Elektronická pošta a Linux

Petr Kolář

Technická univerzita v Liberci

Petr.Kolar@vslib.cz

Příspěvek na seminář LinuxStart 2000
Brno 3.–4. června 2000

Příspěvek poskytuje informace o poštovních protokolech a standardech RFC-822, MIME, SMTP a jeho rozšířeních, o programech pro dopravu pošty zejména o programu sendmail, a o programu procmail. Jsou také zmíněny otázky ochrany proti spamu.

Formát e-mailů

Formát elektronických dopisů definuje standard RFC-822 z roku 1982. Zpráva podle RFC-822 smí obsahovat pouze ASCII znaky (sedmibitové), začíná hlavičkami (povinně musí obsahovat hlavičky `From:`, `Date:` a `To:` nebo `Bcc:`, pořadí hlaviček je libovolné, doporučuje se však, aby bylo takové, jako je v ukázce zprávy níže) a pokračuje tělem, které je od hlaviček odděleno prázdným řádkem:

```
Return-Path: <vvlezly@e-dnes.cz>
Received: from SpoolDir by TYTO (Mercury 1.44); 20 Mar 00 07:34:02 +0100
Received: from bubo.vslib.cz (147.230.16.1) by tyto.vslib.cz
    (Mercury 1.44) with ESMTP; 20 Mar 00 07:33:55 +0100
Received: from hermes.iol.cz (hermes.iol.cz [194.228.2.36])
        by bubo.vslib.cz (8.9.3/8.9.3) with ESMTP id HAA21655
        for <petr.kolar@vslib.cz>; Mon, 20 Mar 2000 07:33:33 +0100
Received: from [194.228.147.3] by hermes.iol.cz (Post.Office MTA
      v3.5.3 release 223 ID# 631-66640U65000L65000S0V35) with SMTP
      for <petr.kolar@vslib.cz>; Mon, 20 Mar 2000 07:33:12 +0100
Date: Mon, 20 Mar 2000 07:33:12 +0100
From: Vladimir Vlezly <vvlezly@e-dnes.cz>
Subject: info 01/00
Sender: vlezlik@mafra.cz
To: petr.kolar@vslib.cz
Cc: jan.klc@spsselib.hiedu.cz, adamec@bubo.vslib.cz (Petr Adamec)
Message-ID: <20000320063312.AXE11865@hermes.iol.cz@[194.228.147.3]>

I vy jste v databazi, kterou nam prodala jedna sikovna firma.
Nakonec vam nezbyde nez pouzit sluzeb firmy CESKOMORAVSKA
POHREBNICKA www.hrbitov-online.cz

S uctou Vladimir Vlezly
```

RFC-822 stanovuje formáty hlaviček, adres, časových údajů apod..

Adresy mají tvar *lokální-část*@*doména*, kde *doména* je plně kvalifikované doménové jméno (v nouzi to může být číselná IP adresa uzavřená v hranatých závorkách). RFC-822 dovoluje i skupinové adresy tvaru *skupina*: *adresa,adresa*;. Takzvané route adresy `<@domena1.top1,@domena2.top2:user@domena3.top3>` se doporučuje nepoužívat.

Hlavička `Return-Path:` obsahuje adresu odesilatele přidanou posledním programem pro dopravu pošty. Ve zprávách o nedoručitelnosti pošty je nastavena na `<>`, což signalizuje, že se zpráva o nedoručitelnosti nedoručitelné zprávy o nedoručitelnosti `:-)` nemá posílat.

Hlavičky `Received:` se čtou odspoda a udávají, jakými počítači zpráva prošla. Pokud hlavička zabírá více řádků, musí pokračovací řádky začínat aspoň jednou mezerou nebo tabelátorem.

Hlavička `Date:` obsahuje datum a čas odeslání zprávy; RFC-822 povoluje pouze poslední dvojčíslí roku (Y2K problém), v RFC-1123 je doporučován zápis roku čtyřčíslím. Odečtením časového posunu od časového údaje dostaneme čas v UTC.

