



CCERT Report on Spam

Chinese American Networking Symposium,
Nov. 2005

Dr. Quang-Anh Tran
CERNET Computer Emergency Response Team
Email: qa@ccert.edu.cn

Agenda

- Brief introduction of CCERT
- Evolution of spam
- Statistical rules for spam filter
- Suggested framework for spam filter

Part I

Brief Introduction of CCERT

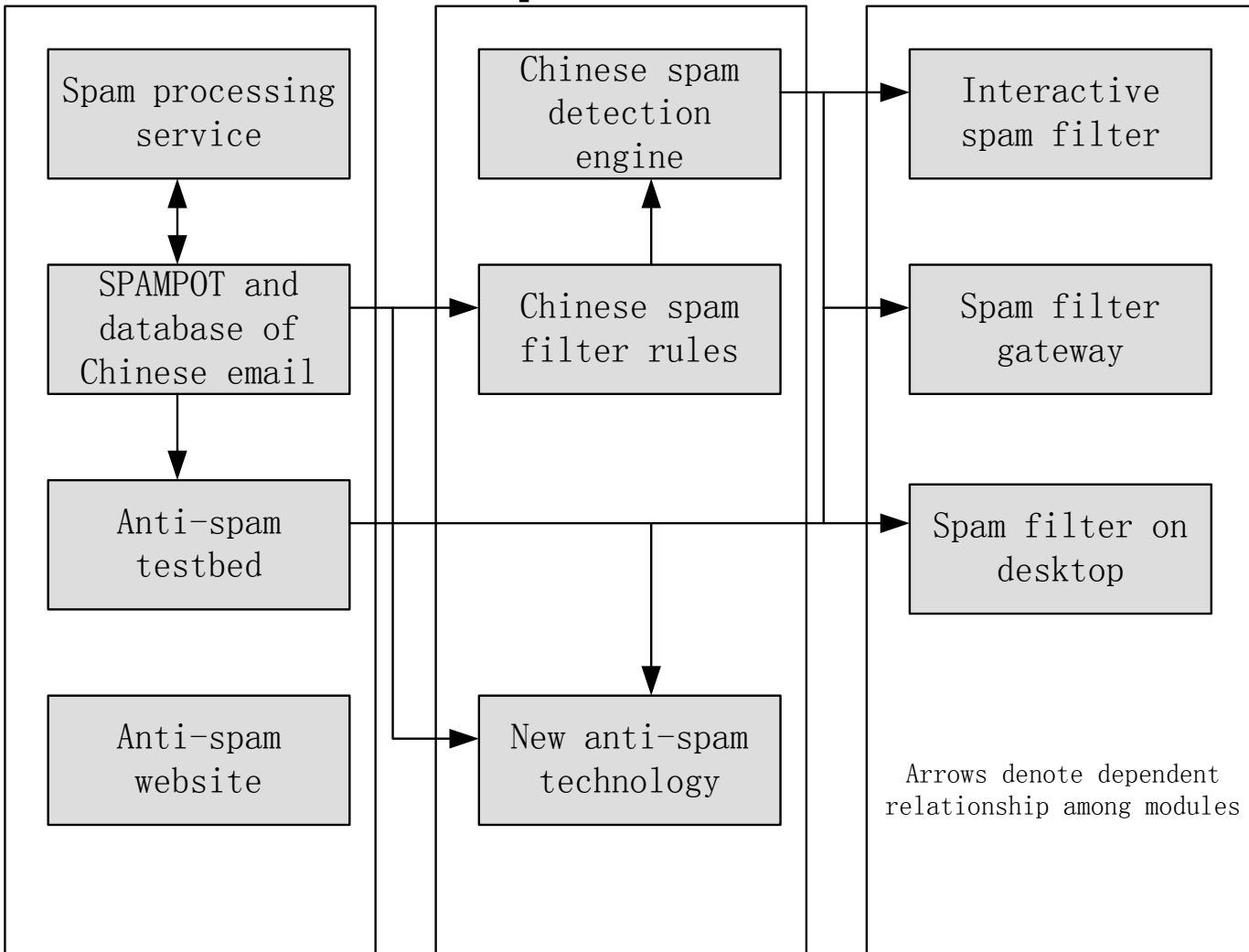
CERNET Computer Emergency Response Team

- Incident response service
 - Spam, attack, abuse handling
 - Virus/Worm control
- Research and development
 - Anti-spam
 - Intrusion detection
 - Virus Worm
 - IPv6 Security
 - Access Control
- Security training
- More details
 - Website: <http://www.ccert.edu.cn>

CCERT anti-spam history

- 1996 Deal with open-relay in China
- 1999 Founded; spam@ccert.edu.cn
- 2002 Establish CERNET anti-spam management rules
- 2003 Manager of technical Workgroup in Internet Association of China
- 2003 Book: Spam and Anti-spam technology
- 2004 Statistical rules for Chinese spam filter
- 2005 Chinese spam database

