Záchrana a obnova dat

Záchrana dat je tvůrčí práce, která nemá svá pevná pravidla, pouze cíl - vyřešit situaci, kdy uživatel uložil informace na nějaké zařízení či médium a z různých důvodů k nim ztratil přístup.

Krásným námětem pro odborné články jsou zařízení vybavená měřicími body, požadovanými a skutečnými průběhy signálů a postupem analýzy směřující k identifikaci vadné součásti a k její následné výměně. Podobný postup v podstatě aplikují samoopravné nástroje paměťových zařízení a operačních systémů. Vyrobit například povrchy pevných disků absolutně bez chybných míst, tzv. drop outs, není možné ani dnes, ale místo papírku se seznamem nepoužitelných bloků jsou disky vybaveny blokem záložních sektorů a mechanismem, který je automaticky použije místo sektorů poškozených. Zbývá doplnit, že první vadný sektor hlášený operačním systémem znamená, že naznačený relokační mechanismus nedokáže chybu vyřešit typicky protože již má všechny záložní bloky obsazené a že to je správná chvíle pro výměnu disku.

Opravné utility jako chkdsk ve Windows pracují o patro výše a sledují konzistenci všech částí souborového systému. Obvykle je cílem jejich autorů umožnit připojení poškozeného svazku a to i za cenu ztráty některých souborů, které v lepším případě ukládá jako found001 atd. v adresáři vytvořeném pro tento účel. Použití takové utility má i skrytá nebezpečí, z nichž za zmínku stojí fakt, že při každém běhu zapisují na opravované médium. Pokud je chyba způsobena například problémy s adresací bloků na zařízení s velkou kapacitou, dojde k nevratnému poškození dat i značného rozsahu.

Společnou nevýhodou a nedostatkem obou naznačených mechanismů je zcela obecný přístup k datům souborového systému. Obsahem a významem souborů se zabývají pouze aplikace, které s nimi pracují a uživatelé často nemají k dispozici informace, které by jim umožnily rozlišit, které soubory jsou skutečně nezbytné pro obnovu funkce účetnictví a které nikoliv. Tento problém je obzvláště zvýrazněn při návrhu systému zálohování, které má se záchranou dat mnoho společného – cílem je obnovit 100 % funkci po výpadku a to v co nejkratším čase.

Záchranu dat lze rozdělit na 3 základní činnosti:

- záchrana z poškozených médií, která například nelze číst, při čtení dochází k chybám, jsou mechanicky, tepelně nebo elektricky poškozena atd. Vždy je pořízen pracovní binární obraz, určený pro další práci bez rizika poškození původního zařízeni, s nímž lze pracovat mnohem rychleji než s původním, poškozeným médiem. Pouze případy velkých diskových polí jsou zpracovány bez celkového obrazu, z praktických a časových důvodů se pořizují pouze obrazy poškozených disků. Vzácně lze kopii zařízení rovnou použít, nejčastěji ale taková kopie postupuje do fronty strukturálně poškozených médií, protože v okamžiku selhání nemohl operační systém zapsat všechny aktuální změny.
- záchrana z bezchybně pracujících médií, kde je poškozena struktura souborového systému a svazek lze opravit tak, aby s ním mohl pracovat původní operační systém. Velmi časté je to v případě poškození některého životně důležitého sektoru, který například NTFS nebo FAT nemá zálohován. Pak je celý svazek odmítán jako neznámý formát a nelze ho ani opravovat prostředky operačního systému. Chybějící údaje lze odvodit z ostatních částí svazku někdy aplikací zkušenosti s chováním dané verze operačního systému na médiu stejného typu a velikosti. Po doplnění chybějících informací lze svazek obvykle již snadno opravit a zprovoznit. Jak bylo zmíněno výše, není po běhu opravných utilit obvykle cesty zpět, proto je velmi důležité je nezkoušet dokud neexistuje jistota správnosti doplněných údajů nebo spolehlivá kopie.
- záchrana dat z obecného bezchybného média, kde je kladen důraz na obnovu konkrétních dat konkrétního formátu nebo kde důležitá data tvoří jen zlomek celkové velikosti disku nebo kde oprava svazku do funkčního stavu není možná.

