PŘIHLÁŠKA VYNÁLEZU

zveřejněná podle § 31 zákona č. 527/1990 Sb.

(21) Číslo dokumentu:
2000 -3776

(13) Druh dokumentu: **A3**

(51) Int. Cl.[7]:
H 04 Q 7/38

(19)
ČESKÁ
REPUBLIKA

ÚŘAD
PRŮMYSLOVÉHO
VLASTNICTVÍ

(22) Přihlášeno: **11.02.2000**

(32) Datum podání prioritní přihlášky: **11.02.1999**

(31) Číslo prioritní přihlášky: **1999/19905884**

(33) Země priority: **DE**

(40) Datum zveřejnění přihlášky vynálezu: **14.03.2001**
(Věstník č. 3/2001)

(86) PCT číslo: **PCT/EP00/01136**

(87) PCT číslo zveřejnění: **WO00/48418**

(71) Přihlašovatel:
DETEMOBIL DEUTSCHE TELEKOM MOBILNET GMBH, Bonn, DE;

(72) Původce:
Hager Rolf, Remagen-Oedingen, DE;
Mathar Rudolf, Aachen, DE;
Hellebrandt Martin, Aachen, DE;
Töx Reinhold, Aachen, DE;

(74) Zástupce:
Všetečka Miloš JUDr., Hálkova 2, Praha 2, 12000;

(54) Název přihlášky vynálezu:

Způsob rozpoznání zákaznického zneužití služeb provozovatele sítě pomocí online analýzy datových souborů, vztažených k zákazníkovi

(57) Anotace:

V předloženém řešení se zkoumají datové soubory provozovatele sítě k chování zákazníků v mobilní radiotelefonní síti pro čestné zákazníky a zákazníky s podvodnými úmysly na základě stochastických modelů. Pomocí jednodimenzionálních filtrů se vyvíjejí způsoby k rozpoznání podvodníků, podporované počítačem. 1. S pomocí analýzy hlavních komponent se nabízí grafický nástroj, se kterým se s pomocí dvojdimenzionálních diagramů a příslušného zbarvení nechají lehce identifikovat data podvodníků. Toto se děje pomocí vizuální inspekce nebo pomocí automatického, vnitřního počítačového testu. 2. Na základě Fisherovy diskriminantní analýzy se navrhuje automatický test, který simultánně z vysoce dimenzionálních datových souborů identifikuje takové, které patří podvodníkům. Oba způsoby se vyznačují při malých nárocích na výpočet malými pravděpodobnostmi chyb 1. a 2. typu. Toto znamená vysokou pravděpodobnost detekování u malého počtu chybových alarmů.

CZ 2000 - 3776 A3

advokát
120 00 Praha 2, Hálkova 2

ZPŮSOB ROZPOZNÁNÍ ZÁKAZNICKÉHO ZNEUŽITÍ SLUŽEB PROVOZOVATELE SÍTĚ POMOCÍ ONLINE ANALÝZY DATOVÝCH SOUBORŮ, VZTAŽENÝCH K ZÁKAZNÍKOVI

Oblast techniky

Vynález se týká způsobu podle úvodní části nároku 1, jakož zařízení pro zpracování dat k provádění způsobu.

Dosavadní stav techniky

Sice jsou již známy způsoby typu, uvedeného na začátku, tyto ale trpí nevýhodou, že pracují pomalu a nepřesně a relativně často vytvářejí nežádoucí chybný alarm.

Vynález má proto za úkol, vytvořit podstatně vyšší pravděpodobnost detekování zneužití vůči existujícím způsobům, přičemž se má dosáhnout menší pravděpodobnost chybného alarmu (falešně podezřelý normální zákazník) vzhledem k existujícím způsobům.

Podstata vynálezu

K řešení uvedeného úkolu slouží technická nauka, reprodukovaná v nároku 1. Další provedení a modifikace myšlenky vynálezu jsou předmětem vedlejších nároků.

V předloženém popisu vynálezu se datové soubory provozovatele mobilní sítě pro chování uživatele v mobilní radiotelefonní síti rozkládají na soubory pro čestné

zákazníky a na soubory pro zákazníky s podvodnými úmysly, logicky a s pomocí výpočetní techniky na základě stochastických modelů. Způsoby k počítačově podporovanému rozpoznání podvodu se vyvíjejí nad jednodimenzionálními filtry, implementovanými do MEGS.

1. S pomocí analýzy hlavních komponent se nabízí grafický nástroj, se kterým se pomocí dvojdimenzionálních diagramů a příslušného zabarvení nechají lehce identifikovat data podvodníků. Toto se děje pomocí vizuální inspekce nebo automatického, vnitřního počítačového testu.

2. Na základě Fisherovy diskriminantní analýzy se navrhuje automatický test, který simultánně identifikuje z vysoce dimenzionálních datových souborů takové, které patří podvodníkům.

Oba způsoby se při nepatrných nárocích na výpočet vyznačují malou pravděpodobností chyby 1. a 2. typu. Toto znamená vysokou pravděpodobnost detekování při nepatrném počtu chybných alarmů.

Způsob se v podstatě sestává z následujících kroků, které se, řízeny programem, provádějí v zařízení na zpracování dat:

1. Zjištění vstupních datových souborů z následujících složek:

1.1 kumulované datové soubory přes pevný časový interval, např. 30-ti denní datové soubory;

1.2 data jednotlivých spojení poslední doby, povolená z hlediska zákonné ochrany dat (t.č. 5 dní):

cílové telefonní číslo, doba trvání hovoru, typ spojení atd.);

1.3 data, specifická pro zákazníka (stáří zákazníka v síti, typ platby atd.).

2. Akumulování vstupních znaků podle tříd (typ cílového telefonního čísla, počty hovorů, typy hovorů atd.): každý datový soubor z akumulovaných dat se znázorňuje vysoce dimenzionálním reálným vektorem.

3. Provádění analýzy hlavních komponent:

3.1 Provedení analýzy hlavních komponent na datových souborech již rozpoznaných podvodníků;

3.2 spektrální rozklad příslušné kovarianční matice;

3.3 určení relevantních hlavních komponent;

3.4 klasifikace hlavních komponent, relevantních pro podvodné chování.

4. Transformace hlavních komponent nedetekovaných datových souborů na základě spektrálního rozkladu kovarianční matice v kroku 3.

5. Znázornění (eventuálně grafické) hlavních komponent datových souborů a (eventuálně vizuální) diskriminace příp. podvodného chování.

6. Odhad a zjištění empirických kvantil hlavních komponent k řízení pravděpodobnosti chyb prvního a druhého typu při automatickém detekování a generování alarmu.

7. Fisherova diskriminantní analýza k určení dělicí nadroviny mezi datovými soubory identifikovaných

podvodníků a normálních zákazníků s grafickým znázorněním.

8. Odhad a zjištění empirických kvantil množství promítnutých dat k řízení pravděpodobností chyb 1. a 2. typu při automatickém detekování a generování alarmů.

Přednosti způsobu podle vynálezu jsou následující:

1. jednoduchá schopnost výpočtu za podmínek reálného času: nutné jsou pouze vektorové součty, násobení a maticová inverze;
2. způsob je nezávislý na architektuře počítače a protokolu;
3. způsob zahrnuje možnost automatického, vnitřního počítačového testu na podvodný úmysl s následným automatickým alarmem;
4. datové soubory, korespondující se zneužitím, se mohou za účelem detekování znázorňovat graficky: optická rozlišitelnost;
5. způsob se učí z minulých datových souborů, aktuální datové soubory se tak mohou lépe diskriminovat;
6. způsob je schopný adaptace na nové podvodné profily.

Definování problému

Cílem tohoto vynálezu je včasné rozpoznání podvodných úmyslů ze strany uživatelů mobilní radiotelefonní sítě. Tím se mají omezit finanční ztráty provozovatele. Přitom se pozorují nejenom případy úmyslného zneužití, nýbrž také takové, ve kterých je třeba předvídat, že účastník nebude

s to, hradit své náklady na telefon. V této práci se dále mezi těmito dvěma rozdílnými aspekty nerozlišuje.

Chování účastníků a stanovení možného podvodného úmyslu se charakterizují s pomocí statistických metod a modelů. Za tímto účelem se používají datové soubory, které se evidují k vyúčtování hovorů ze strany provozovatele sítě. Přitom se jedná o data účastníků, jejichž informace se vztahují na časový interval 80 dní. V těchto takzvaných 80-ti denních datových souborech se nacházejí informace o jednotlivých celkových denních obratech zákazníků, denní obraty při mezinárodních a roamingových spojeních, jakož i příslušnost účastníků k třídě, která zobrazuje jeho věk v síti, tedy dobu jeho příslušnosti k síti. K testovacím účelům byly vyšetřovány 80-ti denní datové soubory zákaznických segmentů 19 a 28, to znamená zákazníků, jejichž mobilní radiotelefonní číslo začíná 19 nebo 28.

K dispozici jsou dále data jednotlivých spojení zákazníků, ze kterých se nechají vyčíst příslušná cílová telefonní čísla, začátek, trvání a náklady hovoru, jakož i další informace, jako například buňka, ze které byl telefonát veden. Osobní data účastníka, jako jeho MSISDN číslo a cílové telefonní číslo, jsou přitom přirozeně z důvodu ochrany dat buď zakódována, nebo zkrácena. Data jednotlivých spojení se vztahují na účastnické segmenty 19, 28, 30 a 31 a vztahují se vždy na časový interval jednoho týdne. Segmenty 30 a 31 byly vyhledány speciálně, protože se jedná o již velmi dlouho existující zákaznické segmenty, které dobře zrcadlí chování účastníků bez podvodných úmyslů. Popsaná data účastníka, která jsou podkladem této práce, se dělí na informace, které patří zákazníkům bez podvodných úmyslů a na data od podvodníků, provozovatelem sítě již

rozpoznaných a zablokovaných.

Za účelem analýzy chování účastníků s podvodným úmyslem nebo bez něho se v první části zprávy vyvíjejí statistické modely, s jejichž pomocí se mohou vystihnout různé aspekty chování zákazníka, důležité pro rozpoznání podvodu.

Protože se chování zákazníků s podvodným úmyslem při hovoru znázorňuje jako příliš nehomogenní pro společné zpracování, klasifikují se takoví zákazníci podle znaků jejich hovoru. Pro každou takovou třídu se udávají možnosti k identifikaci. Třídění se přitom uskutečňuje podle chování, kterým se vyznačuje určitá skupina účastníků s podvodným úmyslem a kterým se signifikantně odlišuje od individuálního chování ostatních účastníků. Přitom se používají metody multivariantní statistiky a diskriminantní analýzy. Toto umožňuje, že se chování zákazníků s podvodnými úmysly přesně popisuje jenom s pomocí dvou znaků a podrobuje se testům, které spočívají na těchto znacích.

V poslední části výzkumů se hodnotí určité filtry ze systému rozpoznávání zneužití MEGS vzhledem k jejich kvalitě, to znamená jejich pravděpodobnosti chyby k rozpoznání podvodníků. Údaje se přitom vztahují na empirická data daná k dispozici, která se shromažďovala systémem MEGS.

Tvorba modelu

K vyšetřování chování účastníků s podvodným úmyslem nebo bez něho se určují charakteristické veličiny rozdělení denního obratu. Následovně se zkoumají příslušnosti k třídám

v základních zákaznických segmentech 19 a 28 na jejich využitelnost. Dále se určují empirické rozdělovací funkce a hustoty čítání, potřebné k pozdější analýze filtrů MEGS.

Rozdělení denních obratů

Denní obraty v obou pozorovaných skupinách účastníků se považují za realizace nekorelovaných náhodných proměnných X_e pro normální zákazníky a X_b pro podvodníky. Důležité charakteristické znaky obou rozdělení jsou dány empirickou střední hodnotou a empirickým rozptylem pozorování. Seznam zjištěných veličin se nachází v následující tabulce.

	normální zákazníci	podvodníci
očekávaná hodnota	205.88	11738.1
rozptyl	378174	8.95e+08

Pro segmenty 19 a 28 existují 80-ti denní datové soubory, ze kterých se ještě jednou počítají empirické střední hodnoty denních celkových obratů pro oba segmenty. Přitom se narozdíl od výše uvedeného výpočtu hodnot z dat jednotlivých spojení také zohledňuje, zda účastník vůbec telefonuje ve stejný den. V následující tabulce jsou uvedeny příslušné hodnoty a počty pozorování.

	segment 19...	Segment 28...
očekávaná hodnota	0.22539	1.3335
počet pozorování	512703	5876000

Nápadný rozdíl v odhadech parametrů mezi segmenty 19 a 28 se nechá vysvětlit tím, že se u segmentu 19 jedná o zcela nový zákaznický segment, ve kterém existuje první použití

teprve k časovému okamžiku 11. července 1998. Dále se nechá poznamenat, že mnoho nových účastníků tohoto segmentu nově přibylo teprve po 11. červenci. Tak dostaneme empirickou střední hodnotu prvního použití v jednotce MEGS v segmentu 19 takto:

empirická střední hodnota X (první použití)	1269.33
počet pozorování (=účastník v segmentu 19)	18989

Tím se nechá vysvětlit rozdíl segmentů 19 a 28 ohledně jejich empirických středních hodnot, protože mnoho uživatelů segmentu 19 přichází do úvahy jako uživatelé teprve k relativně dlouhému časovému intervalu, tzn. ve dnech před jejich vstupem do sítě je jejich celkový denní obrat od 11. července roven 0, což má samozřejmě rychle vliv na empirickou střední hodnotu.

Příslušnost k třídám

K dalšímu vyšetření chování normálních zákazníků se nyní pozoruje příslušnost účastníka k třídám. Podle stáří zákazníka jsou účastníkovi provozovatelem sítě přiřazeny třídy A, B, C, D a E, které tvoří dobrý základ pro posuzování platební morálky účastníků.

