PEMBOBOTAN KATA SMS SPAM

Jufriadif Na'am
Dosen Tetap Universitas Putra Indonesia YPTK Padang
Lubuk Begalung, Padang, Sumatera Barat
Email: jupriadifnaam@gmail.com

ABSTRAK

Murahnya biaya pengiriman SMS mengakibatkan tingginya penyalahgunaan dalam pengiriman SMS. Pesan yang disalahgunakan ini disebut dengan SMS Spam. SMS Spam tidak hanya mengganggu tetapi juga melakukan penipuan yang dapat merugikan si penerima. Kategori penipuan SMS Spam dapat dikelompokan menjadi advertising, smishing, remittance, scamming, dan forecasting. Untuk mengetahui kategori penipuan dari sebuah pesan dapat ditentukan dari jumlah bobot kata yang terkandung dalam pesan tersebut.

Kata kunci : SMS Spam, advertising, smishing, remittance, scamming, forecasting, Naive Bayesian, pembobotan, penipuan.

ABSTRACT

Low cost of sending SMS leads to high abuse in the delivery of SMS. Abused messages referred to as SMS Spamming. SMS Spamming is not only annoying but also committed fraud that harm the recipient. SMS Spamming can be categorized as advertising, smishing, remittance, scamming, and forecasting. To determine the category of a message, we can determine it from the weight of words contained in the message.

Keywords : SMS Spam, advertising, smishing, remittance, scamming, forecasting, Naive Bayesian, weighting, fraud.

1. PENDAHULUAN

Perkembangan teknologi *wireless* (nirkabel) dalam era komunikasi data yang semakin cepat dan mengglobal ini telah membawa masyarakat melewati beberapa tahapan pengembangan teknologi sekaligus. Perkembangan teknologi mobile yang semakin meningkat dari setiap generasi ataupun produk terbaru semakin membuat dunia teknologi semakin memanjakan konsumennya. Seiring dengan meningkatnya kebutuhan manusia akan kecanggihan teknologi ini mampu membantu mengatasi permasalahan komunikasi. Salah satu media yang sangat populer digunakan saat ini adalah telepon seluler (ponsel).

Telepon seluler (ponsel) atau telepon genggam (telgam) atau *handphone* (HP) adalah perangkat telekomunikasi elektronik yang mempunyai kemampuan dasar yang sama dengan telepon konvensional saluran tetap, namun dapat dibawa ke mana-mana (portabel;*mobile*) dan tidak perlu disambungkan dengan jaringan telepon menggunakan kabel (nirkabel;*wireless*).

Selain berfungsi untuk melakukan dan menerima panggilan telepon, juga mempunyai fungsi untuk pengiriman dan penerimaan pesan singkat (*short message service;*SMS). SMS seperti yang digunakan pada *handset* modern berasal dari telegrafi radio di pager memo radio menggunakan protokol telepon standar. Ini didefinisikan pada tahun 1985 sebagai bagian dari *Global System for Mobile Communications* (GSM) seri standar sebagai sarana pengiriman pesan hingga 160 karakter ke dan dari *handset* GSM.[4] Meskipun sebagian besar pesan SMS adalah pesan teks *mobile-to-mobile*, dukungan untuk layanan ini telah diperluas untuk mencakup teknologi mobile lainnya, seperti jaringan ANSI CDMA dan Digital AMPS, serta satelit dan jaringan di darat.[4]

SMS adalah aplikasi data yang paling banyak digunakan, dengan perkiraan 3,5 miliar pengguna aktif, atau sekitar 80% dari seluruh pelanggan telepon selular pada akhir tahun 2010[9]. Ini desebabkan karena biaya yang murah dan mudah dalam mengoperasikannya. Pada tahun 2010, 6,1 triliun pesan teks SMS dikirim. [4] Hal ini berarti rata-rata 193.000 SMS per detik.[4] SMS telah menjadi industri komersial besar, berpenghasilan $114.600.000.000 secara global pada tahun 2010.[11] Harga rata-rata untuk satu

pesan SMS adalah $0,11, sementara jaringan *mobile* mengisi interkoneksi antar operator sekitar $0,04 per SMS. [2]