Hlavička `From:` obsahuje adresu autora zprávy, hlavička `Sender:` pak adresu toho, kdo zprávu skutečně poslal. Adresát (nebo adresáti) je uveden v hlavičce `To:`, adresát(i) kopie v hlavičce `Cc:`; zpráva je posílána i na adresy uvedené v hlavičce `Bcc:`, ale tato hlavička bude z dopisu odstraněna (blind carbon copy). Více adres v jedné hlavičce se odděluje čárkami. Adresy mohou obsahovat i komentáře; pokud je část adresy mezi znaky `< >`, vše ostatní je komentář, stejně tak je komentář i obsah kulatých závorek včetně těchto závorek a mezer, které je obklopují.

Adresa, na kterou se má odpovídat, může být uvedena v `Reply-To:`.

Hlavička `Subject:` by měla charakterizovat obsah zprávy, podobný význam mají hlavičky `Comments:` a `Keywords:`. V uvedené zprávě je špatně hlavička `Message-ID:`, která by měla jednoznačně identifikovat zprávu (nesmí obsahovat dva zavináče — toto je specialitka M$).

Při přeposlání dopisu by měly být přidány hlavičky `Resent-From:` (případně i `Resent-Sender:`), `Resent-To:` (`Resent-Cc:`, `Resent-Bcc:`), `Resent-Date:`, případně i `Resent-Reply-To:` a `Resent-Message-Id:`.

Nestandardní hlavičky musí začínat `X-` (například název poštovního programu může být uveden v `X-Mailer:`).

MIME

Standard MIME (Multipurpose Internet Mail Extensions) z roku 1992 rozšiřuje RFC-822 na e-dopisy obsahující jiná data než ASCII text nebo složené z více částí. Použití MIME signalizuje výskyt hlaviček:

```
MIME-Version: 1.0
Content-Type: typ/podtyp
Content-Transfer-Encoding: přenosové-kódování
```

Hlavička `Content-Transfer-Encoding:` udává způsob kódování dat:

`7bit`	řádky ASCII znaků délky max. 998 zakončené CR,LF
`8bit`	jako `7bit`, ale jsou povoleny i znaky s kódy > 127
`binary`	libovolná osmibitová data bez omezení
`quoted-printable`	pro zakódování textových souborů
`base64`	pro zakódování binárních souborů

Hodnoty `7bit`, `8bit` a `binary` udávají, že následují data uvedeného charakteru bez jakéhokoli kódování. Naproti tomu `quoted-printable` udává, že textová data obsahujících i ne-ASCII znaky či příliš dlouhé řádky jsou zakódována do formy ASCII textu s řádky délky nejvýše 76 znaků, `base64` pak udává druh kódování používaný pro binární data.

Hlavička `Content-Type:` udává tak zvaný média typ následujících dat. Tato data mohou být buď složená — typu `multipart` či `message` (tj. tělo dopisu je tvořeno kompletním dopisem včetně dalších hlaviček), nebo jednoduchá — některého z typů `text`, `image`, `audio`, `video` či `application`. Vždy musí být uveden podtyp, a volitelně mohou následovat dvojice `atribut=hodnota` oddělené od předcházejícího textu středníkem — např. český text s diakritikou v kódování ISO-8859-2 by měl být označen

```
Content-Type: text/plain; charset=iso-8859-2
```

Několik významnějších média typů:

`multipart/mixed` — více částí různého charakteru;
`multipart/alternative` — Více částí různého typu, ale stejného obsahu
`multipart/parallel` — Části, které se mají interpretovat současně
`multipart/signed` — Podepsáno
`multipart/encrypted` — Zašifrováno
`message/rfc822` — Kompletní e-mailová zpráva
`message/news` — Newsová zpráva
`message/http` — HTTP odpověď
`text/plain` — Obyčejný text, ne nutně jenom ASCII
`text/html` — HTML
`text/sgml` — SGML
`application/octet-stream` — Blíže neurčená binární data
`application/postscript` — Kupodivu PostScript není mezi texty
`application/zip` — Klasický PKZIP
`application/msword` — Dokument v MS-Wordu (případně i s virem)
`application/pgp-encrypted` — Zašifrováno pomocí PGP
`application/pgp-signature` — Podepsáno pomocí PGP
`application/pgp-keys` — PGP klíče
`video/mpeg` — MPEG video
`video/quicktime` — Quicktime video

Aktuální seznam zaregistrovaných média typů lze získat na FTP serveru `ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/`.