Na základě nedostatečné databáze se tento znak k charakterizování chování účastníků ale dosud nenechal přesně vyšetřit, protože segment 19 obsahuje velmi mladé zákazníky (ve smyslu stáří v síti) a v důsledku toho jsou téměř všichni zákazníci zařazeni do třídy A. Podobně pro segment 28. Zde je zařazen převažující podíl zákazníků již do tříd C a D, takže s tímto datovým souborem není možné učinit žádné přesné výpovědi. Další segmenty nejsou při

dosavadních vyšetřováních k dispozici, čímž se přesné pozorování rozdělení příslušnosti k třídám nemůže uskutečnit bez dodatečných dat. Četnosti příslušností k třídám v segmentech 19 a 28 jsou uvedeny v následující tabulce a mají objasnit výše uvedené vývody.

třída	segment 19...	segment 28...
neklasifikováno	909	77
třída A	18078	6234
třída B	1	6933
třída C	1	13419
třída D	0	46784
třída E	0	3

Charakterizování chování účastníka na základě dat jednotlivého spojení

Po vyšetření 80-ti denních dat se existující data o spojení jednotlivých hovorů zahrnují do pozorování, protože se v těchto datových souborech nechají najít dodatečné informace o chování zákazníka. Zejména se zpracovávají marginální rozložení, potřebná k analýze filtrů MEGS.

K popisu chování účastníka jsou zde především data o zvoleném cílovém telefonním čísle zákazníka, doba trvání hovoru uskutečněného volání a počet volání během dne, jakož i rozdělení obratů zvláštního významu, protože ukazují velmi dobře rozdílné aspekty chování zákazníka. K těmto třem znakům byly zhotoveny stochastické modely a rozdělovací předpoklady.

První vyšetření v tomto rámci platí pro rozdělení doby

trvání hovoru jednotlivých spojení. Na doby trvání hovoru se nahlíží jako na realizace kontinuální náhodné proměnné *D*, ačkoliv v datových souborech existují jako diskrétní hodnoty. Grafické znázornění relativních četností dob trvání hovorů je znázorněno na obrázcích 12 a 13.

Jako další se uskutečňuje sestavení modelu k rozdělení cílových telefonních čísel účastníků. Cíl, který zákazník zvolí, se opět považuje za náhodný. Popisuje se pomocí diskrétní proměnné Z, jejíž profil se dále bude vysvětlovat podrobněji.

Oblast cílových telefonních čísel se na základě velkého počtu možností dělí do různých kategorií, které potom slouží jako body profilu Z. Toto rozdělení je s několika vysvětlivkami shrnuto v tabulce 1.

Předvolby	Popis
0177/0171/0172	předvolby německých poskytovatelů služeb mobilní radiotelefonie
0130/0180/0190	německá čísla předvoleb se speciálními tarify
2…/3883	čísla služeb (service) provozovatele sítě (např. T-Box)
ostatní národní předvolby	všechny ještě nezachycené německé předvolby
roaming	spojení s ostatními mobilními radiotelefonními sítěmi (bez mezinárodního roamingu, žádné oznámení cíle)
MTC	MTC-spojení, žádné oznámení cíle (také call

	forward, international, roaming)
faxové spojení	national, international, roaming
mezinárodní předvolby	celkový počet všech mezinárodních spojení a rozdělení podle jednotlivých států (také mezinárodní roaming)

Tabulka 1: kategorie cílových telefonních čísel

Přitom je třeba dbát na to, že se z dat jednotlivých spojení požadují nejenom informace o cílových telefonních číslech, nýbrž dodatečně také ještě takzvaný „Calltype", který udává, zda se jedná o národní, mezinárodní nebo roamingový hovor a který vykazuje mobile terminated calls (MTC).

Volba bodů profilu se uskutečňuje zejména podle hlediska nabíhajících nákladů na jeden hovor. Toto se týká zejména zvláštních čísel 0130 a 0180 a servisních čísel 2... a 3..., jejichž tarify jsou ve spolkové republice jednotné. Stejný argument platí také pro různé poskytovatele mobilní radiotelefonie v Německu, kteří se berou jako jedna kategorie. Tarify rozdílných spojení 0190 se sice liší mezi sebou, ale všechna spojení tohoto typu se shrnují v jedné skupině, aby se počet bodů profilu vzhledem k dalším pozorováním příliš nezvětšil. Rozdělení na roamingové, mezinárodní hovory a MTC se sice nevyznačují jednotnými náklady, odlišují se ale na základě své tarifní struktury od dosud uvedených kategorií. Aby se nechalo dosáhnout všech cílů, pozorují se dodatečně faxová volání, protože se zde nejedná o hovory, ale o přenos dat. Kategorie ostatních národních cílových telefonních čísel nemá žádnou jednotnou strukturu nákladů. Přesto se tato cílová telefonní čísla musí také zachytit, abychom obdrželi úplné vyhodnocení. Na

základě rozmanitosti národních čísel předvoleb a jejich tarifních struktur se zde nemůže dále rozdělovat do skupin. Profil T_Z náhodné proměnné se tak nechá znázornit jako

T_Z = {'0177','0171','0172','0130','0180','0190',
'Service','national','roaming','MTC','Fax','international',}
=:{t_1, t_2, …, t_{12}}

S pomocí dat jednotlivých hovorů ze segmentů 19, 28, 30 a 31 se nyní určuje empirické rozdělení náhodných proměnných $\hat{Z}$. Výčet vypočítaných relativních četností se nachází v obrázcích 4-11. Aby se nalezl ještě přesnější obraz rozdělení cílových telefonních čísel, dělí se bod profilu t_{12} = 'international' ještě jednou podle různých mezinárodních předvoleb. Dostaneme tedy rozšířený model, který se popisuje pomocí náhodné proměnné Z s profilem

$$T_Z = \{t_1,\ldots, t_{11}, i_1, i_2,\ldots, i_{225}\}$$

t_i, i=1,…, 11 jsou přitom definovány přesně jako nahoře, a body profilu i_k, k=1,…, 225 jsou pro velikost podle 225 různých mezinárodních čísel předvoleb, uspořádaných sestupně.

Také zde je opět empirické rozdělení $\hat{Z}$ obsaženo v tabulkách 12-18.

Poté, co bylo popsáno rozdělení cílových telefonních čísel s pomocí právě popsaného modelu a náhodné proměnné Z, nechají se spolu s modelem rozdělení doby hovoru udat některá marginální rozložení, která vyjadřují pravděpodobnosti tvaru

$$| P(D = x | Z = t_i),\ i=1, \ldots, 12$$

a jejich empirické očekávané hodnoty a rozptyly jsou shrnuty v tabulkách 12-18. Tyto informace se budou používat později, aby se analyzovaly filtry MEGS.

Z diagramů relativních četností počtů volání se nechá odečíst, že N postačuje diskrétnímu rozdělení, jehož hustota čítání má přibližně následující tvar:

$$a(k + b)^c, \text{ pro } k \in N$$

Parametry a, b a c se nechají vypočítat odhadem z existujících dat. Na základě tvaru hustoty čítání ostatně není nutně dána jejich čítatelnost. Obrázek 25 ukazuje relativní četnosti počtů volání u normálních zákazníků a aproximaci empirických hodnot pomocí funkce

$$t(x)=113,5*(4.34+x)^{-3.42}.$$

Parametry t(x) se pro toto zobrazení počítaly numericky. Na základě menšího počtu pozorování ukazují grafy počty volání u známých podvodníků velký rozptyl než u zákazníků bez podvodného úmyslu. Příslušné zobrazení se nachází v obrázku 21.

Jak bylo již zmíněno u rozdělení doby trvání hovoru, jsou také u rozdělení počtu volání na den a u rozdělení obratu znázorněna příslušná společná rozdělení pomocí empirické střední hodnoty a empirického rozptylu na obrázcích 12 až 18.

Stochastická závislost cílových telefonních čísel

Aby se mohly analyzovat filtry systému MEGS, použité provozovatelem sítě, jsou nutné dodatečné informace k rozdělení cílových telefonních čísel.

Tak jsou některé filtry dimenzovány na to, aby zachytily ty účastníky, kteří v jednom dni vícekrát telefonují do určité světové zóny. Empirické pravděpodobnosti převážného telefonování do jedné ze světových zón se u výše popsaného modelu mohou odečítat z obrázků. Pokud by jevy, že hovor spadá do jedné z popsaných kategorií, byly stochasticky nezávislé, mohla by se pravděpodobnost pro to, že se v jednom dni k-krát telefonuje do určité cílové třídy, vyjádřit k-násobným součinem empirických pravděpodobností příslušných tříd u rozdělení Z popř. $\hat{Z}$.

Jak jsme se již domnívali, kategorie cílových telefonních čísel ale nejsou stochasticky nezávislé. To se ukazuje, tím, že párově stochastickou nezávislost dvou tříd vyvracíme s pomocí testu, založeném na kontingenčních tabulkách.

V dalším se mají pozorovat náhodné proměnné X a Y s nominálním vyjádřením, přičemž X udává kategorii posledního zavolání a Y kategorii příštího hovoru. Mají tedy oba z profilů $T=T_2\backslash\{'MTC'\}$. Pod nominálním vyjádřením se přitom rozumí hodnoty, které nepodléhají žádnému marginálnímu rozložení a nejsou srovnatelné, jako třeba body profilu '0171', '180', '0190'.

Ke znázornění nominálních náhodných proměnných a jejich

vyšetřování se používá koncept kontingenčních tabulek. K tomu mají X I a Y J nominální vyjádření. Existuje tedy M=I*J možných kombinací k popsání vztahu mezi X a Y. Pozorování (X,Y) páru znaků mají pravděpodobnostní rozdělení, které se znázorňuje v tabulce s I řádky a J sloupci. Buňky tabulky reprezentují I*J možných výsledků. Jejich pravděpodobnosti se označují p_{ij}, přičemž p_{ij} udává pravděpodobnost, že pozorování spadá do buňky (i, j). Pokud jsou v buňkách četnosti páru znaků, hovoří se o (I x J) kontingenční tabulce.

Pravděpodobnostní rozdělení p_{ij} je společným rozdělením X a Y. Marginální rozložení, které dostáváme jako řádkové popř. sloupcové součty přes p_{ij}, se označují pomocí

$$p_{i.} = \sum_j p_{ij} \quad \text{a} \quad p_{.j} = \sum_i p_{ij}$$

Obecně se přitom musí ještě splnit následující vedlejší podmínka:

$$\sum_i p_{i.} = \sum_j p_{.j} = \sum_i \sum_j p_{ij} = 1$$

Aby se mohla testovat nezávislost, kontroluje se hypotéza

$$H_0 : p_{ij} = p_{i.} * p_{.j} \quad \text{popř.} \quad m_{ij} = \frac{m_{i.} * m_{.j}}{n}$$

přičemž n je celkový počet všech pozorování a m_{ij}, $m_{i.}$, $m_{.j}$ udávají příslušné očekávané četnosti. m_{ij} se může vypočítat z pozorovaných četností pomocí odhadu Maximum-Likelihood

$$\hat{m}_{ij} = \frac{n_{i.} * n_{.j}}{n}$$

Hypotéza H_0 se odmítá k úrovni α, když hodnota statistiky testu

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

je větší než $(1-\alpha)$ kvantilu příslušného χ^2 rozdělení s $(I-1)*(J-1)$ stupňů volnosti.

S pomocí tohoto testu se může pro každou běžnou α úroveň odmítnout hypotéza párově stochasticky nezávislých kategorií cílových telefonních čísel při zavolání v jednom dni.

Plánované vyšetření dat jednotlivých spojení

U vyšetřování chování účastníků s a nebo bez podvodného úmyslu je zvláště zajímavá, jak již bylo zmíněno výše, příslušnost zákazníka k třídě, která dokumentuje jeho platební morálku. Toto rozdělení, provedené provozovatelem sítě, bohužel není k dispozici jako informace v datech jednotlivých spojení, takže pro segmenty 30 a 31 se ohledně toho nemohla provést žádná vyšetření. Rozdělení do tříd jsou k dispozici jenom v 80-ti denních datových souborech, tedy pro segmenty 19 a 28. Segment 19 se ovšem nehodí, z důvodů uvedených na úvod, pro vyšetřování tohoto znaku chování. Použití těchto informací je možné teprve tehdy, když je k dispozici dostatečně mnoho dat. Proto se k pozorování tohoto znaku chování vyžaduje příslušnost ke třídě spolu s daty jednotlivého spojení, abychom dostali jisté výsledky.

Dále je plánováno vyšetřování buněk, ze kterých vedou zákazníci své telefonické hovory. Má se zjistit, zda existují buňky, tedy geografické oblasti, se zvláště vysokými počty podvodů. Také k tomu ovšem není existující databáze poznaných podvodníků postačující, abychom získali jisté výsledky.

Zákazníci s podvodným úmyslem

Nyní se má chování podvodníků analyzovat a matematicky popisovat. Informační bázi pro následující pozorování tvoří data již poznaných podvodníků z různých segmentů, daná k dispozici provozovatelem sítě.

Denní obraty 80-ti denních dat

Z předložených dat podvodníků k 80-ti denním datovým souborům není možné, dělat výpověď o rozdělení pro denní obrat u účastníků s podvodným úmyslem, protože k informacím zákazníků 18989 popř. 73450 segmentu 19 popř. 28 existují celkem jenom data o 8 popř. 4 podvodnících příslušných skupin účastníků. Mimoto byli 3 z těchto 12 zákazníků pomocí doplňkových informací, které nejsou k dispozici, rozpoznání jako podvodníci, protože jejich celkové obraty činí v pozorovaných 80 dnech jenom 1 až 6 DM, čímž pro odhad rozdělení přichází do úvahy jenom 8 pozorování.

Chování podvodníků na základě dat jednotlivých spojení

K charakterizování chování podvodníků existují data jednotlivých spojení 57 podvodníků, již rozpoznaných provozovatelem sítě. Bohužel se jedná přitom jenom o

informace z 3708 jednotlivých hovorů, což je oproti více než miliónu datových souborů k normálnímu chování zákazníků velmi malé číslo. Na základě malého počtu pozorování nebude stále možné, pro všechny aspekty chování podvodníků verifikovat rozdělení pravděpodobnosti. Odhad prvního a druhého momentu ale zůstává možný a účelný.

Klasifikace rozpoznaných podvodníků

Na základě existujících informací o podvodnících, již rozpoznaných provozovatelem sítě, se v dalším provádí rozdělení účastníků s podvodnými úmysly do tříd. Tím se má od sebe oddělit chování určitých skupin podvodníků, aby se takto mohly uvést ostré testy k jejich rozpoznání.

