

bley

intelligentes Greylisting ohne Verzögerung

Evgeni Golov

Institut für Informatik
Rechnernetze und Kommunikationssysteme
Heinrich-Heine-Universität Düsseldorf

13.11.2010

Was ist Spam?



Was ist Spam?

- Unerwünschte E-Mails

Was ist Spam?

- Unerwünschte E-Mails
- Meist mit Werbung für Produkte und Dienstleistungen

Was ist Spam?

- Unerwünschte E-Mails
- Meist mit Werbung für Produkte und Dienstleistungen
- Versand ohne Erlaubnis

SMTP

- Simple Mail Transfer Protocol (RFC 821, aktuell RFC 5321)

SMTP

- Simple Mail Transfer Protocol (RFC 821, aktuell RFC 5321)
- Textbasiertes Protokoll zur Übertragung von E-Mails

SMTP

- Simple Mail Transfer Protocol (RFC 821, aktuell RFC 5321)
- Textbasiertes Protokoll zur Übertragung von E-Mails
- Keinerlei Authentifizierung

SMTP

- Simple Mail Transfer Protocol (RFC 821, aktuell RFC 5321)
- Textbasiertes Protokoll zur Übertragung von E-Mails
- Keinerlei Authentifizierung
- “gesprochen” zwischen dem sendenden und empfangenden Mail-Server

SMTP Beispiel

alice@a.example.org → bob@b.example.org

Empfänger: 220 b.example.org running some SMTP daemon

Sender: **HELO** a.example.org

Empfänger: 250 hello a.example.org

Sender: **MAIL FROM:** <alice@a.example.org>

Empfänger: 250 <alice@a.example.org> ... Sender ok

Sender: **RCPT TO:** <bob@b.example.org>

Empfänger: 250 <bob@b.example.org> ... Recipient ok

Sender: **DATA**

Empfänger: 354 End data with <CR><LF>.<CR><LF>

Sender: Mail-Body

Sender: .

Empfänger: 250 Ok

Sender: **QUIT**

Empfänger: 221 Bye

Wie wird Spam verschickt?

- Zustellung ganz normal via SMTP

Wie wird Spam verschickt?

- Zustellung ganz normal via SMTP
- Großteil über Botnets (87.9%) ¹

¹MessageLabs Intelligence: Q3/September 2009

Wie wird Spam verschickt?

- Zustellung ganz normal via SMTP
- Großteil über Botnets (87.9%) ¹
- Botnets bestehen aus infizierten Computern

¹MessageLabs Intelligence: Q3/September 2009

Wie wird Spam verschickt?

- Zustellung ganz normal via SMTP
- Großteil über Botnets (87.9%) ¹
- Botnets bestehen aus infizierten Computern
- Aber auch über eigene Server und online Kontakt-Formulare

¹MessageLabs Intelligence: Q3/September 2009

Wie wird Spam verschickt?

- Zustellung ganz normal via SMTP
- Großteil über Botnets (87.9%) ¹
- Botnets bestehen aus infizierten Computern
- Aber auch über eigene Server und online Kontakt-Formulare
- Versand meist nicht RFC-konform

¹ MessageLabs Intelligence: Q3/September 2009

Wie wird Spam gefiltert?

- Pre-MX Spam-Filter
 - **vor** der Annahme der E-Mail durch den Mail-Server
 - Analysieren nur IP-Adresse, Absender und Empfänger
 - Schnell und ressourcenschonend
 - Allerdings auch fehleranfällig

Wie wird Spam gefiltert?

- Pre-MX Spam-Filter
 - **vor** der Annahme der E-Mail durch den Mail-Server
 - Analysieren nur IP-Adresse, Absender und Empfänger
 - Schnell und ressourcenschonend
 - Allerdings auch fehleranfällig
- Post-MX Spam-Filter
 - **nach** der Annahme der E-Mail durch den Mail-Server
 - Analysieren die ganze E-Mail
 - Langsam und ressourcenhungrig (da viele Informationen zu verarbeiten sind)
 - Sehr genau
 - Hier nicht weiter besprochen

Blacklisting

- Datenbanken von bekannten Spam-Versendern (oft DNSBLs)