V případě složených částí (`multipart`, `message`) je v hlavičce `Content-Type:` uveden parametr `boundary=hranice` definující řetězec použitý v oddělovačích jednotlivých částí, které mají tvar CR,LF,`--hranice`,CR,LF.

Poštovní programy

Uživatelské poštovní programy (označované jako MUA — Mail User Agents) neprovádějí doručení pošty adresátovi; odesílané zprávy předávají některému z programů pro dopravu pošty (MTA — Message Transfer Agents). Toto oddělení poštovních programů dovoluje, aby si uživatel mohl vybrat, jaký poštovní program bude používat, a aby správce počítače mohl nezávisle zvolit, jaký program je pro daný počítač nejvhodnější, jaký umí spravovat apod.. Tvůrci MUA se mohou zaměřit na uživatelskou přívětivost a nemusí se zabývat otázkami směrování a doručování pošty (které navíc v případě výpadku sítě může trvat déle, než je odesilatel dopisu přihlášen a má spuštěn svůj poštovní program).

MUA umožňují uživateli práci s e-dopisy: příjem (načítání z lokálních souborů nebo z POP či IMAP serverů), filtrování a třídění, čtení, psaní odpovědí a nových dopisů, tisk.

Úkolem MTA je převzít zprávu, rozhodnout o jejím dalším směrování (mechanismus a server), odeslat zprávu, v případě dočasných problémů opakovat pokusy o doručení nebo se pokusit přesměrovat zprávu. Při neodstranitelných problémech ohlásit chybu (návratovým kódem, vrácením e-dopisu s chybovou zprávou apod.).

Další skupinu programů tvoří servery, které zpřístupňují obsah uživatelovy poštovní schránky (POP a IMAP servery). Zvláštní postavení má program fetchmail, který načítá dopisy z POP nebo IMAP serveru a rozesílá je dále. Dále existuje množství pomocných programů pro třídění, konverze a další operace s poštou.

Dosud nejrozšířenějším programem na dopravu pošty v prostředí Unixu je sendmail. Pochází z roku 1982, kdy současné standardy teprve vznikaly, a může být použit jako brána mezi různými poštovními systémy (umožňuje rozsáhlé transformace adres a pomocí externích programů lze dodávat další mechanismy pro dopravu pošty), ale má celou řadu nectností: jedná se o jeden velký program, který: přebírá poštu od jiných poštovních programů, slouží jako SMTP server, SMTP klient, provádí přepisování adres, směrování pošty i konverze těla dopisu. Nebyl navrhován se zvláštním zřetelem k bezpečnosti, takže obsahoval množství bezpečnostním děr, a navíc se chová špatně při velkém zatížení.

Existuje celá řada alternativ sendmailu:

- Postfix `http://www.postfix.org`
- Qmail `http://www.qmail.org`
- Exim `http://www.exim.org` — vychází z smail 3
- Zmailer `http://www.zmailer.org/` — náhrada sendmailu, vytvořen v roce 1988; nižší zatížení počítače, méně bezpečnostních děr
- MMDF `http://www.mmdf.org` — standardní mailer v SCO Unixu
- Smail 3 pro SMTP sítě
- Smail 2.5 pro UUCP sítě
- PP neboli Isode Message Switch — pro RFC-822 a X.400; kdo neviděl, nepochopí
- Delivermail (předchůdce sendmailu; zastaralý)

Jako nejnadějnější se jeví Qmail, Postfix a Exim. Ve všech třech případech se jedná o stále vyvíjené modulární programy, psané s ohledem na bezpečnost a dobrou účinnost i při velkém zatížení.