Jako motivace k tomu může sloužit srovnání celkového denního obratu a denního obratu na číslech 190 z obrázku 26. Je možno zřetelně rozeznat, že množství bodů se rozpadá do dvou tříd, totiž za prvé na body podél hlavní diagonály pro zákazníky, jejichž celkový obrat se skládá z hovorů na číslech 190. Za druhé odpovídají body podél osy x datovým souborům s jenom nepatrným obratem na číslech 190.

Ke klasifikaci zákazníků s podvodným úmyslem provádíme analýzu hlavních komponent odhadnuté kovarianční matice R_b. Podkladem je datový soubor na den a účastníka, sestávající ze 30 znaků. Pozorují se denní obraty, počet hovorů na den a denní doby trvání hovorů. Každý z těchto tří hlavních znaků se dělí na deset již jmenovaných kategorií cílových telefonních čísel. Uvnitř datového souboru se vždy používají v pořadí čísel 171-, 172-, 177-, 180-, 190-, mezinárodní hovor, ostatní národní spojení, roaming, služby (service) a MTC volání pro hlavní znaky. K tomu označuje $yi \in R^{30}$, i=1,

..., n, vektor se znaky denního datového souboru zákazníka s podvodným úmyslem a n jejich počet. Dále budiž

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

odhadnutá očekávaná hodnota. Potom použijeme odhadu Maximum-Likelihood kovarianční matice

$$R_b = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(y_i - \bar{y}\right)^T$$

Nyní představíme matici R_b jako

$$R_b = T\Lambda T'$$

s ortogonální maticí T a diagonální maticí Λ, která jako diagonální zápisy obsahuje ty o velikosti podle vzestupně uspořádaných vlastních hodnot

$$\sigma_1^2 \geq \ldots \geq \sigma_n^2$$

Přitom dostaneme

$\sigma_1^2 \ldots, \sigma_{30}^2$ = (8e + 08, 2e + 08, 7e + 06, 3e + 06, 2e + 06,
1e + 06, 1e + 05, 6e + 04, 4e + 04, 1e + 04,
1e + 04, 4406, 2434, 480, 308, 147, 115, 115, 35.3,
19.4, 7.3, 1.8, 1.1, 0.6, 0.3, 0.1, 0.02, 0.01,
0.002).

Podíl celkové variability, popsaný pomocí prvních obou hlavních komponent, leží očividně při

$$\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \ldots + \sigma_{30}^2} = 98.6\%.$$

To znamená, že transformované datové soubory

$$\tilde{y}_i = (\tilde{y}_{i,1}, \ldots, \tilde{y}_{i,30}) = T'(y_i - \bar{y}), \; i = 1, \ldots, n$$

leží až na zanedbatelně malou chybu v dvojdimenzionálním podprostoru, který je napnut pomocí prvních dvou jednotkových vektorů. Dále má j-tá hlavní komponenta $\tilde{y}_{i,j}$ očekávanou hodnotu 0, rozptyl σ_i^2 a různé hlavní komponenty jsou nekorelované.

Vlastní vektory t_i matice R_b jsou sloupci ortogonální transformační matice T, to znamená

$$T = (t_1, \ldots, t_{30}).$$

Kvůli, v absolutní hodnotě největším, záznamům prvního a druhého vektoru se nechají diskriminovat podvodníci, tím že se na kategorie, patřící k těmto komponentám, hledí jako na klasifikující znaky pro podvodné chování. Podle vlastních vektorů, patřících k σ_1^2 a σ_2^2, dostaneme rozdělení účastníků s podvodným úmyslem na takové zákazníky, jejichž zneužití se vyznačuje čísly 0190 nebo zahraničními hovory. Na obrázku 27 je skupina podvodníků 0190 znázorněna první hlavní komponentou, která probíhá ve směru osy x. Ostatní podvodníci, kteří jsou nápadní zahraničními hovory, se nechají rozpoznat pomocí druhé hlavní komponenty ve směru osy y.

Testy k rozpoznání podvodníků

Po diskriminování podvodníků, rozpoznaných již provozovatelem sítě pomocí analýzy hlavních komponent, se nyní udávají statistické způsoby testování k rozpoznání podvodníků. K tomu se nahlíží na pozorování chování účastníků bez podvodného úmyslu jako na náhodné a označuje se náhodnými proměnnými $X \sim (\mu, R_e)$, jejichž rozložení má očekávanou hodnotu μ a rozptyl R_e. Matice R_e se dále rozkládá na

$$R_e = S\Theta S',$$

přičemž $S = (s_1 \ldots, s_{30})$ je ortogonální matice.

$$\Theta = \mathrm{diag}\left(\vartheta_1^2, \ldots \vartheta_{30}^2\right) \quad \text{s} \quad \vartheta_1^2 \geq \ldots \geq \vartheta_{30}^2$$

označuje diagonální matici uspořádaných vlastních hodnot R_e, jejichž odhady jsou dány pomocí

$\vartheta_1^2, \ldots, \vartheta_{30}^2$ = (2e+05, 7e+04, 7e+04, 6e+04, 2e+04,
2e+04,1e+04, 1e+04, 7226, 6757,
4732, 2922, 1099, 983, 646, 440, 322, 293, 162, 4.4
3.9, 1.7, 0.8, 0.4, 0.4, 0.3, 0.2, 0.16, 0.01,
0.002).

S pomocí ortogonální matice T z transformace hlavních komponent z R_b tvoříme

$$\tilde{X} = T'(X - \mu),$$

z čehož s linearitou očekávané hodnoty vyplývá $E[\tilde{X}] = 0$ a $E[\tilde{X}\tilde{X}'] = T' R_e T$. Při použití vlastních vektorů t_1, t_2 k oběma

větším vlastním hodnotám kovarianční matice R_b podvodníků z předchozího úseku vyplývá

$$E\left[t_1'\bar{X}\right] = 0 \quad \text{a} \quad \mathrm{Var}\left(t_1'\bar{X}\right) = \sum_{i=1}^{30} t_{1,i}^2 \vartheta_i^2$$

popřípadě

$$E\left[t_2'\bar{X}\right] = 0 \quad \text{a} \quad \mathrm{Var}\left(t_2'\bar{X}\right) = \sum_{i=1}^{30} t_{2,i}^2 \vartheta_i^2$$

$t_{1,i}$ a $t_{2,i}$ označují přitom i-tou komponentu vektoru t_1 popř. t_2. Zejména platí $\mathrm{Var}\left(t_1'\tilde{X}\right) =\leq \vartheta_1^2 \forall i = 1,\ldots,30$. Dále je střední hodnota datových souborů Y_i při výše uvedené transformaci

$$\frac{1}{m} = \sum_{i=1}^{m} T'\left(y_i - \mu\right) = T'\left(\bar{Y} - \mu\right)$$

a rozptyly souhlasí s hodnotami ϑ_1^2, ..., ϑ_{30}^2 z Λ.

Protože hodnoty μ a R_e nejsou známy, odhadují se z existujících n pozorování X_i s pomocí odhadu Maximum-Likelihood

$$\mu = \bar{X}, \text{ a } R_e = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)$$

Použitelnost rozkladu odhadnuté kovarianční matice na transformaci hlavních komponent je zajištěna.

Výsledky vyšetřování podvodníků z posledního odstavce jsou nyní shrnuty spolu s právě provedenou transformací

hlavních komponent dat normálních zákazníků na obrázku 21. V tomto obrázku se nechá dobře rozpoznat zřetelně menší rozptyl hlavních komponent u normálních zákazníků $\mathrm{Var}(t_1'\tilde{X}) \leq \vartheta_1^2 = 2 * 10^5$ na rozdíl od $\vartheta_2^2 \geq 2 * 10^9$, který spolu s posuvem očekávaných hodnot pro zákazníky s podvodným úmyslem ve směru větších hlavních komponent umožňuje použití testu na útlum.

Za účelem uvedení testu k rozpoznání podvodu se vždy pozorují obě hlavní komponenty odděleně, to znamená že se vyvíjí speciální test pro 190 a pro zahraniční podvodníky. Nejdříve se odhaduje empirický (1-α) kvantil $Q_{1-\alpha}^i$, i=1, 2 pro obě hlavní komponenty. Používá se k tomu n stanovených pozorování

$$X_{1:n}^i, X_{2:n}^i, \ldots, X_{n:n}^i$$

Následovně se určuje to číslo k, pro které platí

$$k = \begin{cases} n(1-\alpha) & \text{pokud } n(1-\alpha) \text{ je celočíselné} \\ n[(1-\alpha)+1] & \text{pro ostatní} \end{cases}$$

Potom je kvantil $Q_{1-\alpha}^i = X_{k:n}^i$. Pro speciálně existující data vyplývají pro $\alpha = 0.005$ kvantily $Q_{1-\alpha}^1 = -0.0035$ a, $Q_{1-\alpha}^2 = 14.1089$.

S pomocí empirických kvantilů se může nyní uspořádat test na podvodný úmysl pomocí transformace hlavních komponent dat účastníka jednoho dne. Pokud leží hodnota transformace dat zákazníka nad jedním z obou vypočítaných kvantilů $Q_{1-\alpha}^i$, předpokládá se, že se jedná o podvodníka.

Parametr α přitom udává pravděpodobnost chyby pro to, že zákazník bez podvodného úmyslu se nesprávně rozpozná jako podvodník. Tento omyl je přirozeně třeba udržovat co nejmenší. K použití testu není nutné, pokaždé znovu počítat kvantily $Q^i_{1-\alpha}$, nýbrž stačí periodický nový výpočet těchto hodnot. Vedle tohoto testu ještě existuje možnost, najít podle grafického znázornění transformace hlavních komponent ty účastníky, jejichž hodnoty se nenacházejí na vypočítaných transformačních osách a kteří tak jsou nápadní svým chováním, protože podvádějí kombinací volání 0190 a zahraničních hovorů. Příslušná data zákazníků se mohou potom manuálně překontrolovat ohledně podvodného úmyslu, protože právě popsaným testem nejsou odfiltrováni jako podvodníci (příklady k tomu je možno rozpoznat na obrázku 21).

Dosud popisovaný postup transformace hlavních komponent a následného testu podvodníků se nechá použít nejenom na denní datové soubory účastníků, nýbrž data, která jsou základem, se mohou sbírat po libovolný časový interval a vyhodnotit. Tím dostáváme možnost, automaticky zkoumat chování účastníků přes různé časové intervaly.

Nápadnost účastníků byla dosud měřena jenom podle prvních dvou dominujících hlavních komponent, to znamená 190 a zahraničních spojení, protože již rozpoznaní podvodníci v datovém materiálu, který je k dispozici, se vyznačují jenom těmito oběma aspekty v jejich chování. Účelné ale je rozpoznání každého typu abnormality v chování vzhledem k účastníkům bez podvodných úmyslů. Z tohoto důvodu se používá test na vícedimenzionální útlum. Toto spočívá na Mahalanobis-distanci

$$(X_i - \overline{X}) \cdot R_e^{-1} (X_i - \overline{X}),$$

která měří odchylku chování od chování normálních zákazníků. Explicitní údaj testu se nemůže uskutečnit, protože jsou k dispozici jenom data o podvodnících obou již jmenovaných tříd. Jako alternativa k odchylce chování účastníka od chování normálních zákazníků se může přirozeně také měřit odchylka od podvodného chování použitím R_b^{-1}. Pomocí těchto distancí jsme v situaci, že můžeme lépe přehlédnout chování účastníků a testovat odchylky. Na základě omezeného počtu odlišitelných typů podvodníků v datech, která jsou k dispozici, ale není možné další pozorování vícedimenzionálních testů na útlum k současnému časovému okamžiku.

Jako další postup testu k rozpoznávání podvodníků se může používat Fisherova diskriminantní analýza, která je založena na oddělení podvodníků od normálních zákazníků pomocí roviny. Hledá se lineární funkce $a'x$, která maximalizuje poměr čtverců vzdáleností mezi oběma skupinami účastníků a tyto tak odděluje od sebe. Vektor a je přitom vlastní vektor k největší vlastní hodnotě matice $W^{-1}B$, která se vypočítává z

$$W = n_1 R_e + n_2 R_b$$

a

$$B = \left(\frac{n_1 n_2}{n}\right) dd' \text{ přičemž } d = \mu - \bar{Y}.$$

Přitom n_1 dává počet dat normálních zákazníků a n_2 počet dat podvodníků. Příslušný vlastní vektor a vyplývá jako $a = W^{-1}d$. Navrhuje se statistika testu

$$d' W^{-1}\{X - \frac{1}{2}(\mu + \bar{Y})\}$$

která používá polohu bodu ke střednímu bodu spojovací dráhy mezi μ a $\bar{Y}$. Pro použití k definování problému rozpoznání podvodu je ostatně u provozovatele sítě účelné, tuto testovací hodnotu používat v jiném měřítku, což samo o sobě nemá žádný vliv na správnost postupu. Výsledky použití Fisherovy diskriminantní analýzy jsou znázorněny na obrázku 28. Na ose y jsou vyneseny hodnoty statistiky (1) hodnot i/n_1 pro čestné uživatele a hodnot j/n_2 pro podvodníky.

Ukazuje se, že přibližně 30 % dat podvodníků leží pod čarou danou y = 0.0025. Nechá se očekávat, že pomocí příslušného diskriminantního testu vynikne vysoký podíl podvodníků. Všimněme si, že různé datové soubory patří stejné osobě; v pojednávaném případu existuje 213 datových souborů k 57 rozpoznaným podvodníkům. 30 % odpovídá přibližně 70 datových souborů, které byly rozpoznány jako podvodné.

Algoritmus k rozpoznání podvodu

Jsou dána nová pozorování X a kvantily $Q^i_{1-\alpha}$

1. Vypočítej transformaci hlavních komponent $\tilde{X} = T'(X - \mu)$

2. Otestuj transformaci na

$$\tilde{X}_i \leq Q^i_{1-\alpha},\ i = 1,2$$

$\tilde{X} > Q^i_{1-\alpha}$ pro i — X je podvodník, jdi na 3.

$\tilde{X}_i \leq Q^i_{1-\alpha}$ pro i — X je normální zákazník, jdi na 4.

