA kapacitu zjistit z /proc/ide/hd<X>/geometry a /proc/ide/hd<X>/capacity. Zjištěné parametry použít na příkazovém řádku programu dd – viz samostatný obecný dokument o RAIDu. Specielně pro disky poznamenané Promise RAIDem autor sepsal perlový skript, který funguje v Linuxu – viz http://sweb.cz/Frantisek.Rysanek/pdcraid/clean_promise_raid_drive.pl