Obálka a hlavičky

Pro dopravu sítí se dopis doplňuje takzvanou obálkou (envelope), což je adresa odesilatele a jedna nebo více adres příjemců. Počáteční obsah obálky se získá z hlaviček dopisu a během dopravy sítí se může měnit:

- Pokud je dopis určen více uživatelům na různých počítačích, musí se v některém místě sítě cesty jednotlivých kopií dopisu rozejít. Z obálky jednotlivých kopií zprávy budou odstraněny ty adresy, pro které byl dopis poslán jinou cestou. Naopak v hlavičkách zůstanou pro informaci příjemců všechny adresy (kromě `Bcc:`).

- Uživatel může pomocí souboru `.forward` ve svém domovském adresáři předepsat, na jakou adresu má být dopis automaticky forwardován. Aby měl adresát kontrolu, na jakou adresu byl dopis původně zaslán, bude změněna pouze adresa v obálce.

- Při rozesílání dopisu z mailing listu bude adresa odesilatele v obálce nahrazena adresou správce listu. Nedoručitelné dopisy budou vraceny správci listu, který může například adresu, která je delší dobu nedosažitelná, vyškrtnout ze seznamu abonentů listu.

- Pokud není možné elektronický dopis doručit, bude jeho odesilateli e-poštou vrácena zpráva o chybě. V hlavičkách je jako odesilatel uveden například `MAILER-DAEMON@uzel`, aby byl odesilatel původního dopisu informován, kde k chybě došlo. Může se však stát, že i zpráva o nedoručení je nedoručitelná, což by mohlo vést k nekonečnému pingpongu nedoručitelných dopisů. Proto je v chybové zprávě prázdná adresa odesilatele v obálce, což je informace pro MTA, že se nemají pokoušet vracet chybové zprávy.

Při použití protokolu SMTP je obálka tvořena adresami v SMTP příkazech `MAIL FROM:` a `RCPT TO:`.

Protokol SMTP

Pro dopravu pošty se v Internetu používá protokol SMTP (Simple Mail Transfer Protocol) definovaný RFC-821 z roku 1982. Protokol je založen na architektuře klient–server: počítač, který přijímá poštu, funguje jako server (naslouchá na TCP portu 25); počítače, které mu potřebují zaslat poštu, navazují spojení na TCP port 25 jako klienti.

Vlastní dialog podle protokolu SMTP probíhá takto (`C:` je klient, `S:` server):

```
S: 220 bubo.vslib.cz ESMTP Sendmail 8.9.3/8.9.3;
       Wed, 24 May 2000 09:02:21 +0200 (MET DST)
C: HELO ns.felk.cvut.cz
S: 250 Hello ns.felk.cvut.cz, pleased to meet you
C: MAIL From:<csTeX-Mgr@cs.felk.cvut.cz>
S: 250 <csTeX-Mgr@cs.felk.cvut.cz>... Sender ok
C: RCPT To:<MILAN.KERSLAGER@VSLIB.CZ>
S: 250 <MILAN.KERSLAGER@VSLIB.CZ>... Recipient ok
C: RCPT To:<VIT.ROCEK@VSLIB.CZ>
S: 250 <VIT.ROCEK@VSLIB.CZ>... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: (hlavičky a tělo zprávy)
C: .
S: 250 Ok
C: QUIT
S: 221 bubo.vslib.cz closing connection
```

Server po přijetí příkazu `QUIT` uzavře TCP spojení.

Další příkazy: `VRFY` a `EXPN` slouží k verifikaci adresy případně k rozvinutí aliasů. `NOOP` je prázdný příkaz. `RSET` slouží ke zrušení zasílání dopisu (může být použit místo příkazu `RCPT To:` nebo `DATA`). `HELP` vypíše seznam příkazů, které zná server.