3. Modifikace očekávané hodnoty a kovarianční matice u podvodníků

$$\bar{Y} \leftarrow \frac{nY + X}{n+1}$$

$$R_b \leftarrow \frac{nR_b + (X - \bar{Y})(X - \bar{Y})'}{n+1}$$

4. Modifikace očekávané hodnoty a kovarianční matice u normálních zákazníků

$$\bar{\mu} \leftarrow \frac{n\mu + X}{n+1}$$

$$R_e \leftarrow \frac{nR_e + (X - \mu)(X - \mu)'}{n+1}$$

Posouzení zvolených filtrů MEGS

V tomto oddíle se vyhodnocují zvolené filtry systému MEGS ve smyslu pravděpodobností chyb 1. a 2. typu, které se získávají z empirického modelu rozdělení telefonních čísel.

Každý filtr i se může chápat jako test hypotézy

H^i_0 : účastník není podvodník

H^i_1 : účastník je podvodník

tzn. zákazník je podchycen filtrem i, tedy se akceptuje

hypotéza H_1^1, tedy jeho podvodný úmysl.

U testů hypotéz tohoto typu se mohou udělat dvě různé chyby. Normální zákazník se může jednak mylně rozpoznat jako podvodník. Tento omyl se nazývá chyba 1. typu, nebo také α - chyba. Jednak se může přirozeně přihodit, že se u testovaného účastníka jedná o podvodníka a test se přesto rozhoduje pro hypotézu H_0^1. Toto se potom označuje jako chyba 2. typu, nebo jako β - chyba. Ve smyslu stanovení problému je účelné, omezit chybu 1. typu, aby se zbytečně neztráceli zákazníci, jejichž připojení bylo mylně zablokováno.

Dříve zmíněné pravděpodobnosti chyb by se nyní měly popsat na základě zjištěných empirických pravděpodobností, abychom tak získali měřítko hodnocení použitých filtrů. Následující tabulka ukazuje soupis výsledků tohoto vyšetřování.

popis filtru	α-chyba	β-chyba
'0190' 100 DM/den	0.000042	0.987325
'roaming' ≥ 500 DM/den	0.000004	1
světová zóna 2+9 ≥ 300 DM/den	0	0.995955
1 zavolání světová zóna 2	0.000462	0.837108
1 zavolání světová zóna 8	0.000389	0.989753
1 zavolání světová zóna 9	0.000189	0.898056
2 zavolání/den ve světové zóně 2+5	0.000160	0.991640
2 zavolání/den ve světové zóně 2+8	0.000162	0.994337
2 zavolání/den ve světové zóně 2+9	0.000468	0.990291
2 zavolání/den ve světové zóně 5+8	0.000121	0.994876
2 zavolání/den ve světové zóně 5+9	0.000430	0.989482
2 zavolání/den ve světové zóně 8+9	0.000431	0.992179
5 zavolání/den ve světové zóně 2	0.000016	0.995146
5 zavolání/den ve světové zóně 5	0.000012	0.997303
5 zavolání/den ve světové zóně 9	0.000062	0.994876

α-chyba se počítá jako relativní četnost příslušného filtru u účastníků bez podvodného úmyslu a β-chyba jako 1-relativní četnost filtru u podvodníků.

Nechá se rozpoznat, že pozorované filtry vykazují velmi malou pravděpodobnost chyby 1. typu, ale také velmi vysokou β-chybu.

Provozovatel sítě používá ještě velký počet dalších filtrů, které se ale k tomuto okamžiku nemohou vyhodnocovat, protože tyto používají stáří zákazníka, tedy jeho příslušnost k třídě, která bohužel není obsažena v datech, která jsou k dispozici.

Přehled obrázků na výkresech

Vynález bude blíže vysvětlen prostřednictvím konkrétních příkladů provedení znázorněných na výkresech, na kterých představuje

obr. 1 relativní četnost obratů hovorů u normálních zákazníků,

obr. 2 relativní četnost obratů hovorů u podvodníků,

obr. 3 relativní četnost počtu volání na den u podvodníků,

obr. 4 tabulka hodnot empirického rozdělení cílových telefonních čísel,

obr. 5a a 5b tabulka dat rozšíření modelu pro různá mezinárodní telefonní čísla,

obr. 6 tabulka četnosti volání z ciziny do Německa podle cílových telefonních čísel (účastníci bez podvodného úmyslu),

obr. 7 tabulka četnosti volání z ciziny do Německa ve srovnání s ostatními cílovým zeměmi (účastníci bez podvodného úmyslu),

obr. 8 tabulka četnosti zavolání z ciziny do Německa od účastníků s podvodným úmyslem,

obr. 9 tabulka četnosti zavolání z ciziny od účastníků s podvodným úmyslem,

obr. 10 a obr. 11 empirické hodnoty cílových telefonních čísel účastníků s podvodným úmyslem,

obr. 12 až 16 empirická střední hodnota a empirický rozptyl doby rozhovoru u normálních zákazníků a u podvodníků,

obr. 17 až 18 empirické hodnoty rozdělení obratu u normálních zákazníků a u podvodníků,

obr. 19 kontingenční tabulka pro test na párovou nezávislost telefonních čísel v určitém časovém intervalu pozorování,

obr. 20 znázornění hlavních komponent, přiřazených 612 nebo 622,

obr. 21 výsledek k Fisherově diskriminantní analýze,

obr. 22 přístrojově-technické znázornění průběhu způsobu podle obr. 23,

obr. 23 vývojový diagram způsobu k rozpoznání podvodu s přístroji podle obr. 22,

obr. 24 četnosti denních obratů u normálních zákazníků (výřez),

obr. 25 relativní četnosti počtů volání na den u

normálních zákazníků a aproximace pomocí funkce (výřez),

obr. 26 celkový denní obrat vzhledem k obratu na číslech 190,

obr. 27 hlavní komponenty, přiřazené σ_1^2 popř. σ_2^2, a

obr. 28 výsledek Fisherovy diskriminantní analýzy.

Příklady provedení vynálezu

Tabulka podle obr. 4 obsahuje hodnoty empirického rozdělení cílových telefonních čísel pro profily $T_z = \{t_1, t_2, \ldots, t_{12}\}$. Přitom je třeba respektovat, že různé kategorie, jako například 'Service' a 'MTC' se ještě dále člení, aby se mohly lépe extrahovat z předložených dat jednotlivých spojení. Ve druhé tabulce podle obr. 5 jsou obsažena data rozšíření modelu pro různá mezinárodní telefonní čísla. Přitom se jedná o rozštěpení kategorie 'mezinárodní spojení'. Databáze se skládá z 1391739 pozorování.

V tabulce podle obr. 6 a obr. 7 jsou pro účastníky bez podvodného úmyslu uvedena speciálně ještě jednou volání z ciziny do Německa, tedy ta s předvolbou 0049. Relativní četnosti se zde vztahují na příslušný počet spojení 0049 v předcházející tabulce.

Zcela analogicky k tabulkám účastníků bez podvodného úmyslu jsou na obrázcích 8, 9 a 10 vypsány empirické hodnoty cílových telefonních čísel podvodníků ve třech tabulkách.

Obr. 8 přitom uvádí četnost zvolených cílových kategorií, zatímco obrázek 9 (s pokračováním pomocí obrázku 10) ukazuje předvolby, zvolené podvodníky z ciziny, a jejich četnost. Na obr. 11 je zobrazena cílová kategorie, zvolená podvodníky jako nejčastější.

V tabulkách obrázků 13 až 18 jsou znázorněny empirické hodnoty pro rozdělení u určených kategorií cílových telefonních čísel. Obrázek 12 ukazuje rozdělení doby trvání rozhovoru u normálních zákazníků a obrázek 13 u podvodníků.

Obrázek 14 ukazuje počet volání na den u normálních zákazníků a obrázek 15 u podvodníků.

Obrázek 16 a 17 ukazuje rozdělení obratu u normálních zákazníků a obrázek 18 u podvodníků.

Obrázek 19 ukazuje konečně test nezávislosti pro kategorie cílových telefonních čísel. Tabulka ukazuje kontingenční tabulku pro test na párovou nezávislost kategorií telefonních čísel. Přitom je třeba dbát toho, že nebyla pozorována žádna čísla '0130', jedná se tedy o tabulku 9x9, která má příslušné X^2 (Chi-Quadrat) rozdělení, tedy 64 stupňů volnosti. Při vysoké hodnotě statistuky restu se hypotéza nezávislosti přirozeně odmítá pro všechny vhodné úrovně α. Z odmítnutí párové stochastické nezávislosti vyplývá, že nemůže platit žádná společná stochastická nezávislost kategorií cílových telefonních čísel. Toto platí adekvátně pro všechny dny.

Obrázek 20 ukazuje grafické znázornění hlavních komponent, přiřazených σ_1^2 popř. σ_2^2, zatímco obrázek 21 ukazuje výsledek Fisherovy diskriminantní analýzy jako

grafické znázornění.

Obr. 22 ukazuje přístrojově-technické provedení způsobu podle vynálezu, které je znázorněno jako blokové schéma na obr. 23.

Ve znázorněném kroku 1 jsou příkladně znázorněny telekomunikační síťové členy. Označení MSC znamená mobilní ústřednu, znázorněnou počítačem ústředny, zatímco VMS je takzvaný Voice Mail System, se kterým se vytvářejí síťové hlasové výstupy závislé na uživateli. Přístrojové jednotky VAS-NE znamenají dodatečné síťové členy, jako např. členy zpoplatňování a další. Datové soubory provozovatele sítě, které se vytvářejí v tomto přístrojovém prostředí, se přenášejí přes signalizační systém č. 7 (File Transfer Access and Management) k počítači k rozpoznání zneužití. Tento počítač se také označuje jako server zákaznických dat. Zde zmíněný signalizační systém č. 7 (FTAM) je 7-vrstvý protokol, který přenáší celé svazky datových souborů najednou. Jedná se tedy o datové pole (File), ve kterém je obsaženo mnoho tisíc datových souborů, které se Online přenášejí na server zákaznických dat.

V kroku 2 se nechá probíhat celý způsob, jehož blokový diagram je na obr. 22 a 23 vyznačený jako krok 2. Důležité je, že kroky výpočtu, jako transformace hlavních komponent (Fisherova diskriminantní analýza a všechny ostatní kroky výpočtu) probíhají v reálném čase na tomto počítači.

Vypočítají se výsledky a mohou se v kroku 3 přenášet v reálném čase na zákaznickou pracovní stanici (Customer Care Workstation). U konzole této stanice nyní sedí obsluhující osoba, a na její obrazovce vzniká optický a/nebo

akustický alarm, pokud se rozpoznalo zneužití. Obsluhující osoba potom může ještě během probíhajícího neoprávněného hovoru zasáhnout a například tento hovor přerušit, nebo zabránit opakované volbě atp. Může se rovněž vydat akustická výstraha, týkající se zneužití.

Existuje také zpětné hlášení od zákaznického serveru na počítač, týkající se zneužití. Obsluhující osoba může například změnit na zákaznickém počítači (při vyvolaném alarmu) práh alarmu nebo jiná kritéria zakročení. Tato data se sdělují počítači, týkajícího se zneužití, který se z nich učí a zahrnuje je do svých výpočtů.

Obrázek 23 ukazuje vývojový diagram způsobu podle vynálezu. Zde probíhající kroky jsou uloženy ve znacích nároku 1.

Důležité je zpětné vedení z funkčních bloků, uspořádaných v obrázku na spodním okraji. Přes toto zpětné vedení se uskutečňuje aktualizace vypočítaných kovariančních a středních hodnot. Systém má tedy schopnost se sám učit.

Vyšetřování datového materiálu, který je k dispozici, ukázalo, že se chování účastníků s podvodným úmyslem nechá v podstatě charakterizovat jenom dvěma znaky. Dosažené výsledky v příkladném provedení a znázorněných tabulkách se vztahují na příkladný datový soubor provozovatele sítě. Na základě existujících dat se mohly určit pouze dvě různé skupiny účastníků s podvodným úmyslem.

Zastupuje:
Dr. Miloš Všetečka v.r.

PATENTOVÉ NÁROKY

1. Způsob rozpoznání zneužití služeb provozovatele sítě zákazníkem pomocí online analýzy datových souborů, vztažených k zákazníkovi, s následujícími kroky:

1. Online zjištění souboru vstupních dat ze síťových elementů z následujících komponent:

1.1 kumulované datové soubory přes pevný časový interval, např. 30-denní datové soubory;

1.2 data jednotlivých spojení poslední doby, povolená z hlediska zákonné ochrany dat, ve dnech (t. č. 5 dní): cílové telefonní číslo, doba trvání hovoru, typ spojení atd;

1.3 data, specifická pro zákazníka (stáří zákazníka v síti, typ platby atd.);

2. Akumulování vstupních znaků podle tříd (typ cílového telefonního čísla, počty hovorů, typy hovorů atd.);

3. Provádění analýzy hlavních komponent:

3.1 Provádění analýzy hlavních komponent na datových souborech již rozpoznaných podvodníků;

3.2 spektrální rozložení příslušné kovarianční matice;

3.3 určení relevantních hlavních komponent;

3.4 klasifikace hlavních komponent, relevantních pro podvodné chování;

4. Transformace hlavních komponent nedetekovaných datových souborů na základě spektrálního rozkladu kovarianční matice v kroku 3

5. Znázornění hlavních komponent datových souborů a diskriminace popř. podvodného chování

6. Odhad a zjištění empirického kvantilu hlavních komponent k řízení pravděpodobností chyb prvního a druhého typu při automatickém detekování a generování alarmů;

7. Fisherova diskriminantní analýza ke zjištění dělicí nadroviny mezi datovými soubory identifikovaných podvodníků a normálních zákazníků, s grafickým znázorněním
8. Odhad a zjištění empirických kvantilů promítnutých dat k řízení pravděpodobností chyb 1. a 2. typu při automatickém detekování a generování alarmu.

2. Způsob podle nároku 1, ***vyznačující se tím***, že v kroku 2 způsobu se každý datový soubor z akumulovaných dat znázorňuje vysoce dimenzionálním reálným vektorem.

3. Způsob podle nároku 1 nebo 2, ***vyznačující se tím***, že v 5. kroku způsobu se hlavní komponenty datových souborů znázorňují graficky.