Pesan SMS memiliki beberapa kerentanan keamanan karena fitur dan masalah SMS palsu yang dapat dilakukan dengan mudah. Pada bulan Oktober 2005, peneliti dari Pennsylvania State University menerbitkan sebuah analisis kerentanan di jaringan selular SMS.[10] Para peneliti berspekulasi bahwa penyerang mungkin mengeksploitasi fungsi terbuka jaringan ini sehingga pengguna terganggu. Jenis serangan itu antara lain :

a. *Spoofing* SMS

   Industri GSM telah mengidentifikasi sejumlah serangan yang berpotensial penipuan pada operator seluler yang dapat disampaikan melalui penyalahgunaan layanan pesan SMS. *Spoofing* SMS terjadi ketika penipu memanipulasi informasi alamat dalam rangka untuk meniru pengguna yang telah berkeliaran ke jaringan lain dan mengirimkan pesan ke jaringan lokal. Sering, pesan ini ditujukan kepada tujuan di luar jaringan lokal dengan SMS lokal pada dasarnya sedang dibajak untuk mengirim pesan ke jaringan lain.

b. *Limitation* SMS

   Dalam upaya untuk membatasi telemarketer yang telah diambil untuk membombardir pengguna dengan pengiriman pesan bertubi-tubi tetapi tidak diinginkan oleh si penerima. Seperti di India telah menerapkan peraturan baru pada bulan September 2011, yang hanya boleh 3.000 pesan SMS per pelanggan per bulan, atau rata-rata 100 per pelanggan per hari.[10] Karena representasi yang diterima dari beberapa penyedia layanan dan konsumen, Trai (*Telecom Regulatory Authority of India*) telah meningkatkan batas ini menjadi 200 pesan SMS per *sim card* per hari dalam kasus layanan prabayar, dan sampai 6.000 pesan SMS per *sim card* per bulan kasus layanan pasca bayar. Tetapi dinyatakan tidak konstitusional oleh pengadilan tinggi Delhi tetapi tetap ada beberapa keterbatasan.

c. *Flash* SMS

   Sebuah SMS *Flash* adalah jenis SMS yang muncul langsung di layar utama tanpa interaksi pengguna dan tidak secara otomatis disimpan dalam kotak masuk.[11] Hal ini dapat digunakan dalam keadaan darurat seperti alarm kebakaran atau kasus kerahasiaan, seperti dalam memberikan salah *one-time password*.

d. *Silent* SMS

   Silent SMS, juga dikenal sebagai SMS diam atau siluman sms atau ping siluman. SMS ini bekerja untuk menemukan seseorang berada dimana dan dengan demikian untuk membuat profil gerakan. Pesan tidak muncul pada layar dan hanya memicu sinyal akustik saat diterima. Tujuan utamanya untuk memberikan layanan khusus dari operator jaringan untuk setiap ponsel. SMS ini sering dilakukan atas perintah kepolisian untuk menangkap data seperti identifikasi pelanggan. Di Jerman pada tahun 2010 hampir setengah juta Silent SMS dikirim oleh polisi federal, bea cukai dan dinas rahasia *Verfassungsschutz* (kantor untuk perlindungan konstitusi).[11]

e. SMS *Flooding*

   Sejumlah SMS yang datang dengan volume besar yang dikirimkan oleh aplikasi. Adapun asal datangnya SMS tersebut dapat dari jaringan lokal atau antar jaringan.

f. SMS *Faking*

   SMS yang berasal dari hasil manipulasi *originating address* dan smsc (*sms centre*) *address*. Hampir sama dengan SMS *spoofing*, si pelaku memanipulasi smsc sehingga menjadi smsc yang ilegal atau tidak dikenal. Dikarenakan ilegal, maka tidak bisa dilakukan penagihan terminasi *rate* terhadap SMS yang datang.

g. SMS *Masking*

SMS *Masking* atau SMS *Sender Id* atau SMS *Bulk* adalah jenis SMS dimana identitas yang muncul pada *handphone* penerima SMS adalah alphanumeric yang biasanya digunakan untuk menampilkan nama perusahaan atau nama *brand* suatu produk

2. SMS SPAM

SMS Spam adalah SMS yang tidak diinginkan untuk diterima (*Unwanted messages*). Jika suatu operator menerima SMS Spam dari operator lain, sementara skema bisnis antar operator tersebut adalah berbayar atas *incoming sms*, bisa dipastikan timbul kerugian pada operator yang menerima SMS Spam tersebut.