Protokol ESMTP — rozšíření SMTP

Obsahuje mnoho rozšíření popsaných v několika RFC. Nové příkazy:

- `EHLO` nahrazuje `HELO`; odpovědí je seznam volitelných nebo rozšířených příkazů nebo rozšíření podporovaných serverem (povinné jsou `HELO`, `MAIL`, `RCPT`, `DATA`, `RSET`, `VRFY`, `NOOP` a `QUIT`, mezi volitelné příkazy patří `SEND`, `SOML`, `SAML`, `EXPN`, `HELP`, `TURN`, mezi rozšířené `ETRN`, `BDAT`)
- `ETRN jméno`; způsobí odeslání obsahu fronty pro zadaný stroj nebo (začíná-li jméno zavináčem) doménu

Některé příkazy (`MAIL FROM:`, `RCPT TO:`) mají další parametry. Pro rozšíření viz `http://www.isi.edu/in-notes/iana/assignments/mail-parameters`.

Jednotlivá rozšíření:

- `PIPELINING` — klient může odesílat další SMTP příkazy bez čekání na odpověď na předchozí příkaz
- `8BITMIME` — možnost příjmu textových osmibitových zpráv; je-li použito, bude mít příkaz `MAIL FROM:` parametr `BODY=8BITMIME`
- `SIZE` — při použití jako odpověď na příkaz `EHLO` udává parametr maximální velikost zprávy, která může být přijata; parametr `SIZE=` příkazu `MAIL FROM:` udává horní odhad velikosti předávané zprávy
- `CHUNKING` — umožňuje přenos binárních zpráv; po příkazu `MAIL FROM:` s parametrem `BODY=BINARYMIME` se místo příkazu `DATA` se použije jeden nebo více příkazů `BDAT velikost [LAST]`, kde velikost udává počet bytů, které budou následovat, a `LAST` signalizuje, že se jedná o poslední kus zprávy; standardní sendmail dosud toto rozšíření neimplementuje
- `DSN` — (delivery status notification) definuje mechanismus, kterým je možné sdělit odesilateli, zda byl jeho dopis doručen; v příkazu `MAIL FROM:` mohou být použity parametry:
  - `RET={FULL,HDRS}` — vrátit celou zprávu nebo jen hlavičky
  - `ENVID=envelope-identifier`

  v příkazu `RCPT TO:` parametry
  - `NOTIFY={NEVER,SUCCESS,FAILURE,DELAY}` — o čem informovat; s výjimkou `NEVER` lze uvést více možností oddělených čárkami
  - `ORCPT=address-type;original-recipient` — původní adresa příjemce; pro Internetovou poštu je address-type RFC822

Ochrana proti spamu

- Zamezení příjmu dopisů, které nesou neklamné příznaky spamu.
- Zamezení příjmu pošty z počítačů, z nichž byl posílán spam v minulosti, a jejichž správce účinně nereagoval.
- Zamezení relayingu (posílání spamu přes další uzel, v tomto případě přes váš).

První dvě opatření chrání lokální uživatele, nejdůležitější je ale třetí — uzly, které nebrání relayingu, se mohou dostat na černou listinu stejně jako spammeři.

Proč spammeři používají relaying?

- Přenášejí zátěž se směrováním a distribucí pošty na jiné
- Obcházejí jednodušší prostředky proti spamu (TCP wrappery, apod.)
- Snáze mohou skrýt svoji identitu (starší sendmaily a některé jiné MTA nepřidávají do hlaviček `Received:` IP adresu odkud dopis převzaly)