Blacklisting

- Datenbanken von bekannten Spam-Versendern (oft DNSBLs)
- Meist anhand der IP-Adresse identifiziert

Blacklisting

- Datenbanken von bekannten Spam-Versendern (oft DNSBLs)
- Meist anhand der IP-Adresse identifiziert
- Wird die IP-Adresse des Senders in einer Blacklist gefunden, wird die Annahme verweigert

Blacklisting

- Datenbanken von bekannten Spam-Versendern (oft DNSBLs)
- Meist anhand der IP-Adresse identifiziert
- Wird die IP-Adresse des Senders in einer Blacklist gefunden, wird die Annahme verweigert
- Sehr effektiv

Blacklisting

- Datenbanken von bekannten Spam-Versendern (oft DNSBLs)
- Meist anhand der IP-Adresse identifiziert
- Wird die IP-Adresse des Senders in einer Blacklist gefunden, wird die Annahme verweigert
- Sehr effektiv
- Leider landen auch Unschuldige in solchen Listen und ihre E-Mails gehen verloren

Whitelisting

- Technisch genauso aufgebaut wie Blacklisten

Whitelisting

- Technisch genauso aufgebaut wie Blacklisten
- Jedoch wird Sendern vertraut und die E-Mail angenommen

Whitelisting

- Technisch genauso aufgebaut wie Blacklisten
- Jedoch wird Sendern vertraut und die E-Mail angenommen
- In geschlossenen Benutzer-Gruppen kann man alle E-Mails die nicht von einem vertrauenswürdigen Sender stammen ablehnen

Whitelisting

- Technisch genauso aufgebaut wie Blacklisten
- Jedoch wird Sendern vertraut und die E-Mail angenommen
- In geschlossenen Benutzer-Gruppen kann man alle E-Mails die nicht von einem vertrauenswürdigen Sender stammen ablehnen
- Im Normalfall ist dies nicht praktikabel

Whitelisting

- Technisch genauso aufgebaut wie Blacklisten
- Jedoch wird Sendern vertraut und die E-Mail angenommen
- In geschlossenen Benutzer-Gruppen kann man alle E-Mails die nicht von einem vertrauenswürdigen Sender stammen ablehnen
- Im Normalfall ist dies nicht praktikabel
- Deswegen meist als Absicherung gegen falsche Blacklist Einträge verwendet

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt
- Beim zweiten wird die E-Mail angenommen

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt
- Beim zweiten wird die E-Mail angenommen
- Je nach Konfiguration dauert dies zwischen 5 und 60 Minuten

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt
- Beim zweiten wird die E-Mail angenommen
- Je nach Konfiguration dauert dies zwischen 5 und 60 Minuten
- Effektiv, da Spammer ungeduldig sind und häufig keinen zweiten Zustellversuch unternehmen

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt
- Beim zweiten wird die E-Mail angenommen
- Je nach Konfiguration dauert dies zwischen 5 und 60 Minuten
- Effektiv, da Spammer ungeduldig sind und häufig keinen zweiten Zustellversuch unternehmen
- E-Mails können theoretisch nicht verloren gehen

Greylisting

- Erster Zustellversuch wird grundsätzlich abgelehnt
- Beim zweiten wird die E-Mail angenommen
- Je nach Konfiguration dauert dies zwischen 5 und 60 Minuten
- Effektiv, da Spammer ungeduldig sind und häufig keinen zweiten Zustellversuch unternehmen
- E-Mails können theoretisch nicht verloren gehen
- Allerdings muss man u.U. lange auf eine wichtige E-Mail warten

SPF

- Sender Policy Framework (RFC 4408)

SPF

- Sender Policy Framework (RFC 4408)
- Definiert welche Server E-Mails mit einer bestimmten Domain als Absender verschicken dürfen

SPF

- Sender Policy Framework (RFC 4408)
- Definiert welche Server E-Mails mit einer bestimmten Domain als Absender verschicken dürfen
- Wenn ein Server nicht unter der Domain senden darf, ist es ein Anzeichen für Spam

SPF

- Sender Policy Framework (RFC 4408)
- Definiert welche Server E-Mails mit einer bestimmten Domain als Absender verschicken dürfen
- Wenn ein Server nicht unter der Domain senden darf, ist es ein Anzeichen für Spam
- Problematisch bei Weiterleitungen und Verteilern