Disamping itu, masih banyak SMS Spam yang terjadi dan sulit dilacak oleh operator, baik itu dalam operator yang sama maupun antar operator. Tetapi ada SMS Spam yang berdasarkan isi (*content*) yang tidak merugikan operator maupun operator lain, tetapi merugikan si penerima. Kerugin disebabkan oleh isi dari SMS tersebut tidak diinginkan oleh penerima, sehingga penerima dapat saja dirugikan bahkan tertipu.

Berikut ini beberapa kategori SMS Spam yang ada:

1. SMS *Scamming*
   SMS yang dikirimkan dengan isi pesan agar si penerima SMS melakukan pemanggilan ulang dengan biaya panggilan yang mahal (*premium call*). Seperti "Dewi bingung mas mau nulis apa,dari pada aq nulis panjang2 tapi serba salah,mending langsung ngobrol di tlpn aja deh mas ini no tlpon aq 0809-1000-888 :_".
2. SMS *Smishing*

   SMS yang dimanipulasi pengirimnya seakan berasal dari perusahaan atau pihak tertentu dengan tujuan mengelabui si penerima SMS. Seperti : "Selamat! No anda mendpt hadiah dr GEBYAR TELKOMSELpoin (pin anda t577rs9) Info klik www.hadiahtelkomsel.weebly.com".
3. SMS *Advertising*

   SMS yang bertujuan untuk mengiklankan suatu produk, barang atau jasa. Seperti : "Bminat VIDEO:Mario TEGUH,YusufMansur,Mualaf2 Dunia,HARUNYahya. Rata2@130episod.@Rp200.000an(Byr Stl Dtg),Utk Lptop/PC sj.SMS Nm Almat Kmplit+KdPos ke 085203643399".
4. SMS *Remittance*

   Agar si penerima SMS melakukan pengiriman atau transfer uang ke nomor rekening tertentu.
   Seperti : "Kirim aja uangnya di Bank BRI a/n DEDI SURYADI No Rek. 3456-01-035105-530".
5. SMS Forecasting
   SMS yang berisi ramalan, seperti "angka top raja togel edisi kamis 25-07-2013 SGP angka ekor 14/41/71/15 pasang 50rbu per angka setelah tembus kirim pulsa 100rbu di NO 085208521355".

3. METODA YANG DIGUNAKAN

Metoda pengolahan data yang dilakukan adalah sebagai berikut:

1. Pengumpulan data.
   Data SMS Spam dikumpulkan dari pesan SMS yang masuk ke nomor *handphone* operator seluler Telkomsel dengan nomor +628126624701 dari bulan September 2012 sampai dengan November 2013.

2. Pengelompokan berdasarkan jenis.
   Data SMS Spam dikelompokan berdasarkan jenis penipuan yang dilakukan, apakah *advertising*, *smishing*, *remittance*, *scamming* atau *forecasting* dengan menggunakan nalar dari kata-kata yang terkandung dalam SMS tersebut.
3. Mencari frekwensi kemunculan setiap kata.
   Tahap ini dilakukan pencarian frekunsi kemunculan setiap kata yang ada dalam pesan tersebut dengan menggunakan algoritma *Naive Bayesian.*
4. Kesimpulan