Prostředky ochrany proti spamu a relayingu

Objevují se až po roce 1997; u starších verzí MTA nejsou. V sendmailu 8.9.x a 8.10.0 je hlavním prostředkem databáze `access`. Její použití je u sendmailu 8.9.x dosti nepřůhledné, protože slouží k ochraně proti spamu i relayingu a testují se pomocí ní jak jména a IP adresy strojů odkud přichází pošta, tak domény a nadřízené domény adres odesilatele a příjemce v obálce. U sendmailu 8.10.0 lze jednotlivé případy rozlišit předřazením řetězce `Connect:`, `From:` nebo `To:` jednotlivým klíčům. Sendmail s vlastností `access_db` implicitně odkudkoli přijímá poštu pro lokální uživatele, a umožňuje lokálním uživatelům posílat poštu kamkoli; relaying však neprovádí. Databáze se vytvoří ze souboru `/etc/mail/access` jehož řádky mají tvar

```
spsselib.hiedu.cz       RELAY
147.230.16              RELAY
savetrees.com           REJECT
abcdef.cz               OK
```

První řádek umožňuje relaying z a do domény `spsselib.hiedu.cz` i dopisů přicházejících ze strojů z této domény, druhý relaying dopisů přicházejících ze strojů ze sítě 147.230.16, třetí naopak odmítá dopisy se směrovou částí adresy `savetrees.com` i spojení ze strojů z této domény. Poslední řádek umožňuje příjem dopisů z domény `abcdef.cz` i když tato doména není v DNS (nikoli však relaying).

Pro používaní databáze `access` je nutné uvést v souboru `.mc`, ze kterého se generuje `sendmail.cf`, řádek

```
FEATURE(‘access_db’)
```

vygenerovat `sendmail.cf` a restartovat sendmail:

```
m4 soubor.m4 > sendmail.mc
cp -b sendmail.cf /etc
/etc/rc.d/init.d/sendmail restart
```

Po každé změně souboru `/etc/mail/access` je třeba z něj vygenerovat access databázi příkazem

```
makemap hash /etc/mail/access < /etc/mail/access
```

Po změně souboru `sendmail.cf` a dalších souborů, které sendmail načítá v textovém tvaru, je nutný restart sendmailu; po změně databází restart potřeba není.

Doprava pošty do uzlu s komutovaným připojením

Pro dopravu pošty do uzlu, který má pouze komutované (modemové) připojení k Internetu, lze použít různé mechanismy:

- SMTP — Relativně jednoduché; nevýhoda: pošta, kterou podřízený uzel neodebere po dobu jejího udržování v poštovní frontě (obvykle 2–5 dní), se vrací jako nedoručitelná.
- POP nebo IMAP — V poslední době velmi populární, pokud se ale použije jedna schránka pro celý uzel, jsou problémy s rozlišením příjemce.
- UUCP — Obvykle UUCP přes TCP. Ačkoli nemá nevýhody výše uvedených metod, je neprávem opomíjena.

Při použití SMTP je třeba zařídit, aby při navázání spojení s poskytovatelem odeslal jeho poštovní server pro uzel s komutovaným připojením. K tomuto účelu slouží SMTP příkaz `ETRN`, který je možné zadat poštovnímu serveru příkazem

```
( sleep 10; echo EHLO mujstroj.mojedomena.cz; sleep 5; \
  echo ETRN mojedomena.cz; sleep 2; echo QUIT ) | \
  telnet stroj.centrala.cz 25
```

Příkazy `sleep` zabraňují, aby telnet uzavřel spojení předčasně. Lze též použít skript etrn.pl z distribuce sendmailu.

Poštovní schránka pro celou doménu

Pokud se pošta pro celou doménu ukládá u poskytovatele připojení do jedné schránky a dále vybírá pomocí protokolu POP nebo IMAP, je třeba rozlišit, jakým uživatelům v cílovém uzlu mají být jednotlivé zprávy doručeny. Informace v hlavičkách nemusí být dostatečné — pokud se jedná o zprávu z elektronické konference, je obvykle v hlavičkách jako adresa příjemce uvedena adresa konference, a informace z obálky se uložením do poštovní schránky ztratí. Řešením je přidat k dopisu hlavičku (obvykle `X-Envelope-To:`), která bude obsahovat adresy příjemce (nebo příjemců) z obálky. Přidání hlaviček pro více adresátů není triviální. Jedním z řešení je použít v sendmailu virtusertable a pro lokální doručení použít schopností programu procmail.