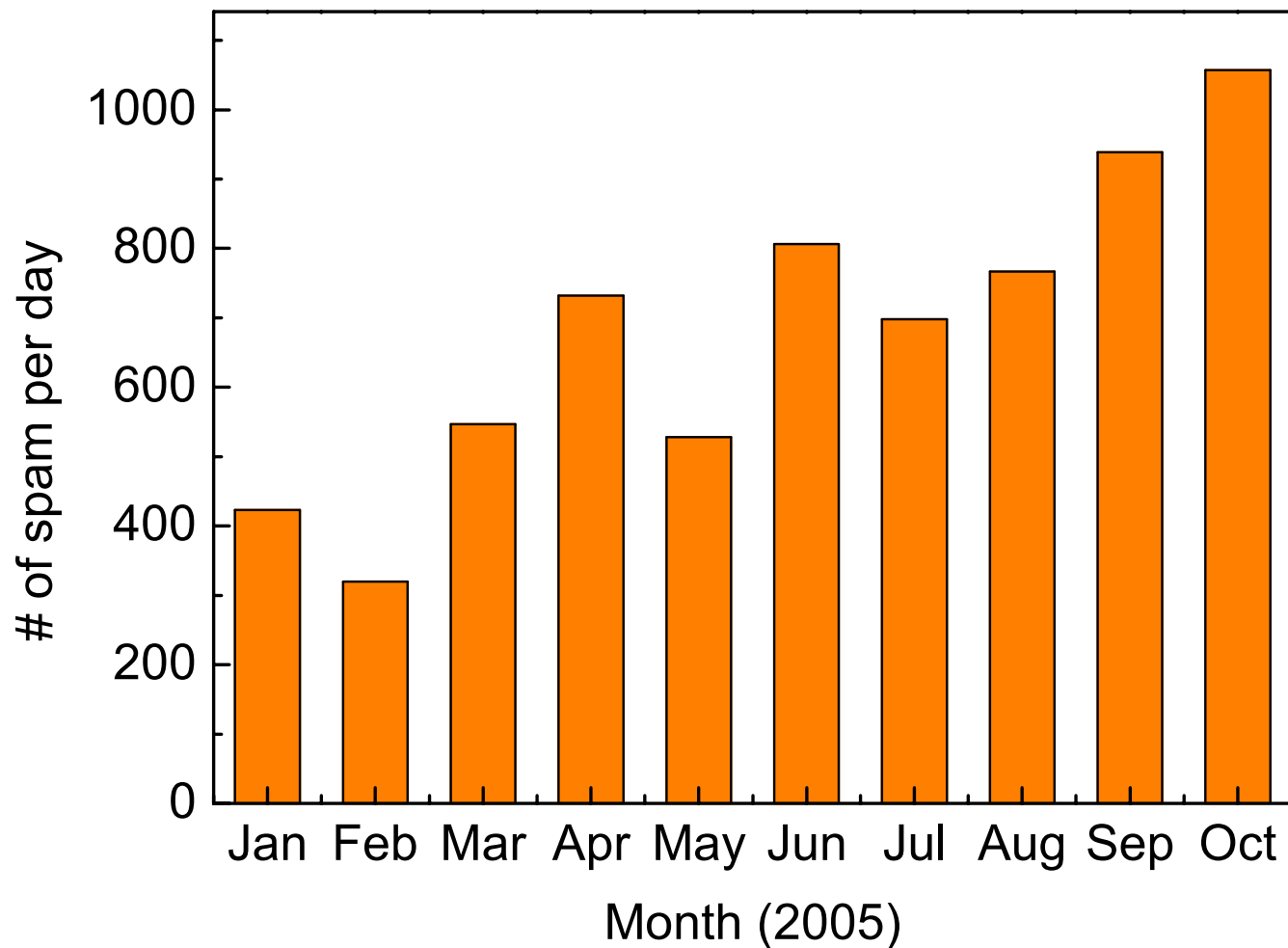
CCERT anti-spam research



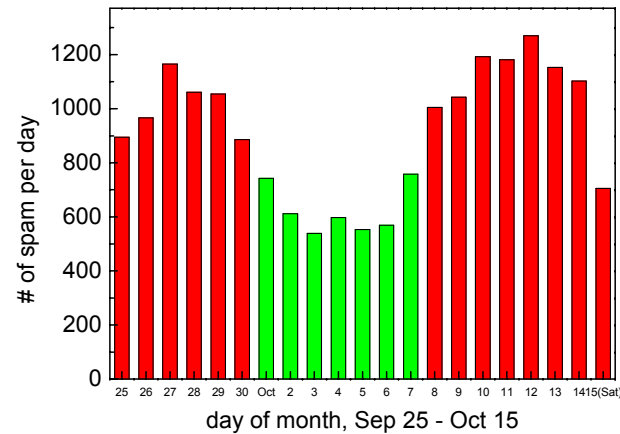