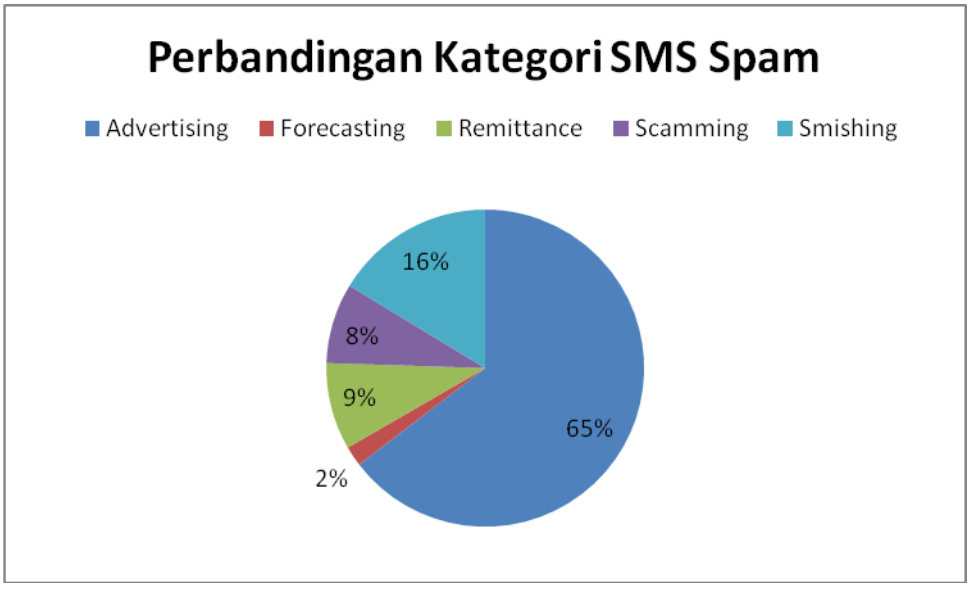
4. PENGOLAHAN DATA

Syarat sebuah pesan dikelompokan kedalam SMS Spam adalah sebagai berikut :

- Nomor pengirim tidak terdaftar pada daftar kontak ponsel atau identitas pengirim tidak dalam bentuk karakter yang tampil dilayar ponsel.
- Pemilik tidak menginginkan pesan tersebut.

Dari hasil seleksi, terkumpul sebanyak 145 pesan yang dikategorikan sebagai SMS Spam. Selanjutnya dilakukan penalaran secara manual untuk mengelompokan pesan tersebut berdasarkan kategori SMS Spam yang telah diuraikan diatas dan hasilnya adalah sebagai berikut :

1. SMS *Advertising* sebanyak 95 pesan.
2. SMS *Smishing* sebanyak 36 pesan.
3. SMS *Remittance* sebanyak 13 pesan.
4. SMS *Scamming* sebanyak 12 pesan.
5. SMS *Forecasting* sebanyak 3 pesan.



Gambar 1. Perbandingan Kategori SMS Spam

Setiap kelompok kategori disimpan dalam sebuah dokumen dan dilanjutkan dengan proses *Naive Bayesian* dengan urutan proses sebagai berikut :

1. *Tokenizing*.

   Pada tahap ini dilakukan proses *tokenizing* yaitu tahap pemotongan *string* data pesan berdasarkan tiap kata yang menyusunnya dan merubah karakter huruf besar menjadi huruf kecil untuk menyamakan pemaknaan dari setiap kata.Klasifikasi Teks

   Setiap dokumen direpresentasikan dalam pasangan atribut (a1, a2, …., a_n), dimana a1 adalah kata pertama, a2 kata kedua dan seterusnya. Sedangkan V adalah label kategori probabilitasnya (V_{MAP}) dengan masukan atribut (a1,a2, …. ,a_n).

   $$\mathbf{V_{MAP} = \underset{Vj \in V}{argmax}\ P(Vj \mid a1, a2, \ldots, an)}$$

   Teorema Bayes menyatakan:

   $$\mathbf{P(B \mid A) = \frac{P(A \mid B)\ P(B)}{P(A)}}$$

   P(a1,a2 … an) nilainya konstan untuk semua Vj sehingga persamaan ini dapat ditulis sebagai berikut :

   $$\mathbf{V_{MAP} = \arg\max_{vje V} P(a1, a2, \ldots, an \mid Vj)\ P(V_j)}$$

   $$\mathbf{P(a1,a2,...,an \mid Vj)\ P(Vj = \prod P(ai \mid Vj)}$$

   Tingkat kesulitan menghitung P(a1, a2 …. an | Vj) menjadi tinggi karena jumlah term P(a1, a2 …. an | Vj) bisa jadi akan sangat besar. Ini disebabkan jumlah term tersebut sama dengan jumlah semua kombinasi posisi kata dikali dengan jumlah kata. Naïve Bayes Classifier menyederhanakan hal ini dengan mengasumsikan bahwa di dalam setiap kelompok, setiap kata independen satu sama lain. Dengan kata lain:

   $$\mathbf{V_{MAP} = \arg\max_{vje V} P(Vj) \prod_i P(a_i \mid V_j)}$$

   - $\mathbf{V_{MAP}}$ = Nilai output hasil kalsifikasi Naive Bayes
   - **P (a1,a2,...,an)** = Peluang kategori A
   - **Vj** = Keadaan atau kategori pesan