4. Způsob podle některého z nároků 1 až 3, ***vyznačující se tím***, že v 5. kroku způsobu se zjištěné podvodné chování znázorňuje vizuálně.

5. Způsob podle některého z nároků 1 až 4, ***vyznačující se tím***, že datové soubory uživatelů se analyzují podle jejich hovorových znaků a člení se do přidělených tříd, a že rozčlenění do tříd se uskutečňuje na základě chování, že určité pozdravy vyznačují zákazníky s podvodným úmyslem a že se podstatně odlišuje od individuálního chování ostatních účastníků.

6. Způsob podle některého z nároků 1 až 5, ***vyznačující se tím***, že chování zákazníků s podvodným úmyslem se vyznačuje následujícími dvěma znaky:
1. zneužití čísel 0190 a zároveň
2. vedení nápadného množství zahraničních hovorů ve zkoumaném období.

7. Způsob podle některého z nároků 1 až 7, ***vyznačující se tím***, že algoritmus k rozpoznání podvodu se skládá z následujících vztahů:

Jsou dána nová pozorování X a kvantily $Q^i_{1-\alpha}$

1. Vypočítej transformaci hlavních komponent $\overline{X} = T'(X - \mu$

Otestuj transformaci na

$$\tilde{X}_i \leq Q^i_{1-\alpha},\ i = 1,2$$

$\tilde{X} > Q^i_{1-\alpha}$ pro *i* — *X* je podvodník, jdi na 3.

$\tilde{X}_i \leq Q^i_{1-\alpha}$ pro *i* — *X* je normální zákazník, jdi na 4.

3. Modifikace očekávané hodnoty a kovarianční matice u podvodníků

$$\overline{Y} \leftarrow \frac{nY + X}{n + 1}$$

$$R_b \leftarrow \frac{nR_b + (X - \overline{Y})(X - \overline{Y})'}{n + 1}$$

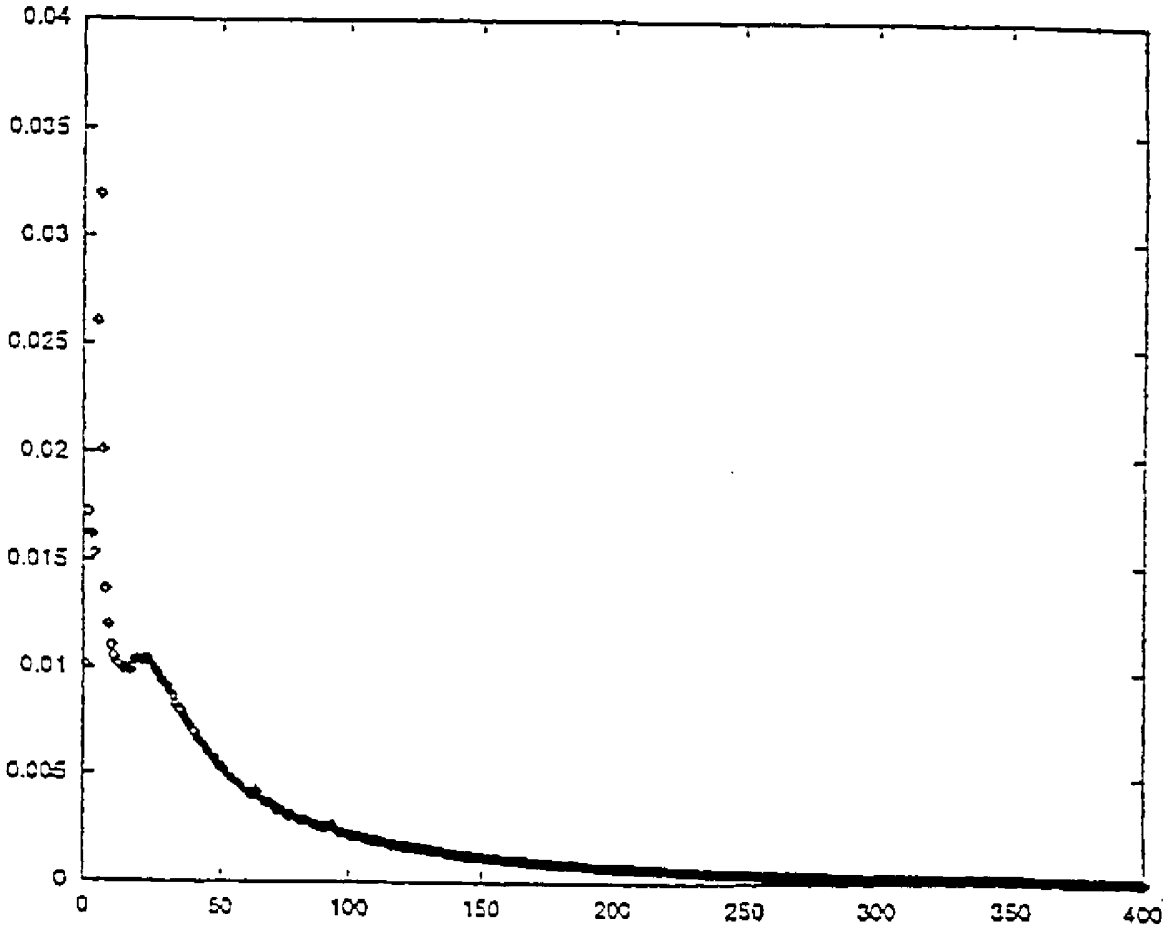
4. Modifikace očekávané hodnoty a kovarianční matice u normálních zákazníků

$$\overline{\mu} \leftarrow \frac{n\mu + X}{n + 1}$$

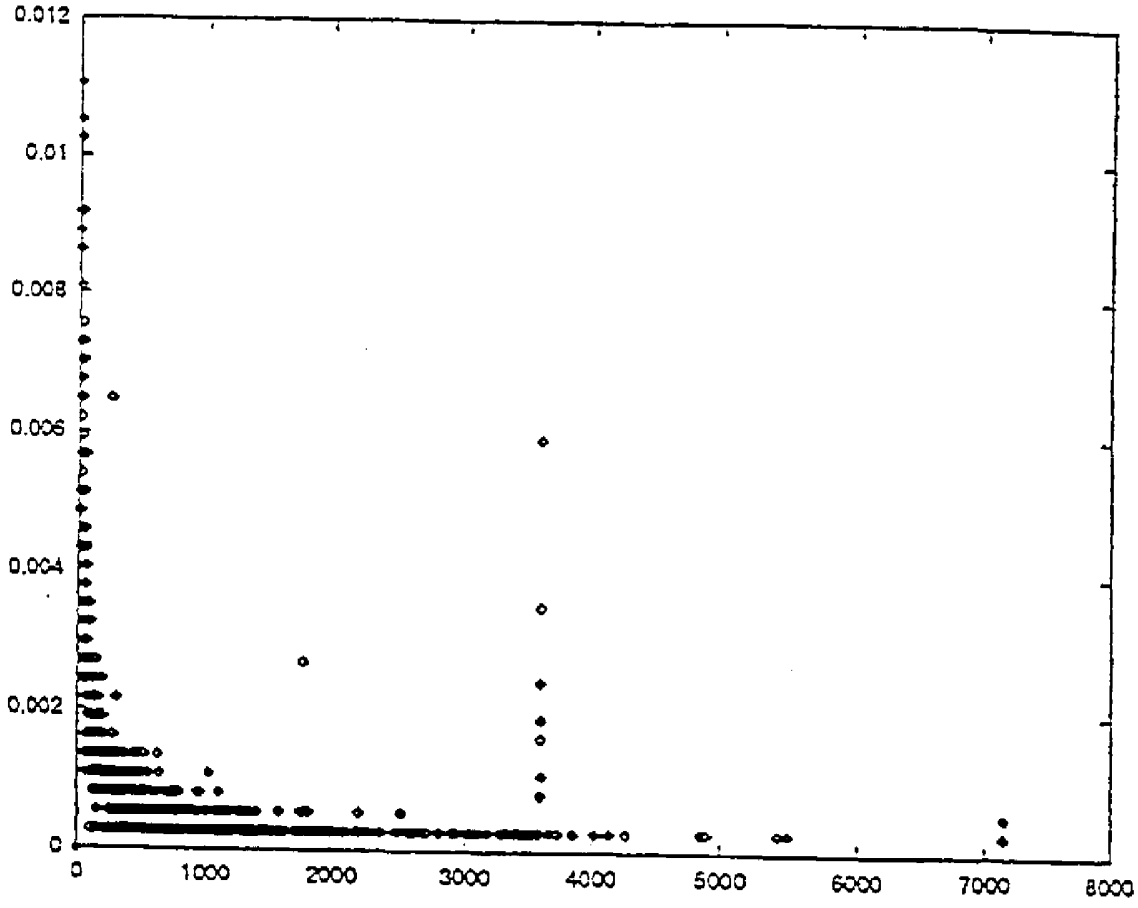
$$R_e \leftarrow \frac{nR_e + (X - \mu)(X - \mu)'}{n + 1}$$

8. Způsob podle některého z nároků 1 až 6, ***vyznačující se tím***, že rozpoznání podvodníků se uskutečňuje Fisherovou diskriminantní analýzou.

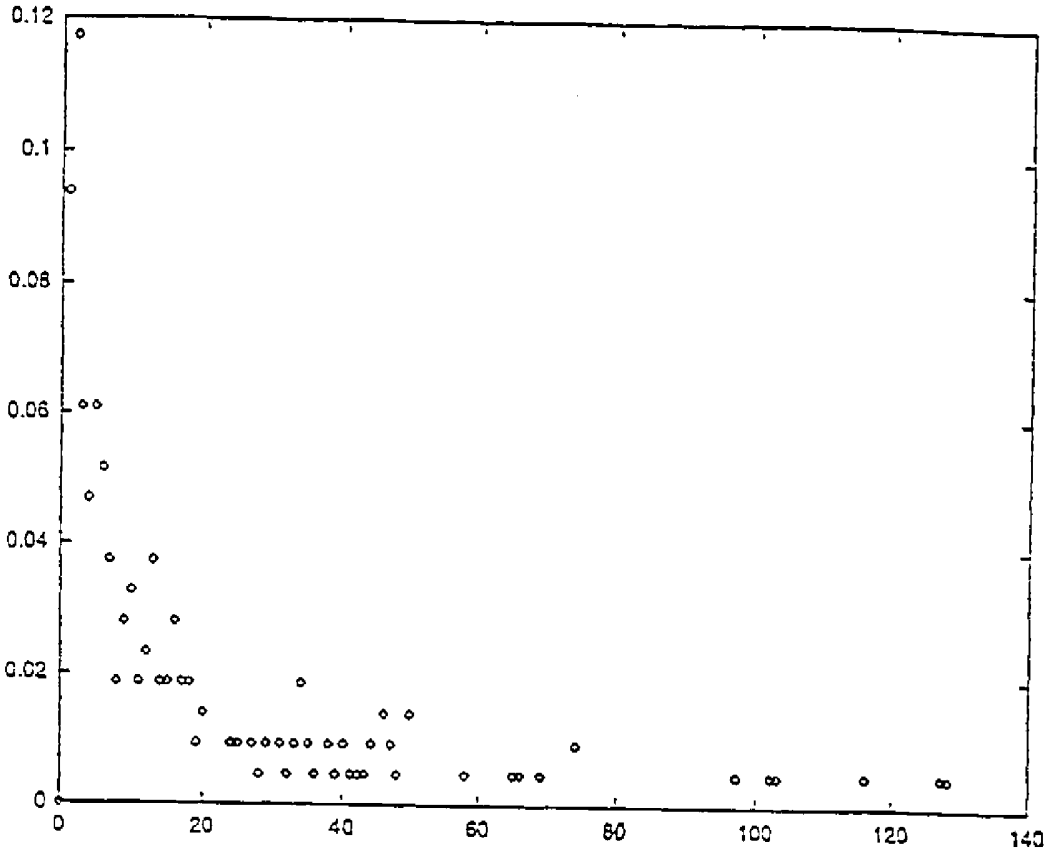
Zastupuje:
Dr. Miloš Všetečka v.r.