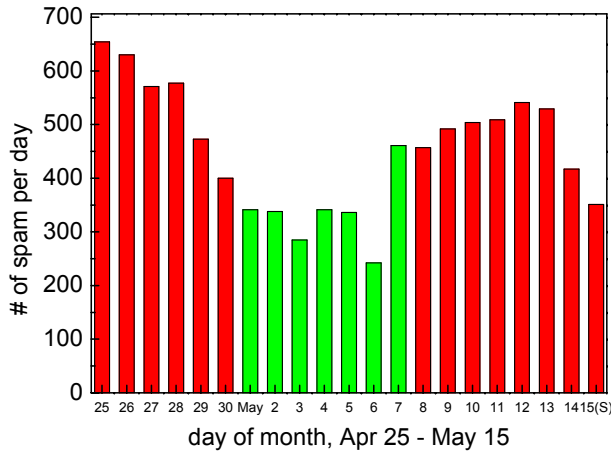
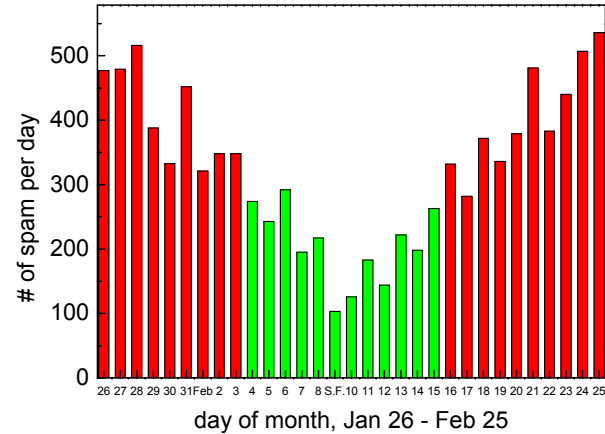
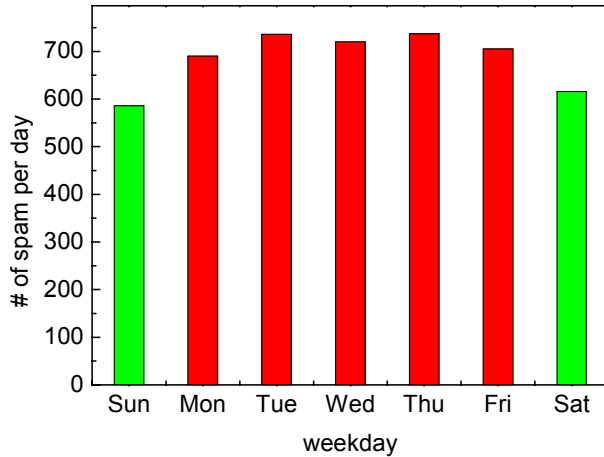
Part II