2. Pemboboton

   Setelah didapatkan **VMAP** dari setiap kata yang terkandung dalam teks, maka dilakukan pembuangan kata-kata sambung dan kata yang tidak mempunyai makna selain dari singkatan (*stop word*). Selanjutnya dilakukan pencarian bobot kemunculan kata dengan rumus :

   Bobot = (VMAP / n)

   n : jumlah pesan dalam setiap kategori

Selanjutnya diambil 10 bobot tertinggi, seperti tabel berikut:

Tabel 1.
Advertising

No	Kata	VMAP	Bobot
1	www	48	0.51
2	Rp	42	0.44
3	promo	39	0.41
4	hub	33	0.35
5	info	30	0.32
6	com	28	0.29
7	tiket	26	0.27
8	Rb	26	0.27
9	sms	24	0.25
10	online	20	0.21

Tabel 2.
Forecasting

No	Kata	VMAP	Bobot
1	angka	6	2.00
2	edisi	3	1.00
3	pulsa	2	0.67
4	togel	2	0.67
5	rb	2	0.67
6	ingat	2	0.67
7	rbu	2	0.67
8	sgp	2	0.67
9	jitu	2	0.67
10	stlh	2	0.67

Tabel 3
Remittance

No	Kata	VMAP	Bobot
1	no	10	0.77
2	rek	8	0.62
3	bank	5	0.38
4	uangnya	4	0.31
5	bni	4	0.31
6	pulsa	4	0.31
7	rb	3	0.23
8	sms	3	0.23
9	mawkdh	2	0.15
10	mhn	2	0.15

Tabel 4.
Scamming

No	Kata	VMAP	Bobot
1	Aq	10	0.83
2	No	7	0.58
3	Mas	7	0.58
4	Mau	7	0.58
5	Ya	7	0.58
6	Anda	4	0.33
7	Saya	4	0.33
8	Abang	3	0.25
9	Bisa	3	0.25
10	Ibu	3	0.25

Tabel 5.
Smishing

No	Kata	VMAP	Bobot
1	anda	28	0.78
2	info	23	0.64
3	pin	21	0.58
4	www	18	0.50
5	no	17	0.47
6	hadiah	16	0.44
7	saya	14	0.39
8	klik	12	0.33
9	com	11	0.31
10	webs	10	0.28

3. Hasil Pengolahan

Untuk menentukan jumlah bobot menggunakan rumus :

JB = ∑ jk x b

Keterangan:

JB : Jumlah bobot setiap kategori pesan

jk : jumlah karakter setiap kategori yang muncul dalam sebuah pesan

b : nilai bobot setiap karakter dalam kelompok kategori SMS Spam

Sebuah pesan dengan kategori SMS Spam tertentu akan memiliki nilai JB yang tertinggi.

4. Pengujian

Pengujian dilakukan terhadap beberapa SMS Spam, dengan hasil sbb :

a. Pesan ke-1:“Bminat VIDEO:Mario TEGUH,YusufMansur,Mualaf2 Dunia,HARUNYahya. Rata2@130episod.@Rp200.000an(Byr Stl Dtg),Utk Lptop/PC sj.SMS Nm Almat Kmplit+KdPos ke 085203643399”

b. Pesan ke-2: “Kirim aja uangnya di Bank BRI a/n DEDI SURYADI No Rek. 3456-01-035105-530”