SPF

- Sender Policy Framework (RFC 4408)
- Definiert welche Server E-Mails mit einer bestimmten Domain als Absender verschicken dürfen
- Wenn ein Server nicht unter der Domain senden darf, ist es ein Anzeichen für Spam
- Problematisch bei Weiterleitungen und Verteilern
- Beispiel uni-duesseldorf.de:

```
"v=spf1 ip4:134.99.128.0/17  
ip4:193.159.219.219/32 ip4:134.99.34.76/32  
mx a:mail.rz.uni-duesseldorf.de ~all"
```

Probleme

Die bisherigen Verfahren

Probleme

Die bisherigen Verfahren

- verlieren entweder wichtige E-Mails

Probleme

Die bisherigen Verfahren

- verlieren entweder wichtige E-Mails
- oder verzögern diese

Probleme

Die bisherigen Verfahren

- verlieren entweder wichtige E-Mails
- oder verzögern diese
- beides ist unerwünscht

Lösung: Intelligent Greylisting

- Selektives Greylisting

Lösung: Intelligent Greylisting

- Selektives Greylisting
- Greylisting nur wenn der Sender sich wie ein Spammer verhält, basierend auf
 - Treffern in Blacklisten
 - negativen SPF Resultaten
 - diversen (RFC) Tests
 - und der Annahme, dass man von dynamischen IP-Adressen normalerweise keine E-Mails verschickt

Implementierung

- Python mit PostgreSQL/MySQL/SQLite3 als Datenbank

²Eigenimplementation mit Hilfe von Twisted

Implementierung

- Python mit PostgreSQL/MySQL/SQLite3 als Datenbank
- DNSWL und DNSBL Tests mit `twisted-names`

²Eigenimplementation mit Hilfe von Twisted

Implementierung

- Python mit PostgreSQL/MySQL/SQLite3 als Datenbank
- DNSWL und DNSBL Tests mit `twisted-names`
- HELO und DynIP Tests

²Eigenimplementation mit Hilfe von Twisted

Implementierung

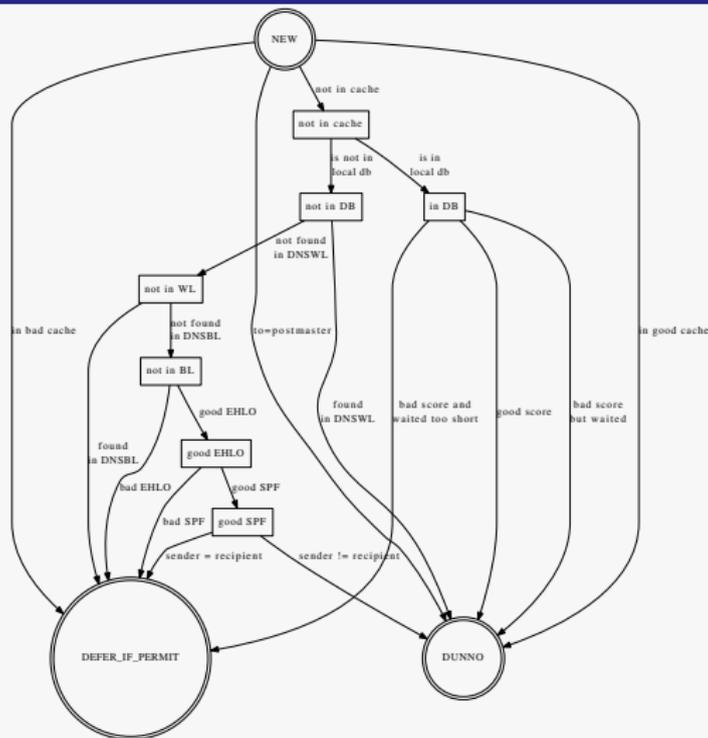
- Python mit PostgreSQL/MySQL/SQLite3 als Datenbank
- DNSWL und DNSBL Tests mit `twisted-names`
- HELO und DynIP Tests
- SPF Tests mit `pyspf`