0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
50
100
150
200
250
300
350
400

obr. 1


0.012
0.01
0.008
0.006
0.004
0.002
0
0
1000
2000
3000
4000
5000
6000
7000
8000

obr. 2



obr. 3

cílová kategorie	počet	relativní četnost
čísla 0177 (E-Plus)	21757	0.015633
čísla 0171 (D1)	229594	0.164969
čísla 0172 (D2)	76415	0.054906
čísla 0130	0	0.000000
čísla 0180	2235	0.001606
čísla 0190	3119	0.002241
2-čísla služeb (Service)	14577	0.010474
3-čísla služeb (Service)	159703	0.114751
zvláštní čísla (nár.)	631715	0.453903
roaming-spojení	35452	0.025473
MTC(Flag:2/2+4)	125321	0.090046
MTC+(int.)+roaming	35457	0.025477
Fax-volání	0	0.000000
mezinárodní spojení	56394	0.040521

obr. 4

předvolba	počet	rel. četnost	předvolba	počet	rel. četnost
004175	22	0.000016	003393	11	0.000008
001809	8	0.000006	001787	1	0.000001
001441	6	0.000004	001345	4	0.000003
001246	2	0.000001	00994	2	0.000001
00974	1	0.000001	00973	3	0.000002
00972	92	0.000066	00971	20	0.000014
00966	47	0.000034	00965	14	0.000010
00964	2	0.000001	00963	35	0.000025
00962	42	0.000030	00961	114	0.000082
00960	2	0.000001	00886	14	0.000010
00880	5	0.000004	00874	2	0.000001
00856	1	0.000001	00855	1	0.000001
00853	4	0.000003	00852	30	0.000022
00677	47	0.000034	00599	290	0.000208
00594	1	0.000001	00593	2	0.000001
00592	5	0.000004	00591	1	0.000001
00507	3	0.000002	00506	4	0.000003
00502	2	0.000001	00421	77	0.000055
00420	678	0.000487	00396	13	0.000009
00389	76	0.000055	00387	260	0.000187
00386	161	0.000116	00385	454	0.000326

obr. 5a

předvolba	počet	rel. četnost	předvolba	počet	rel. četnost
00381	432	0.000310	00380	177	0.000127
00378	2	0.000001	00375	182	0.000131
00374	16	0.000011	00373	15	0.000011
00372	25	0.000018	00371	62	0.000045
00370	120	0.000086	00359	78	0.000056
00358	150	0.000108	00357	22	0.000016
00356	43	0.000031	00355	11	0.000008
00354	1	0.000001	00353	52	0.000037
00352	197	0.000142	00351	265	0.000190
00350	1	0.000001	00299	4	0.000003
00264	11	0.000008	00263	2	0.000001
00262	3	0.000002	00258	2	0.000001
00256	2	0.000001	00254	17	0.000012
00251	2	0.000001	00249	79	0.000057
00248	4	0.000003	00245	12	0.000009
00244	3	0.000002	00243	4	0.000003
00241	12	0.000009	00239	64	0.000046
00237	7	0.000005	00234	34	0.000024
00233	42	0.000030	00232	1	0.000001
00229	2	0.000001	00228	7	0.000005
00225	8	0.000006	00224	1	0.000001
00223	15	0.000011	00221	7	0.000005
00220	13	0.000009	00218	2	0.000001
00216	48	0.000034	00213	30	0.000022
00212	82	0.000059	0098	58	0.000042
0094	806	0.000579	0093	14	0.000010
0092	24	0.000017	0091	81	0.000058
0090	1305	0.000938	0086	70	0.000050
0084	189	0.000136	0082	36	0.000026
0081	188	0.000135	0066	57	0.000041
0065	32	0.000023	0064	5	0.000004
0063	35	0.000025	0062	27	0.000019
0061	45	0.000032	0060	23	0.000017
0058	6	0.000004	0057	20	0.000014
0056	35	0.000025	0055	51	0.000037

obr. 5b

cílová kategorie	počet	relativní četnost
čísla 0177 (E-Plus)	228	0.007291
čísla 0171 (D1)	8926	0.285449
čísla 0172 (D2)	1461	0.046722
čísla 0130	0	0.000000
čísla 0180	19	0.000608
čísla 0190	0	0.000000
2-čísla služeb (Service)	0	0.000000
3-čísla služeb (Service)	0	0.000000
zvláštní čísla (nár.)	20636	0.659930

obr. 7

předvolba	počet	rel. četnost	předvolba	počet	rel. četnost
0054	16	0.000011	0053	3	0.000002
0052	10	0.000007	0051	8	0.000006
0049	31270	0.022468	0048	1674	0.001203
0047	121	0.000087	0046	317	0.000228
0045	329	0.000236	0044	1317	0.000946
0043	2449	0.001760	0041	1490	0.001071
0040	278	0.000200	0039	1700	0.001221
0036	323	0.000232	0034	775	0.000557
0033	1254	0.000901	0032	783	0.000563
0031	1711	0.001229	0030	416	0.000299
0027	52	0.000037	0020	74	0.000053
007	504	0.000362	001	1483	0.001066

obr. 6

cílová kategorie	počet	relativní četnost
čísla 0177 (E-Plus)	56	0.015102
čísla 0171 (D1)	268	0.072276
čísla 0172 (D2)	139	0.037487
čísla 0130	0	0.000000
čísla 0180	3	0.000809
čísla 0190	638	0.172060
2-čísla služeb (Service)	58	0.015642
3-čísla služeb (Service)	73	0.019687
zvláštní čísla (nár.)	652	0.175836
roaming-spojení	29	0.007821
MTC(Flag:2/2+4)	5	0.001348
MTC+(int.)+roaming	4	0.001079
Fax-volání	0	0.000000
mezinárodní spojení	1764	0.475728

obr. 8

předvolba	počet	rel. četnost	předvolba	počet	rel. četnost
001809	18	0.004854	00977	27	0.007282
00974	2	0.000539	00971	14	0.003776
00966	1	0.000270	00965	1	0.000270
00964	3	0.000809	00963	1	0.000270
00962	1	0.000270	00961	68	0.018339
00873	9	0.002427	00852	2	0.000539
00677	2	0.000539	00599	64	0.017260
00593	44	0.011866	00591	1	0.000270
00420	3	0.000809	00389	5	0.001348
00387	10	0.002697	00386	1	0.000270
00385	2	0.000539	00381	160	0.043150
00380	1	0.000270	00375	1	0.000270
00374	48	0.012945	00373	4	0.001079
00371	1	0.000270	00359	1	0.000270
00355	21	0.005663	00351	3	0.000809
00258	2	0.000539	00254	2	0.000539
00251	11	0.002967	00245	1	0.000270
00244	1	0.000270	00243	3	0.000809
00241	4	0.001079	00239	1	0.000270
00237	10	0.002697	00234	57	0.015372
00233	37	0.009978	00232	9	0.002427

obr. 9

předvolba	počet	rel. četnost	předvolba	počet	rel. četnost
00229	10	0.002697	00228	91	0.024542
00227	5	0.001348	00226	35	0.009439
00225	34	0.009169	00224	108	0.029126
00223	14	0.003776	00221	10	0.002697
00220	29	0.007821	00213	1	0.000270
00212	2	0.000539	0098	44	0.011866
0094	19	0.005124	0092	59	0.015912
0091	58	0.015642	0090	80	0.021575
0084	19	0.005124	0081	8	0.002157
0066	1	0.000270	0065	25	0.006742
0061	1	0.000270	0060	1	0.000270
0058	3	0.000809	0057	67	0.018069
0054	1	0.000270	0049	48	0.012945
0048	1	0.000270	0046	9	0.002427
0045	1	0.000270	0044	14	0.003776
0043	4	0.001079	0041	5	0.001348
0039	34	0.009169	0036	10	0.002697
0034	10	0.002697	0033	33	0.008900
0032	12	0.003236	0031	11	0.002967
0030	8	0.002157	0027	1	0.000270
0020	126	0.033981	007	6	0.001618
001	49	0.013215			

obr. 10

cílová kategorie	počet	relativní četnost
čísla 0177 (E-Plus)	0	0.000000
čísla 0171 (D1)	33	0.687500
čísla 0172 (D2)	0	0.000000
čísla 0130	0	0.000000
čísla 0180	0	0.000000
čísla 0190	0	0.000000
2-čísla služeb	0	0.000000
3-čísla služeb	0	0.000000
zvláštní čísla (nár.)	15	0.312500

obr. 11

rozdělení doby hovoru: empirické hodnoty, normální zákazníci			
	střední hodnota	rozptyl	počet pozorování
P(D = x)	89.031	33500.2	1347987
P(D = x, Z = „0171“)	88.9203	29766.6	224455
P(D = x, Z = „0172“)	86.9104	27975.3	74341
P(D = x, Z = „0177“)	68.9554	32455.2	21234
P(D = x, Z = „0180“)	109.505	35386.2	2182
P(D = x, Z = „0190“)	240.873	172160	2990
P(D = x, Z = „zahraničí“)	161.723	75721.4	51968
P(D = x, Z = „ nár. číslo“)	110.289	39925	614483
P(D = x, Z = „roaming“)	129.4	48994.2	31099
P(D = x, Z = „Service“)	33.2223	3412.86	170192
P(D = x, Z = „MTC“)	34.2888	18491	155043

obr. 12

rozdělení doby hovoru: empirické hodnoty, podvodníci			
	střední hodnota	rozptyl	počet pozorování
P(D = x)	419.614	570437	3708
P(D = x, Z = „0171“)	98.7649	39801.6	268
P(D = x, Z = „0172“)	75.2662	12325	139
P(D = x, Z = „0177“)	94.0893	29788.3	56
P(D = x, Z = „0180“)	79.3333	4076.33	3
P(D = x, Z = „0190“)	955.9	1492920	638
P(D = x, Z = „zahraničí“)	441.524	398170	1764
P(D = x, Z = „ nár. číslo“)	162.245	226634	652
P(D = x, Z = „roaming“)	94.931	11477.6	29
P(D = x, Z = „Service“)	72.7176	10024.1	131
P(D = x, Z = „MTC“)	239.786	33575.8	28

obr. 13

hovory během jednoho dne: empirické hodnoty, normální zákaznici			
	střední hodnota	rozptyl	počet pozorování
P(N = n)	3.60083	19.6097	278679
P(N = n, Z = „0171“)	2.27225	7.01537	79865
P(N = n, Z = „0172“)	1.81821	2.861	32318
P(N = n, Z = „0177“)	1.76592	3.36558	9326
P(N = n, Z = „0180“)	1.61302	3.89941	1137
P(N = n, Z = „0190“)	2.24923	7.26093	979
P(N = n, Z = „zahraničí“)	2.4757	9.0543	15842
P(N = n, Z = „ nár. číslo“)	2.66706	8.30287	191501
P(N = n, Z = „roaming“)	3.014	14.922	7645
P(N = n, Z = „Service“)	1.73214	1.96642	80411
P(N = n, Z = „MTC“)	2.31401	9.35516	13162

obr. 14

hovory během jednoho dne: empirické hodnoty, podvodníci			
	střední hodnota	rozptyl	počet pozorování
P(N = n)	17.4085	517.762	213
P(N = n, Z = „0171“)	3.72222	14.7668	72
P(N = n, Z = „0172“)	2.57407	4.62648	54
P(N = n, Z = „0177“)	2.54545	3.68831	22
P(N = n, Z = „0180“)	1.5	0.5	2
P(N = n, Z = „0190“)	9.11429	89.436	70
P(N = n, Z = „zahraničí“)	19.6	770.625	90
P(N = n, Z = „ nár. číslo“)	4.72464	34.055	138
P(N = n, Z = „roaming“)	4.83333	26.9667	6
P(N = n, Z = „Service“)	2.25862	3.6337	58
P(N = n, Z = „MTC“)	3.11111	35.6111	9

obr. 15

rozdělení obratu: empirické hodnoty,normální zákaznici			
	střední hodnota	rozptyl	počet pozorování
P(C= y)	50.8544	28729.6	1347987
P(C= y, Z = „0171“)	29.8012	2438.51	224455
P(C= y, Z = „0172“)	45.763	7093.21	74341
P(C= y, Z = „0177“)	33.8373	6813.77	21234
P(C= y, Z = „0180“)	54.6512	6565.22	2182
P(C= y, Z = „0190“)	361.32	377399	2990
P(C= y, Z = „zahraničí“)	181.384	102244	51968

obr. 16

rozdělení obratu: empirické hodnoty,normální zákazníci			
P(C= y Z = „nár. číslo“)	56.4951	7643.19	614483
P(C= y Z = „roaming“)	169.512	121966	31099
P(C= y Z = „Service“)	12.9313	715.657	170192
P(C= y Z = „MTC“)	31.7846	131584	155043

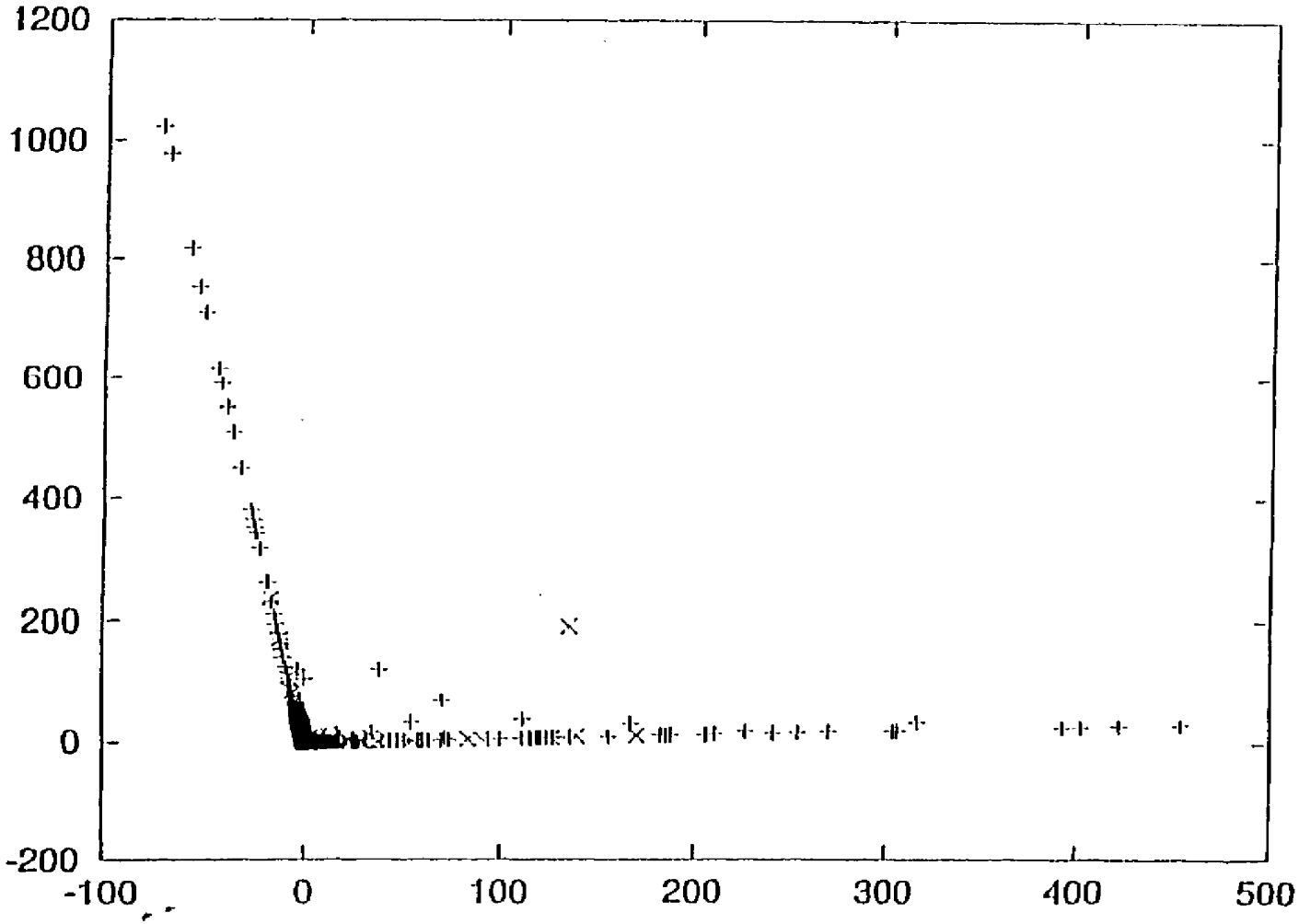
obr. 17

rozdělení obratu: empirické hodnoty, podvodnníci			
	střední hodnota	rozptyl	počet pozorování
P(C = y)	674.278	2662240	3708
P(C = y, Z = „0171“)	25.9254	2111.94	268
P(C = y, Z = „0172“)	25.9353	1373.21	139
P(C = y, Z = „0177“)	31.8571	4171.51	56
P(C = y, Z = „0180“)	39.3333	702.333	3
P(C = y, Z = „0190“)	1903.82	8094190	638
P(C = y, Z = „zahraničí“)	668.791	1239130	1764
P(C = y, Z = „ nár. číslo“)	65.2255	15790.5	652
P(C = y, Z = „roaming“)	127.31	13829.4	29
P(C = y, Z = „Service“)	55.313	6653.39	131
P(C = y, Z = „MTC“)	1425.5	38492800	28