Evolution of Spam

The growth of Chinese spam

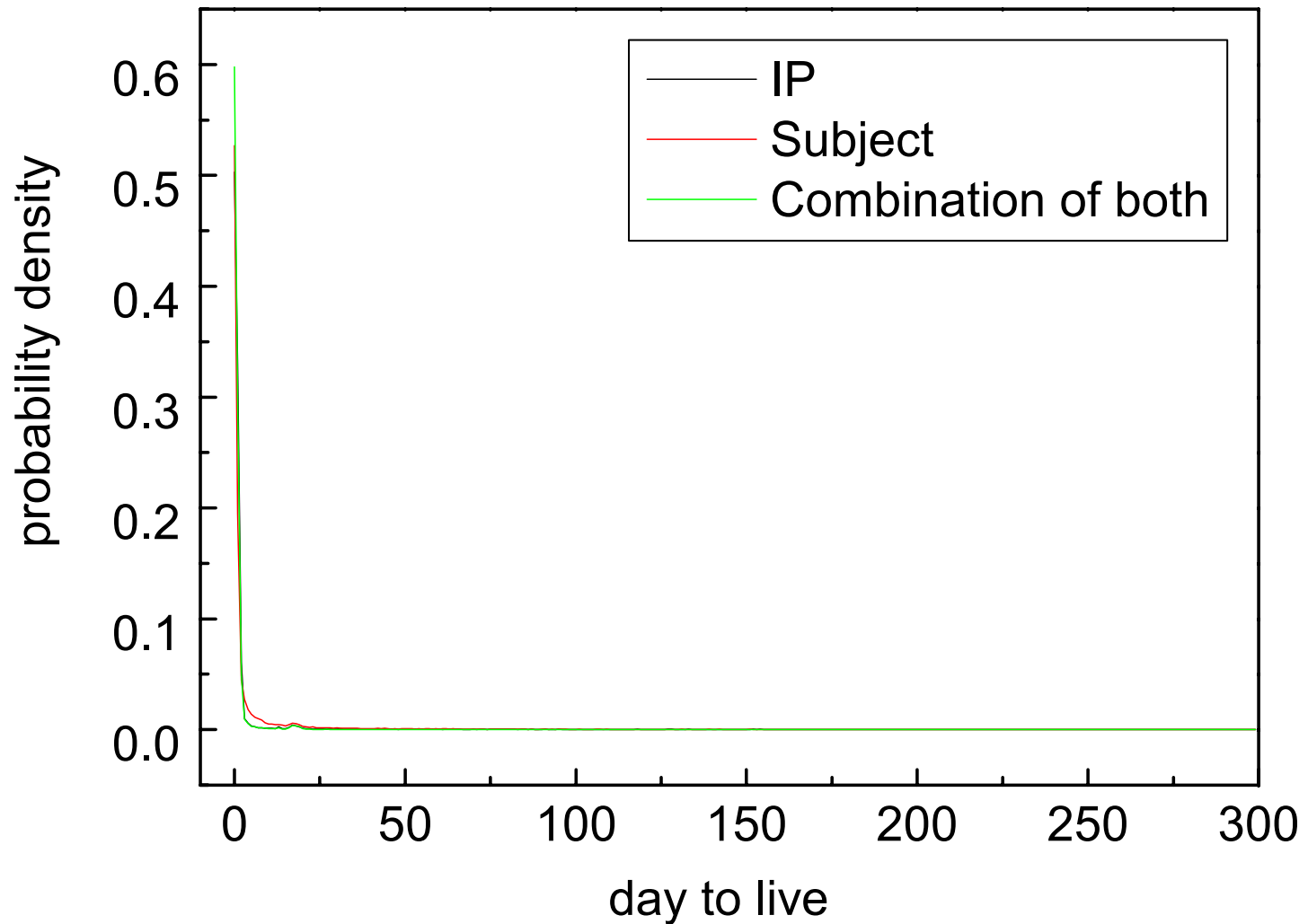


Spam during holiday

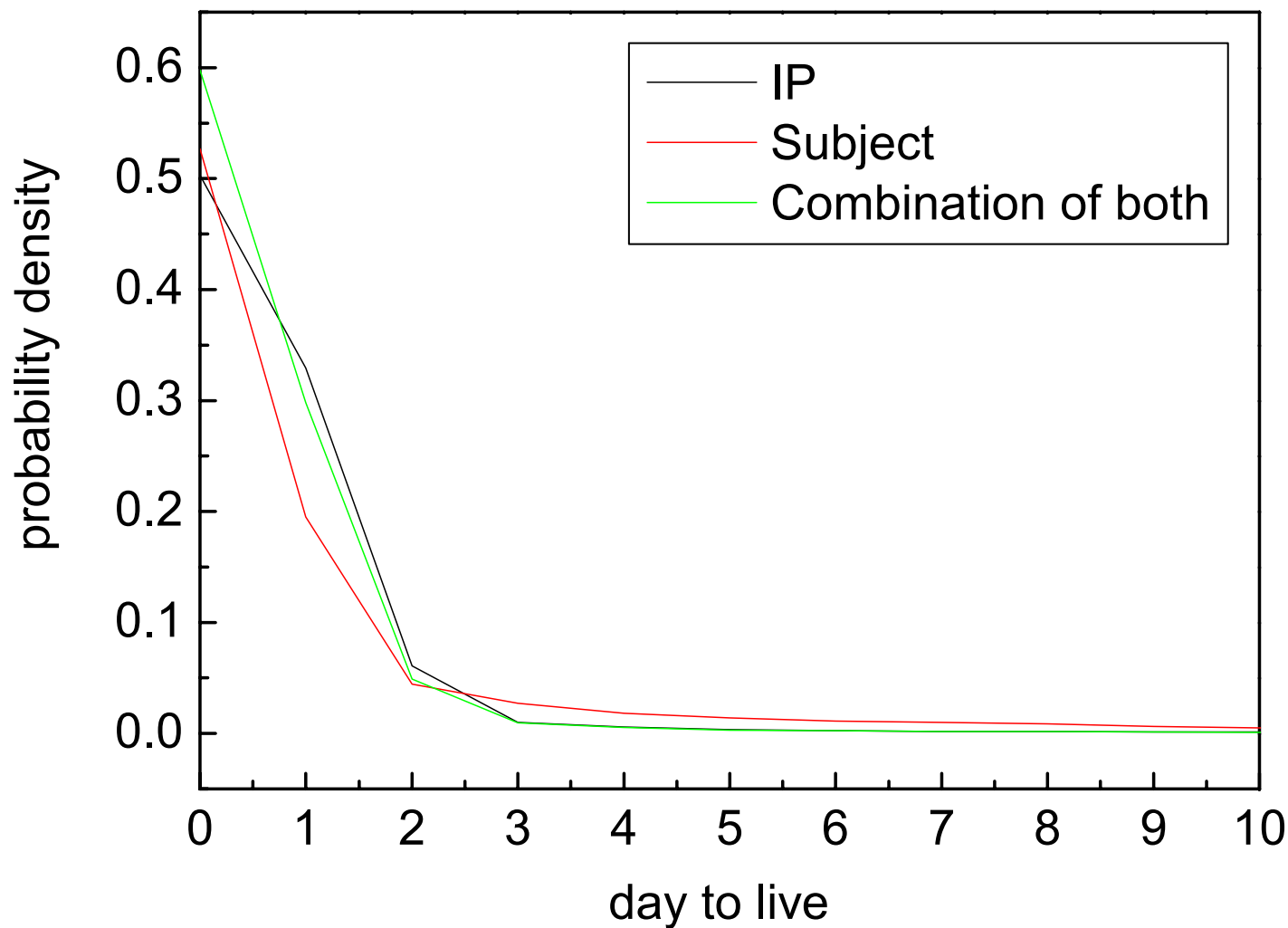


Sasser worm: May 1; CodeRed: Jul; Blaster/Nachi: Aug

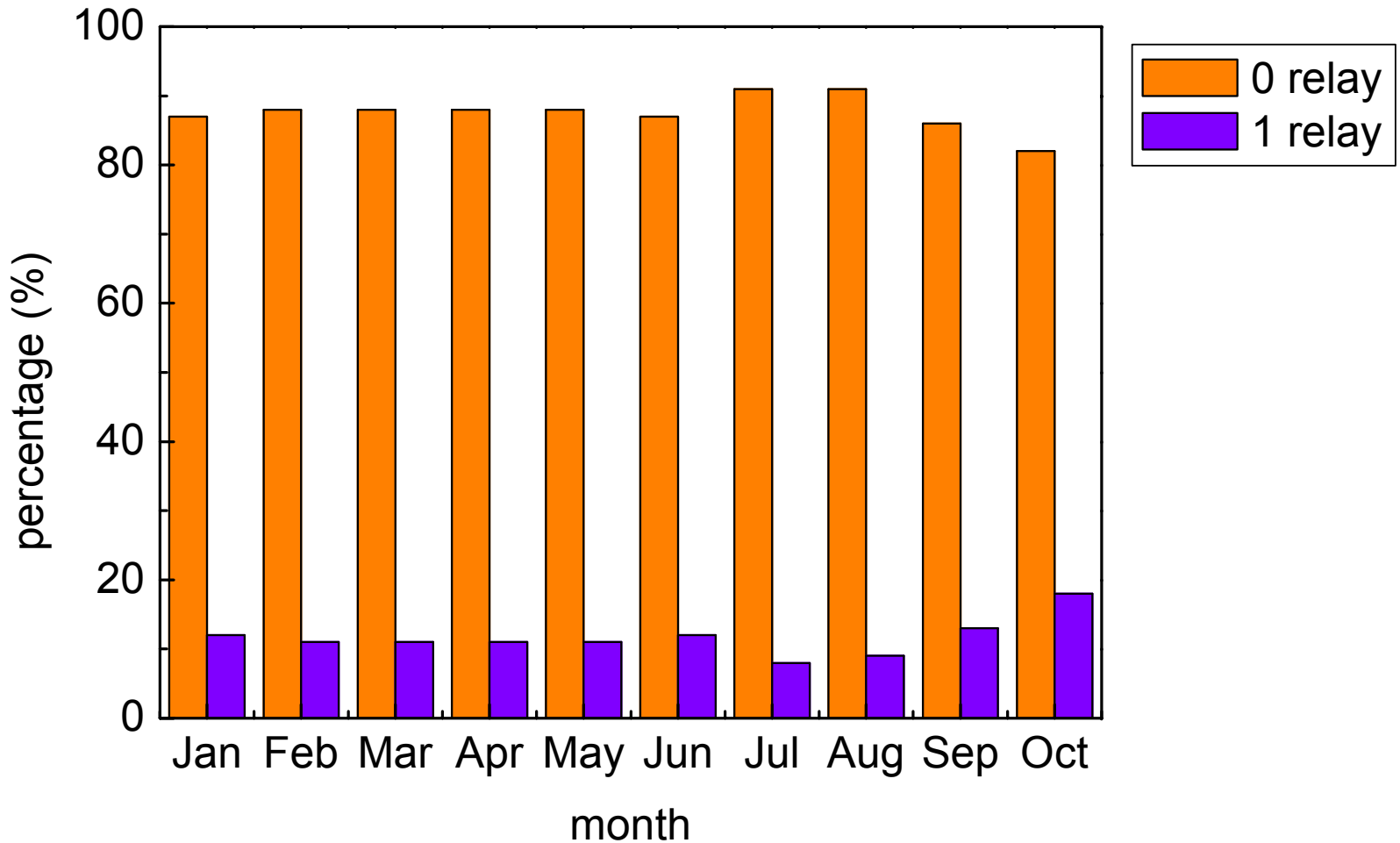
Life time of spam



Live time of spam (cont.)