²Eigenimplementation mit Hilfe von Twisted

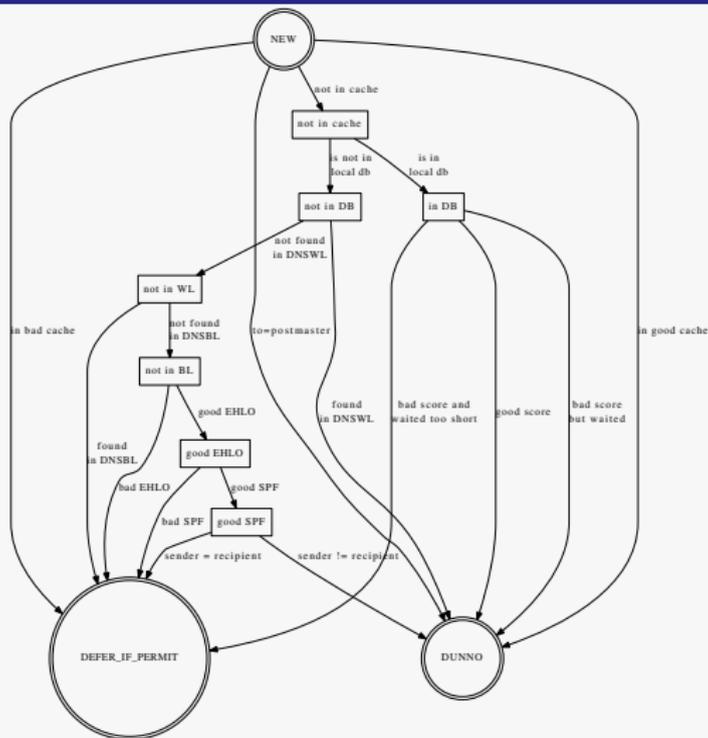
Implementierung

- Python mit PostgreSQL/MySQL/SQLite3 als Datenbank
- DNSWL und DNSBL Tests mit `twisted-names`
- HELO und DynIP Tests
- SPF Tests mit `pyspf`
- Kommunikation mit Postfix via `PostfixPolicy`²

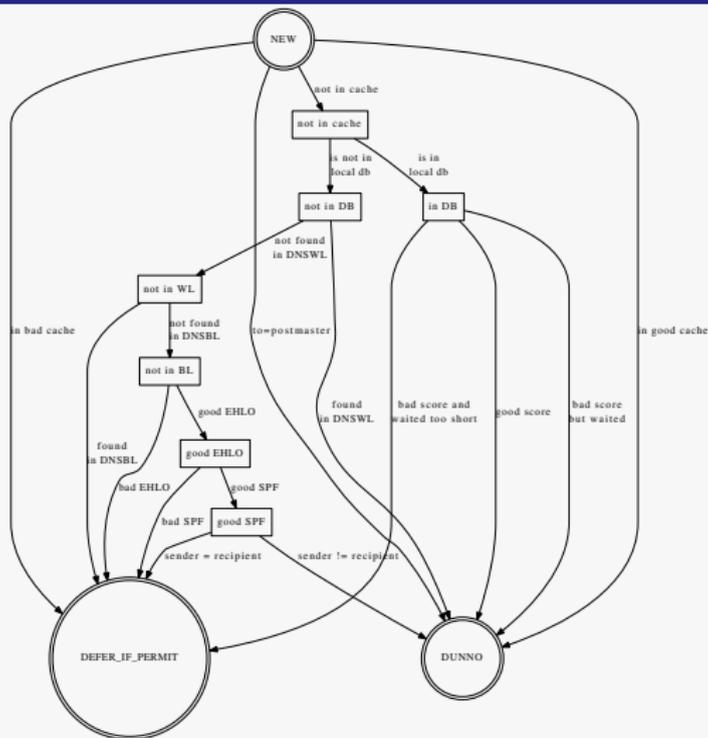
²Eigenimplementation mit Hilfe von Twisted