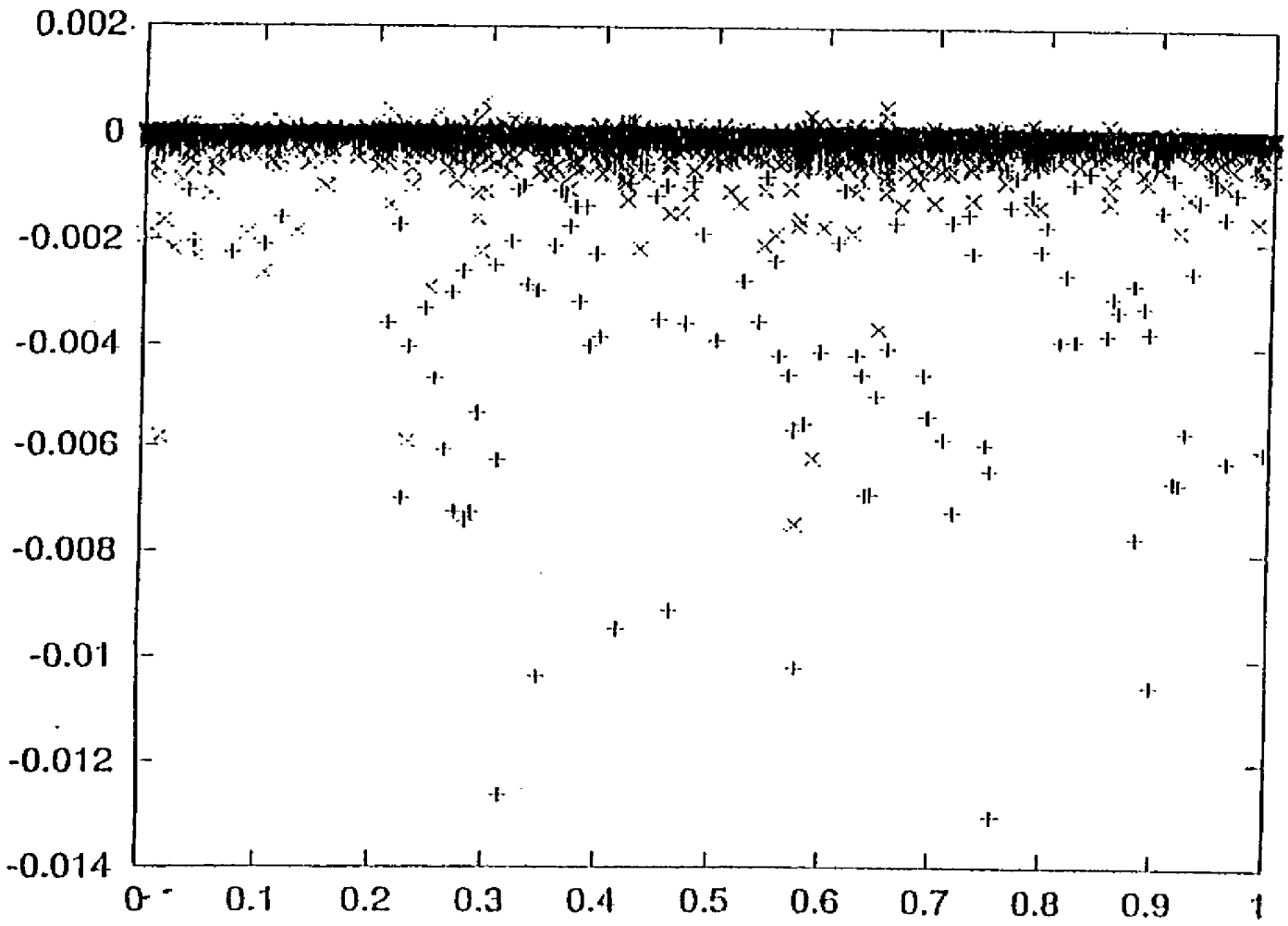
obr. 18

kontingenční tabulka pro 1. srpen 1989, segment 28											
	0177	0171	0172	0130	0180	0190	2...	3...	mezinár.	ostatní	
0177	57	22	4	0	1	0	1	3	1	58	147
0171	23	843	84	0	0	1	11	75	21	549	1607
0172	6	80	224	0	1	2	1	31	8	257	610
0130	0	0	0	0	0	0	0	0	0	0	0
0180	0	0	1	0	2	0	0	1	0	9	13
0190	1	1	0	0	0	15	0	1	0	8	26
2...	2	8	1	0	0	0	8	5	3	33	60
3...	14	137	58	0	1	0	3	355	14	487	1069
mezinár.	1	22	11	0	1	0	2	9	248	130	424
ostatní	59	611	247	0	9	9	34	338	131	4034	5472
	163	1724	630	0	15	27	60	818	426	5565	9428
hodnota Chi-quadrat statistiky testu: 10905.5											

obr. 19

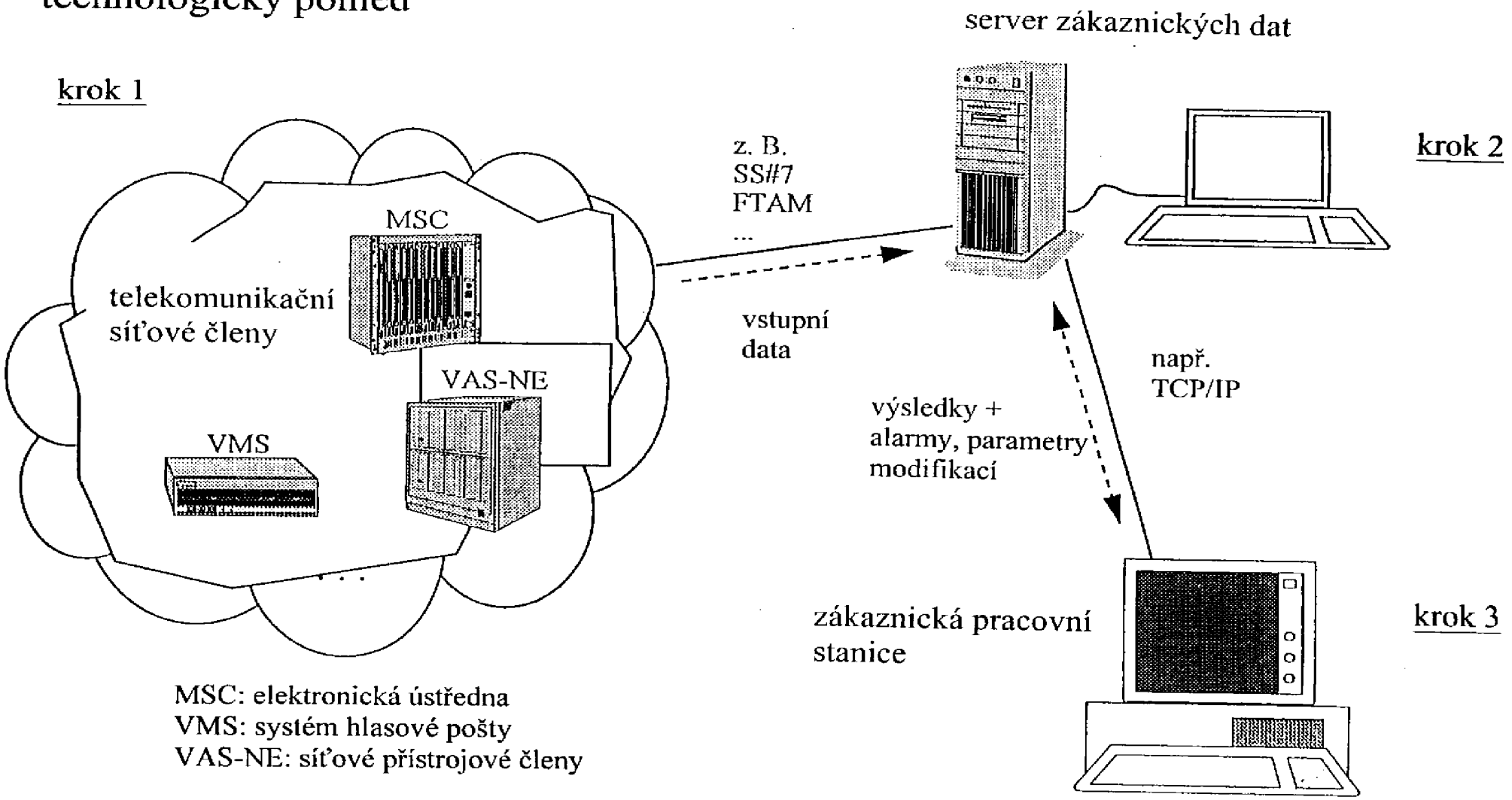


obr. 20


0.002
0
-0.002
-0.004
-0.006
-0.008
-0.01
-0.012
-0.014
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

obr. 21

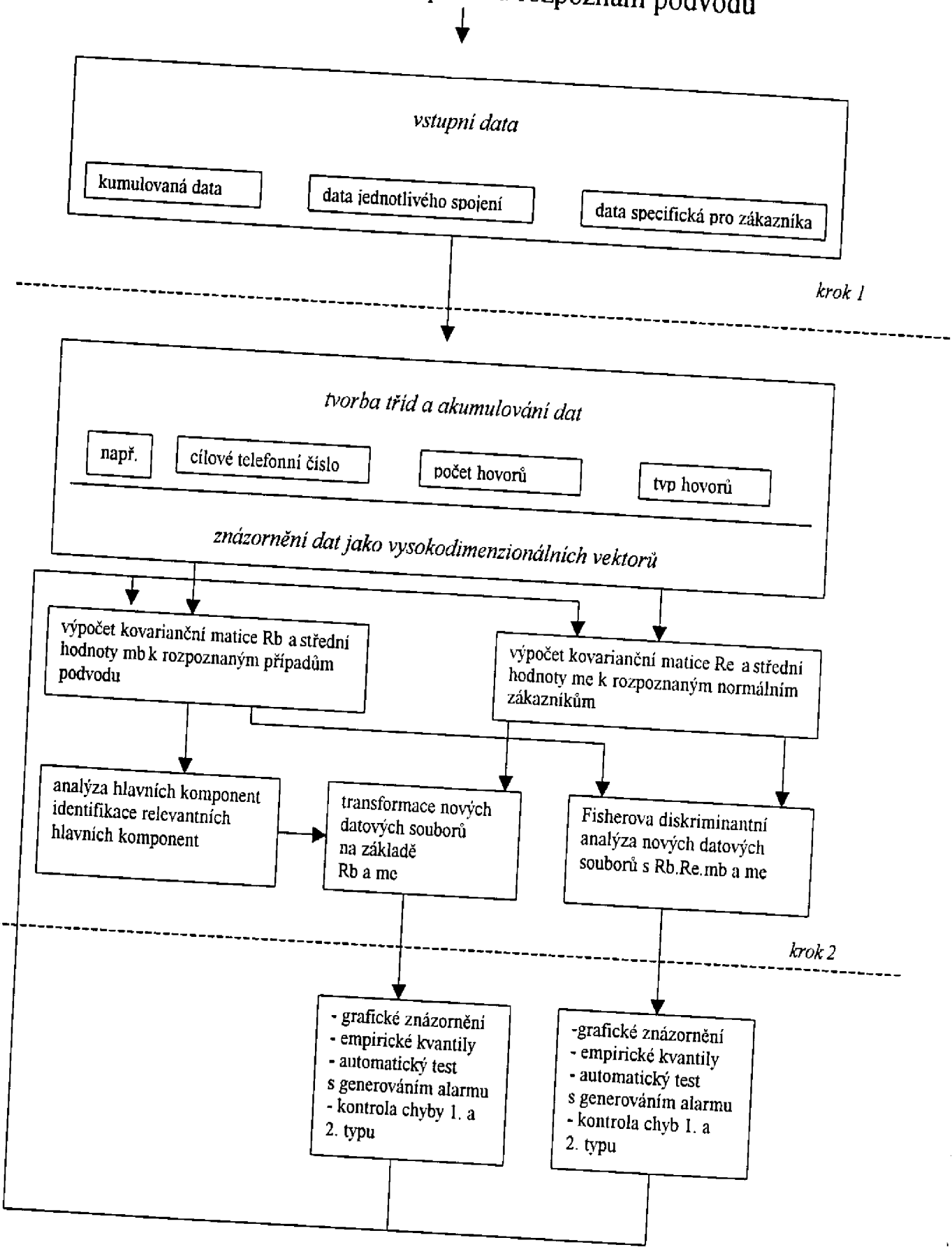
technologický pohled



MSC: elektronická ústředna
VMS: systém hlasové pošty
VAS-NE: síťové přístrojové členy