Spam relay



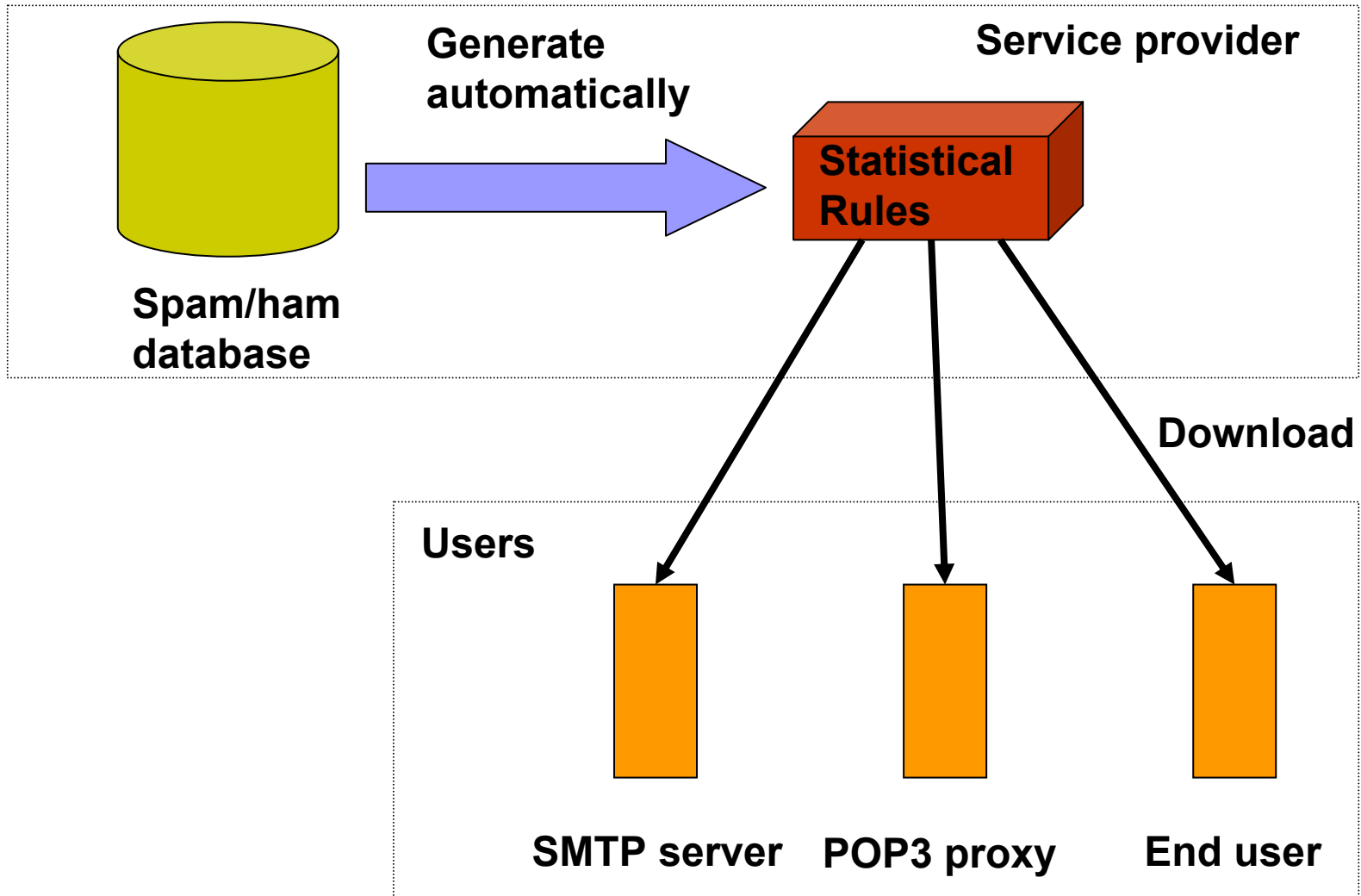
Part III

Statistical rules for spam filter

Statistical rules vs. traditional methods

	Concept	Share ability	Update ability
Rules	Look for spam-liked patterns in an email	YES <i>(popularized quickly)</i>	NO <i>(hard to keep up with the variation of spam)</i>
Statistics	Train the detector upon ham/spam	NO <i>(knowledge obtained from this method is unable to be shared)</i>	YES <i>(possible to make the detector retrained quickly)</i>
Statistical rules	Rules are generated automatically by statistical method	YES	YES