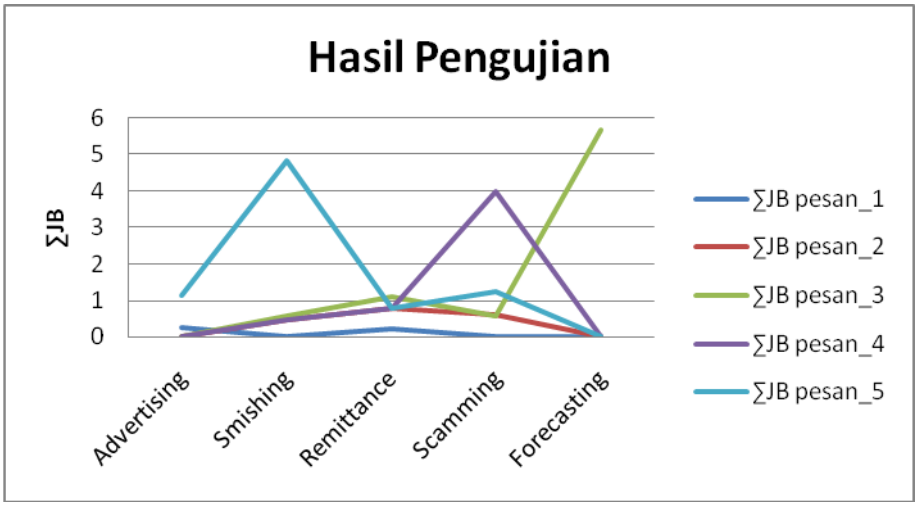
c. Pesan ke-3: “Dewi bingung mas mau nulis apa,dari pada aq nulis panjang2 tapi serba salah,mending langsung ngobrol di tlpn aja deh mas ini no tlpon aq 0809-1000-888 :_”

d. Pesan ke-4:“angka top raja togel edisi kamis 25-07-2013 SGP angka ekor 14/41/71/15 pasang 50rbu per angka setelah tembus kirim pulsa 100rbu di NO 085208521355”

e. Pesan ke-5: “Selamat! No anda mendpt hadiah dr GEBYAR TELKOMSELpoin (pin anda t577rs9) Info klik www.hadiahtelkomsel.weebly.com”

Tabel 6. Hasil Pengujian

Kategori	Advertising	Smishing	Remittance	Scamming	Forecasting
∑JB pesan_1	**0.25**	0	0.23	0	0
∑JB pesan_2	0	0.47	**0.77**	0.58	0
∑JB pesan_3	0	0.57	1.08	0.58	**5.67**
∑JB pesan_4	0	0.47	0.77	**4.00**	0
∑JB pesan_5	1.12	**4.83**	0.77	1.25	0



Gambar 2. Hasil Pengujian

5. KESIMPULAN

Penggunaan SMS terus meningkat setiap saat seiring dengan perkembangan operator seluler saat ini. Peluang ini dimanfaatkan oleh pihak-pihak tertentu untuk melakukan pengiriman pesan yang dapat merugikan penerima yang disebut dengan SMS Spam. SMS Spam tersebut dapat dikategorikan atas *advertising* (iklan), *smishing* (manipulasi), *remittance* (pengiriman uang), *scamming* (premium call), dan *forecasting* (ramalan).

Dengan menggunakan algorithma Native Bayesian dan pembobotan dapat diketahui kata-kata dan pembobotan kata dari sebuah kategori SMS Spam. Sehingga dengan mencari jumlah bobot kata-kata yang terkandung didalam sebuah pesan yang diterima akan dapat diketahui kategori apa dari SMS Spam tersebut.

DAFTAR PUSTAKA

[1] Donegan, Patrick, *The True Cost of SMS Spam*, www.heavyrreading.com, 2013.

[2] GSM Doc 28/85, *Services and Facilities to be provided in the GSM System* , rev2, June 1985.

[3] Hamzah, Amir, *Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis"*, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, Yogyakarta, 2012.

[4] http://en.wikipedia.org/wiki/SMS.

[5] Nirmala Ganapathy, *3,000 SMS a Month Limit in India From Today*, Straits Times Indonesia, September 27, 2011.

[6] Sebastiani, F., *Machine learning in automated text categorization*, ACM Computing Surveys, 2002, 34(1):1-47.

[7] Silver, Katie, *OMG: Text messaging turns 19 this week ... and this is the Brit we have to thank for our sore thumbs*, Daily Mail (London), December 7, 2011.

[8] SMS types on routomessaging.com.

[9] Tomi T. Ahonen, *Time to Confirm some Mobile User Numbers: SMS, MMS, Mobile Internet, M-News*, Blog, January 13, 2011, Retrieved September 16, 2013.

[10] William Enck, Patrick Traynor, Patrick McDaniel, and Thomas La Porta, Exploiting *Open Functionality in SMS-Capable Cellular Networks*, Department of Computer Science and Engineering The Pennsylvania State University, University Park, 2008.

[11] *Zoll, BKA und Verfassungsschutz verschickten 2010 über 440.000 stille SMS*, heise online, Heise.de.