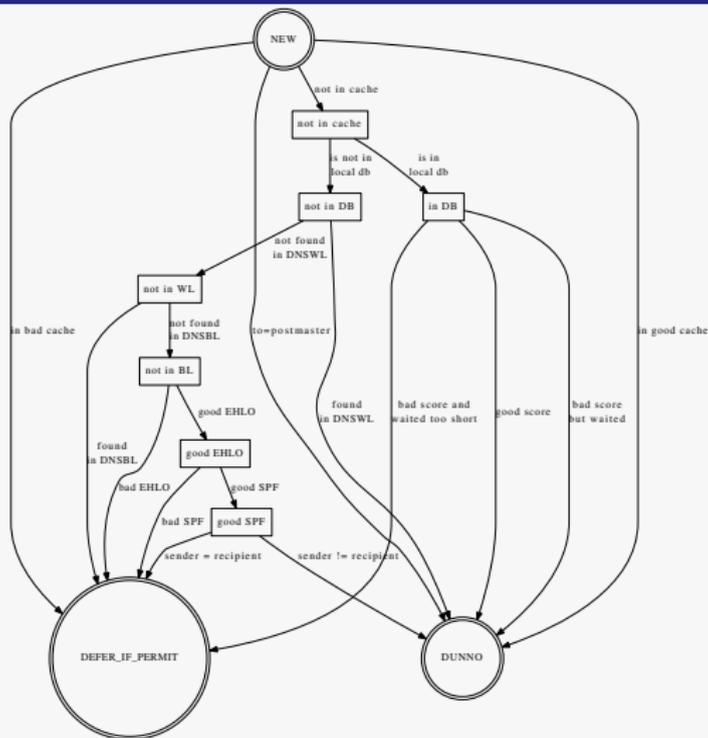
Datenbank speichert Tupel (IP-Adresse, Sender, Empfänger),
Zeitstempel und Aktion



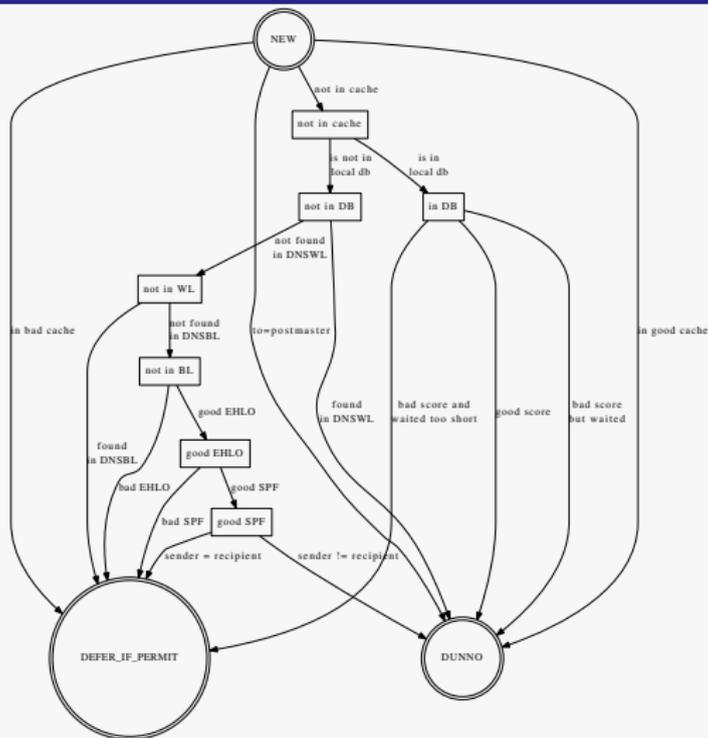
DNSWL/DNSBL Checks gegen konfigurierte Listen



EHLO muss korrekter, auflösbarer FQDN sein



SPF muss "Neutral" oder besser sein



Sender sollte nicht gleich Empfänger sein

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails
- `policyd-weight` (gewichtetes Blacklisting) weist 97.5% davon als Spam ab

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails
- `policyd-weight` (gewichtetes Blacklisting) weist 97.5% davon als Spam ab
- `bley` (intelligentes Greylisting) weist “nur” 97% ab

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails
- `policyd-weight` (gewichtetes Blacklisting) weist 97.5% davon als Spam ab
- `bley` (intelligentes Greylisting) weist “nur” 97% ab
- Dabei gehen aber keine E-Mails verloren