Framework



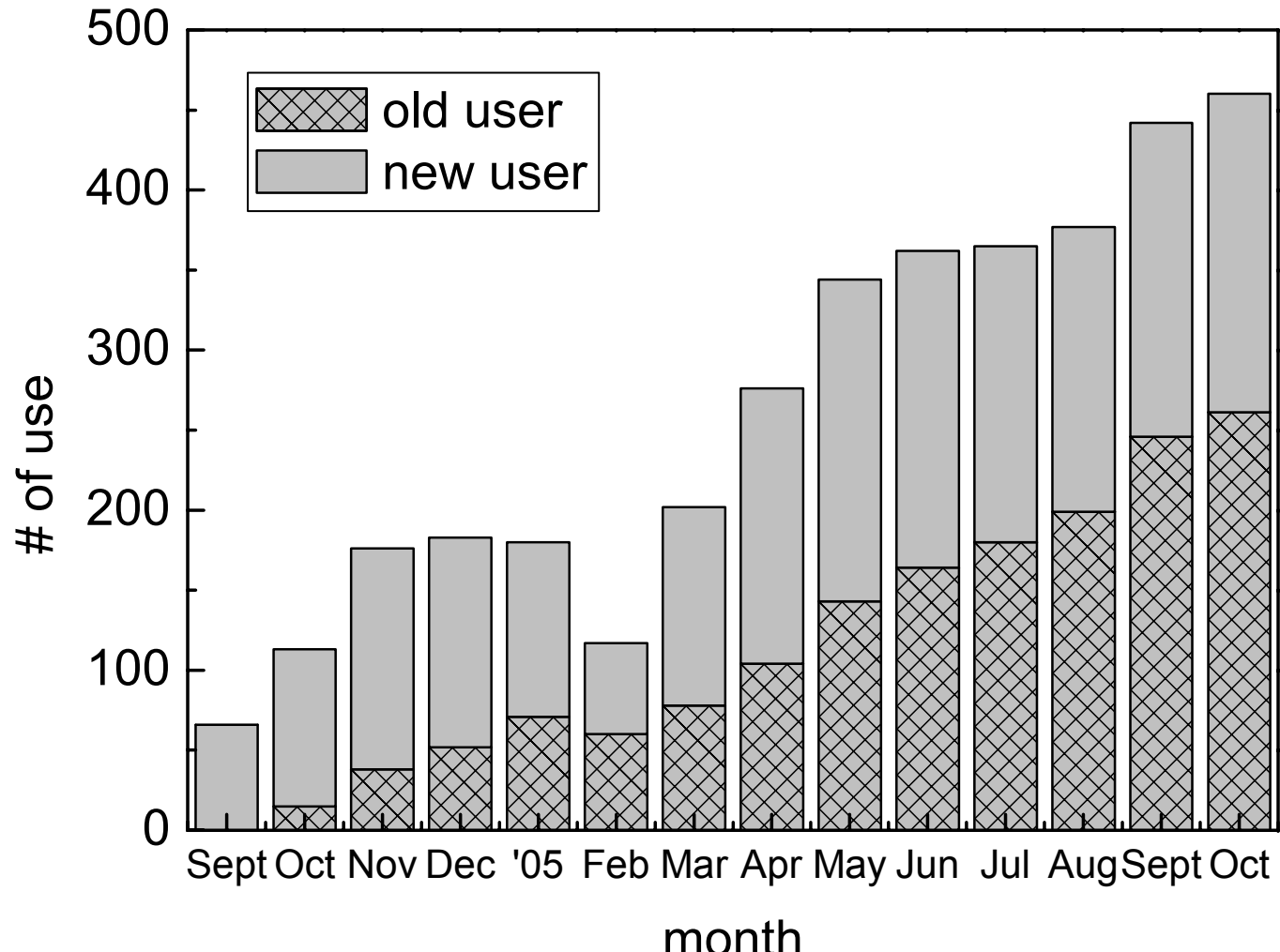
Chinese_rules.cf

- A well-known and widely used statistical rules for SpamAssassin to catch spam written in Chinese (GB2312)
- Website
 - http://www.ccert.edu.cn/spam/sa/Chinese_rules_en.htm

Chinese_rules.cf in progress

- Google search
 - Chinese spam
 - First page
 - 中文垃圾邮件 (“Chinese spam” in Chinese)
 - First result
 - Chinese_rules.cf
 - Nearly 1000 related results

Chinese_rules.cf in progress (cont.)