  Ve všech těchto případech je nutné vyvinout, odladit a použít software, který:
- dokáže číst médium bez funkcí operačního systému, resp. na úrovni fyzických sektorů-bloků. Tuto část softwaru je třeba stále updatovat podle novinek, s nimiž výrobci přicházejí, což je časově velmi náročné často díky nutnosti použít nejen programové, ale i technické prostředky. Provedení záchrany dat to zdržuje pouze v případě výskytu zařízení, které není zpracováno dopředu, jinak jde v podstatě o princip

ovladače, jenž lze použít opakovaně zcela beze změn.

- dokáže lokalizovat bloky-součásti hledaných souborů, což je relativně snadnější u hlaviček souborů a velmi obtížné u bloků „uprostřed“. U některých souborových systémů lze po lokalizaci hlavičky najít správně seřazený seznam dalších bloků a tím soubor získat v původním stavu, jinde je nutné podle struktury a typických bloků informací postupně hledat blok za blokem. Aby nebylo nutné procházet kvůli každému dalšímu bloku vždy celý disk, používáme při znalosti způsobu práce operačního systému, například metody elevator seeking u starších verzí OS Novell Netware nebo využívání bloků s nižšími pořadovými čísly u OS firmy Microsoft, statistické metody pro určení přibližného místa umístění bloku. Tento předpoklad znamená i určitou váhu navíc při rozhodování, který z nalezených bloků odpovídajících formátem a obsahem je nejaktuálnější.
- oddělit aktuální data od šumu dat stejného formátu která se na disku vyskytují například kvůli tzv. zálohování, kdy jsou soubory z jednoho adresáře opakovaně kopírovány do jiného adresáře, což při způsobu přidělování bloků a přepisování existujících souborů generuje na disku kopii alespoň některých částí při každém takovém kopírování. Nejen kvůli možné záchraně dat, ale i jako ochrana před selháním média nebo operačního systému je více než vhodné tímto způsobem zálohovat alespoň na jiný fyzický disk. Rozpoznání aktuálnosti bloku není nutné například v případě tabulky relační databáze, která pomalu roste, záznamy v ní přibývají a již uložené se nemodifikují. Tak se chová třeba účetnictví a jeho hlavní kniha, ze které by mělo být možné vygenerovat většinu informací pro knihy ostatní. V takovém případě je každá kopie souboru zcela stejná až po blok obsahující poslední uložené záznamy, což je dobře patrné pokud pořídíme mapu CRC32 součtů všech bloků svazku (nikoliv sektorů disku) a vhodným způsobem je setřídíme/kategorizujeme.
- po lokalizaci všech bloků umožní celý soubor přečíst se správným řazením bloků a uložit ho na jiné médium. Abychom dokázali efektivně obnovit i strukturu adresářů a vybrali jen některé soubory, používáme techniku plug-inu do souborového manageru FAR, který dovolí zobrazit v jednom z panelů obnovená data a prostě je kopírovat kamkoliv. Akci může v naší laboratoři provést sám vlastník dat aniž by k tomu potřeboval jakékoliv odborné znalosti.

Použití nejvyspělejších metod je časově a tedy i finančně náročné ale dovoluje udržet míru úspěšnosti nad magickými 90-ti %. Ve zdravotnictví má mnoho lidí štěstí a vyléčí si rýmu za několik dnů lékem za stokorun, někteří štěstí nemají a bojují dlouhou dobu s vážnou nemocí za podstatně vyšších nákladů. Na rozdíl od masové produkce bot je záchrana dat i životů především otázkou mnoha hodin práce špičkových specialistů, kterou nelze nahradit. Možná ovšem tento článek inspiruje některou pojišťovnu k nabídce zajištění proti ztrátě dat a nikoliv jen proti ztrátě disku.

Ing. Václav Šamša
Autor pracuje ve společnosti TDP-Ontrack Data Recovery