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails
- `policyd-weight` (gewichtetes Blacklisting) weist 97.5% davon als Spam ab
- `bley` (intelligentes Greylisting) weist “nur” 97% ab
- Dabei gehen aber keine E-Mails verloren
- Sie werden höchstens verzögert

Evaluation

- `mail.die-welt.net` empfängt täglich ca 18.000 E-Mails
- `policyd-weight` (gewichtetes Blacklisting) weist 97.5% davon als Spam ab
- `bley` (intelligentes Greylisting) weist “nur” 97% ab
- Dabei gehen aber keine E-Mails verloren
- Sie werden höchstens verzögert
- Nachfolgend ein paar Zahlen aus dem Zeitraum 01.06.2010-31.07.2010 (1.100.000 Mails, 18.000/Tag)

Evaluation – Abgelehnte E-Mails

%	Grund für Ablehnung
48.00%	Sender in DNSBL
38.38%	Cache Hit
6.58%	Sender benutzt ein nicht RFC-konformes HELO
3.16%	Greylisting aktiv, nicht lang genug gewartet
0.73%	Sender und Empfänger waren identisch
0.22%	IP-Adresse wurde als dynamisch erkannt
0.06%	SPF Test negativ

Sofern Greylisting noch nicht aktiv, wird dies nach der Ablehnung gestartet.

Die restlichen 2.87% wurden angenommen (siehe nächste Folie).

Evaluation – Angenommene E-Mails

%	Grund für Annahme
1.52%	Cache Hit
0.58%	Sender unbekannt, aber keiner der Tests war negativ
0.45%	Sender bekannt, hat sich zuvor schon korrekt verhalten
0.18%	Sender unbekannt, aber in DNSWL
0.14%	Greylisting aktiv, Sender hat genug gewartet

Die restlichen 97.13% wurden abgewiesen (siehe letzte Folie).

Evaluation – Zusammenfassung

- Erkennungsrate fast identisch mit `policyd-weight`

Evaluation – Zusammenfassung

- Erkennungsrate fast identisch mit `policyd-weight`
- Es können aber keine E-Mails verloren gehen

Evaluation – Zusammenfassung

- Erkennungsrate fast identisch mit `policyd-weight`
- Es können aber keine E-Mails verloren gehen
- Laut `SpamAssassin` erreichen pro Tag ca 20-30 Spams die Mailboxen

Evaluation – Zusammenfassung

- Erkennungsrate fast identisch mit `policyd-weight`
- Es können aber keine E-Mails verloren gehen
- Laut `SpamAssassin` erreichen pro Tag ca 20-30 Spams die Mailboxen
- Weitere 20-30 werden nicht von `SpamAssassin` erkannt, sind aber auch Spam

Evaluation – Zusammenfassung

- Erkennungsrate fast identisch mit `policyd-weight`
- Es können aber keine E-Mails verloren gehen
- Laut `SpamAssassin` erreichen pro Tag ca 20-30 Spams die Mailboxen
- Weitere 20-30 werden nicht von `SpamAssassin` erkannt, sind aber auch Spam
- Nach weiterer Analyse: nur 7 gutartige E-Mails im Zeitraum verzögert (meist wegen falschem EHLO)

Fazit

- `bley` ermöglicht effizientes Filtern von Spam

Fazit

- `bley` ermöglicht effizientes Filtern von Spam
- `bley` verzögert, sofern der Sender sich korrekt verhält, keine Mails

Fazit

- `bley` ermöglicht effizientes Filtern von Spam
- `bley` verzögert, sofern der Sender sich korrekt verhält, keine Mails
- `SpamAssassin` ist weiterhin nützlich, muss aber nur noch einen Bruchteil der Mails analysieren

Danke!

Vielen Dank für die
Aufmerksamkeit!

Fragen?

Kontakt

Evgeni Golov

Institut für Informatik
Rechnernetze und Kommunikationssysteme
Heinrich-Heine-Universität Düsseldorf
Universitätsstraße 1
D-40225 Düsseldorf

`evgeni.golov@uni-duesseldorf.de`

`bley.mx`

`www.die-welt.net`

`identi.ca/evgeni`

`twitter.com/zhenech`