Theoretical background of Chinese rules

- Pattern retrieval
 - Chinese word segmentation
- Pattern selection
 - Document Frequency
 - Conditional Probabilities and Bayes's Theorem
 - Information Gain
- Score assignment
 - Genetic algorithms
 - Stochastic Gradient Descent

Part IV

Suggested framework for spam filter

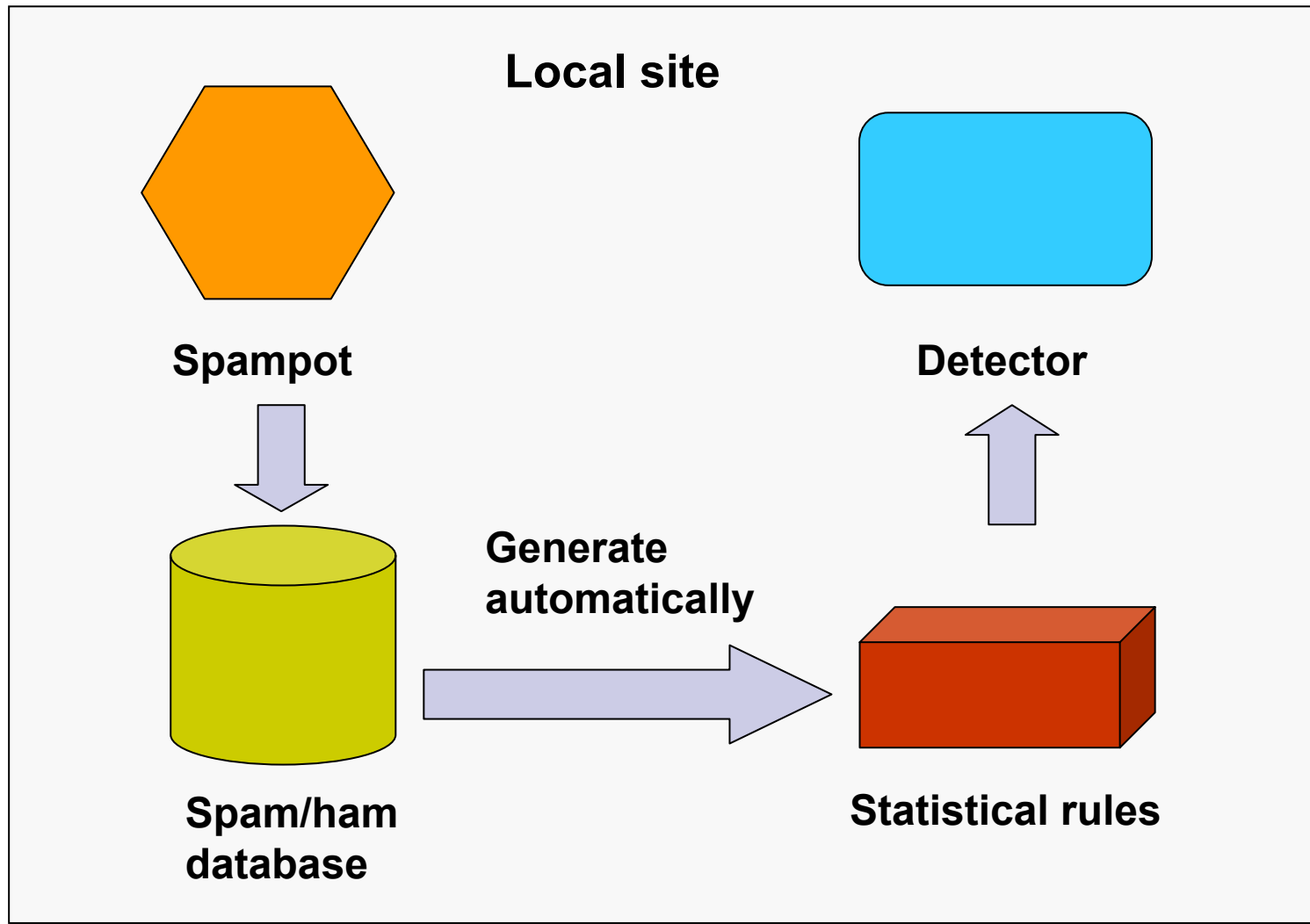
Problem of the present systems

- Spam vary in users and sites
- Dependent on service providers
- Knowledge of spam update slowly

Suggested features

- Local rules
- Statistical rules
- SpamPot (Spam-Honeypot)

Suggested framework



Thank you!