V tabulce virtusertable se uvede položka

```
@domena.cz        schranka+%1
```

a v domovském adresáři uživatele `schranka` bude soubor `.procmailrc` obsahující

```
DOMAIN=domena.cz
ENV_TO=$1

:0f
* ENV_TO ?? .
| formail -i "X-Envelope-To: "$ENV_TO@$DOMAIN

:0fE
| formail -i "X-Envelope-To: UNKNOWN"
```

Program fetchmail je schopen třídit dopisy podle adresy uvedené v hlavičce `X-Envelope-To:`.

Odesílání pošty z uzlu

Je třeba dbát na to, aby adresa odesílatele byla funkční. Většina uzlů v současnosti odmítá dopisy, jejichž zpáteční adresa je nepoužitelná. Pro přepisování adres z vnitřního tvaru na vnější lze (v sendmailu) použít databázi `genericstable`.

Network News

Network News je další klasická služba pro šíření zpráv. Newsový systém zvaný Usenet přesahuje hranice Internetu — existují nebo v minulosti existovaly brány do sítí Bitnet, Fidonet a UUCP sítí.

Na rozdíl od e-poštovních diskusních konferencí nejsou zprávy doručovány do poštovních schránek jednotlivých uživatelů, ale jsou šířeny pomocí newsových serverů. Uživatel se musí připojit k některému news serveru, aby mohl zprávy číst, případně odesílat. Správci news serverů obvykle omezují, z jakých počítačů a domén se lze k jejich news serveru připojovat a číst zprávy, i z jakých domén lze zprávy odesílat. Je možný také přístup pomocí jména a hesla, ale obvykle se nepoužívá.

Celosvětově existují desetitisíce konferencí (v newsech nazývaných skupina — newsgroup). Skupiny jsou kvůli svému obrovskému počtu členěny hierarchicky — například cz.comp.linux je česká skupiny pro oblast počítačů konkrétně zaměřená na Linux. Celkový tok zpráv ve všech newsových skupinách je stovky MB za týden. Zprávy se po několika dnech mažou, takže pokud uživatel nechce o nic přijít, musí newsy sledovat pravidelně. Existují však i různá propojení mezi newsy a e-poštou, pomocí kterých je možné si zasílání zpráv z určitých skupin objednat.

Formát newsových zpráv je popsaný v RFC-1036 a podobá se formátu e-dopisů. Místo hlaviček `Received:` je hlavička `Path:`, která obsahuje jména serverů, kterými zpráva již prošla, oddělená vykřičníkem; místo adresy příjemce je hlavička `Newsgroups:` obsahující seznam skupin, do kterých byla zpráva zaslána, pro určení, do kterých skupis se má odpovídat (obdoba `Reply-To:`), se používá Followup-To:. I pro newsové zprávy se používá rozšíření MIME.

V Internetu se pro dopravu newsů používá protokol NNTP (RFC-977). Slouží pro načítání zpráv klienty ze serverů (někteří klienti umožňují i čtení zpráv z lokálního disku), odesílání zpráv z klientů na servery, i pro přenos zpráv mezi servery. Vzdáleně se podobá SMTP, příkazy jsou však jiné. Zprávu přijatou od klienta rozesílá news server všem sousedním serverům, ti ji rozesílají dále svým sousedům, kteří nejsou uvedeni v hlavičce `Path:` zprávy, atd.. Každý server si pamatuje `Message-Id:` všech zpráv, které přijal v poslední době, a duplicitní zprávy odmítá.

Odlišným způsobem se šíří zprávy v moderovaných skupinách. První server odešle zprávu protokolem SMTP moderátorovi (aby nemusel mít seznam všech moderátorů, používají se aliasy pro jednotlivé hiararchie). Moderátor schválenou zprávu doplněnou o speciální hlavičku odešle opět do toku newsů, odkud se dále šíří stejně jako zprávy v nemoderovaných skupinách.