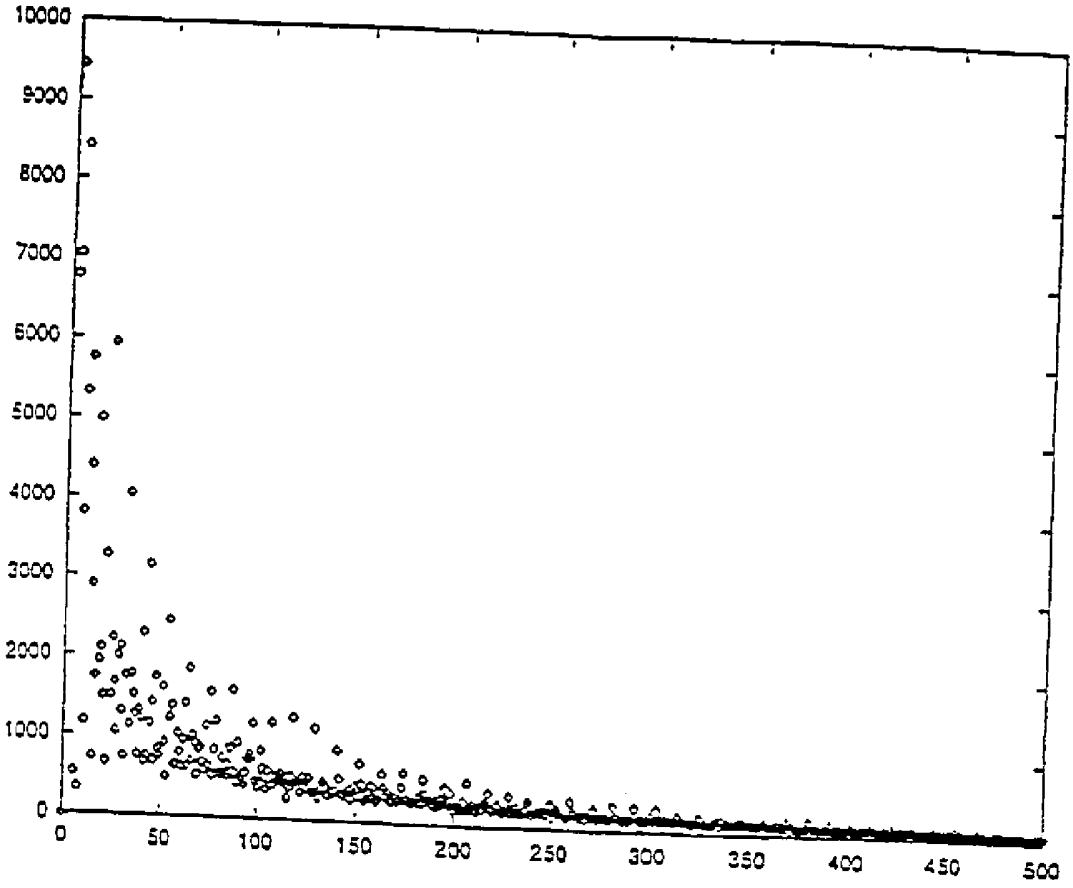
obr. 22

Vývojový diagram způsobu rozpoznání podvodu

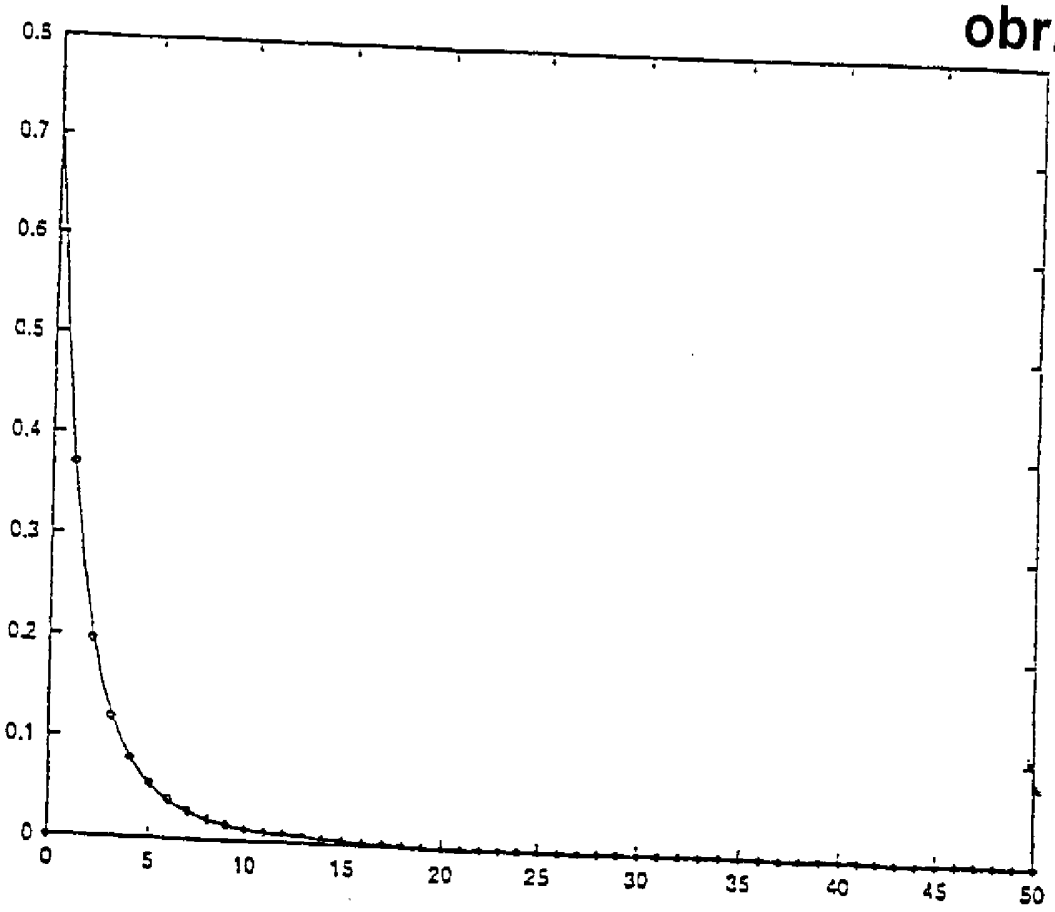


obr. 23

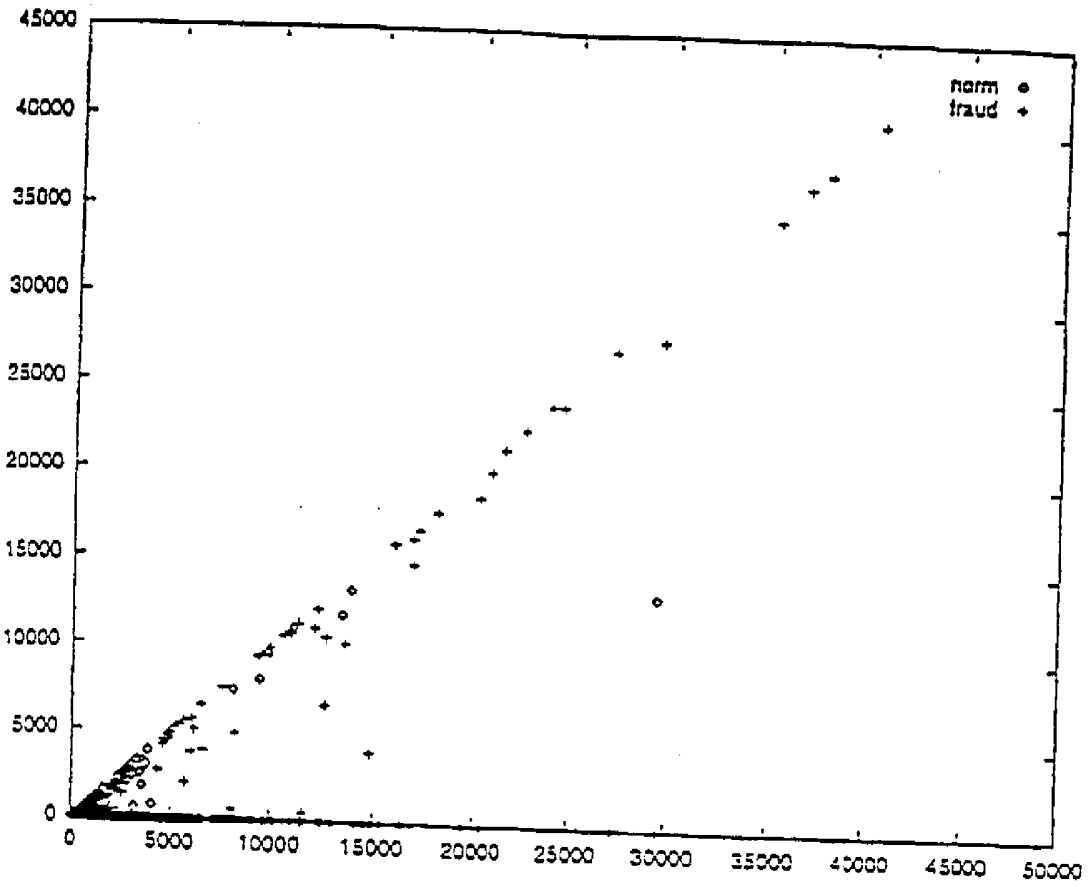
obr. 24


10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
200
250
300
350
400
450
500

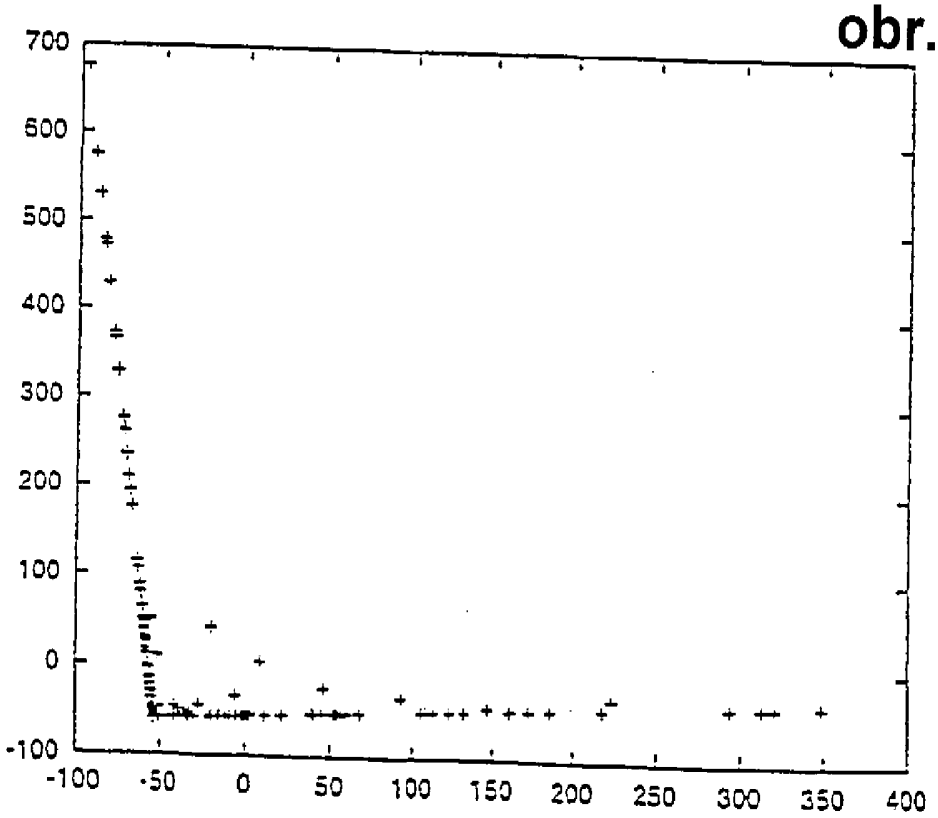
obr. 25

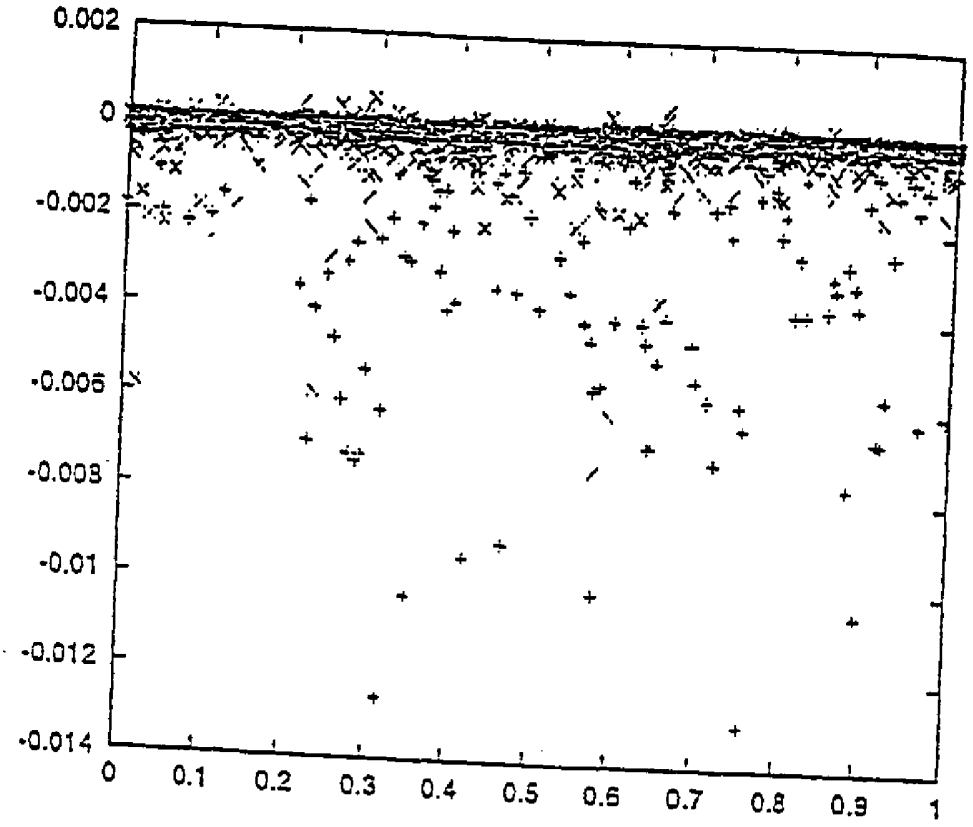

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45
50

obr. 26


norm
fraud
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000

obr. 27


700
600
500
400
300
200
100
0
-100
-100
-50
0
50
100
150
200
250
300
350
400


0.002
0
-0.002
-0.004
-0.006
-0.008
-0.01
-0.012
-0.014
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

obr. 28