Analyses of Real Email Traffic Properties

Kamil MALINKA[1], Petr HANÁČEK[1], Dan CVRČEK[1]

[1] Fac. of Information Technology, Brno University of Technology, Božetěchova 2, 612 66 Brno, Czech Republic

malinka@fit.vutbr.cz, hanacek@fit.vutbr.cz, cvrcek@fit.vutbr.cz,

Abstract. *In this paper, we perform an empirical analysis of email traffic logs obtained from a large university to better understand its impact on the effectiveness and efficiency of anonymous mix remailers. We analyzed data containing records of almost 790,000 emails sent over a period of forty days – the largest dataset we are aware of. The initial analysis of data is followed by an exploration of how variance in message arrival time and size impact the anonymity and efficiency provided by timed and threshold mixes, respectively. The analysis results are subsequently explored for their possible impact on traffic analysis.*

Keywords

SMTP, email traffic, mixing, anonymity, social networks.

1. Introduction

David Chaum introduced the first mix – an instrument for ensuring unlinkability of sender and messages in 1981 [2]. His initial ideas led to many implementations of anonymity systems in the decades since [1, 5, 8, 9, 16]. These systems, in turn, not only bolstered notions of privacy in information systems but also sparked research in developing attack techniques.

One of the first papers on traffic analysis appeared in 1993 [14], Papers on traffic analysis started appearing regularly from 2000 and there is a very good understanding of security limits of anonymity systems. Computational boundaries for attacks on anonymity systems to be successful have been defined [7, 11, 12, 15] in the last few years.

One great challenge for designers of anonymity systems is to aggregate predictable user behavior so that it appears as random noise. Any deviations from pure randomness can often be exploited to undermine security properties of anonymity systems. There are many assumptions about how user behavior may deviate from randomness, yet precious little is known about how users actually operate. For example, one common "rule-of-thumb" is that X% of the effects come from (1-X)% of the causes – so called Pareto principle observable in many not only social types of phenomena. However, no existing analysis of communication patterns has been used to verify this rule for email communication. In this paper, we aim to provide an empirical basis for properties of email communications relevant to the designers of anonymous systems.

To do so, we have analyzed email traffic logs from a large university. Our central findings are:

- Message inter-arrival times do not follow the expected Poisson distribution; instead, the lognormal distribution fits much better.
- Variance in message arrival time and size impact the anonymous sets for time-based mixes and delays imposed on threshold-based mixes.
- We explore efficiency and anonymity trade-offs when setting message block size (for all mixes) and time windows (for time-based ones).
- Most users balance their messages across recipients, but a minority of users does concentrate their messages to a few recipients.

2. Goals of the Analysis

We obtain anonymized logs from a main SMTP server dispatching emails for four faculties of a huge university. These faculties do not have their own SMTP servers and instead rely on the central university computing services. There are approximately 790000 records (email messages) in our log file. Most of the analysis was carried out on the set of 790000 message records as all these messages would be anonymized and delivered by an anonymity system, unless restrictions on data bandwidth are enforced to decrease computational requirements of the anonymity system.

Analysis targeting user behavior was performed on the subset of messages not marked as spam, which composes about 60% of the whole set. The low level of spam in the dataset is caused by an application of gray-listing techniques preceding the data collection. The spam marking is very effective, and the correctness is such that the number of wrongly classified spam messages is negligible with respect to the total number of messages in the data set.

We hoped to obtain the following information:

- Message arrival times – not only the elementary probability distribution of message time arrivals but also

aggregated information useful for design of anonymity systems, such as number of messages routed within time intervals of certain lengths.

- Number of messages sent by individual users – right now, we can only speculate on the characteristics of 'typical' targets for traffic analysis. It could be someone sending a couple of messages per month or a regular sender of emails. We do, however, expect typical user behavior to be worth identifying.
- Size of email messages – this not only impacts the load of the eventual anonymity system, but also influences how much useful information a passive adversary can obtain through monitoring the anonymity system.

We would like to point to a previous work of Diaz et al. [6] that covered some of the goals listed above. There were, however, several limitations of the results that we believe to have mitigated due to a different approach. We use a general email traffic data and we try to find out what would be the impact of global deployment of anonymity systems, e.g. on SMTP servers. The approach eliminates the problems with unpredictability me others, as mentioned further.

Our approach allows predicting achievable properties of anonymity systems, as it highlights several aspects of email traffic that have not been analyzed yet. Also, it shifts the focus from scenarios with very low traffic to real-world scenarios, where the main problems start appearing in relation to the users' behavior.

3. Analysis

3.1 Message Inter-Arrival Times and Sizes

We first study what we anticipated to be a quite uncontroversial and fully expected result: the distribution of arrival times of email messages. One would expect that arrival time would be best fitted with a Poisson distribution, as has been assumed in the literature when designing mixes [10] as well as attacking them [4]. However, email inter-arrival times from our sample (messages of all users combined together) are not consistent with a Poisson distribution (the same result appears in [6]). Instead, there is great variation in arrival times, including delays up to 45 minutes! The tail of the distribution is much longer than for any Poisson distribution. It is not clear, whether the main reasons for the skewed distribution are the traffic variations or dependencies between email messages.

Using maximum-likelihood estimation, we attempted to fit the arrival times to a number of distributions. The best-performing distribution is a lognormal distribution, a skewed distribution which accounts for longer times. Tab. 1 gives relevant properties of best-fit Poisson and lognormal distributions, along with results of the Kolmogorov-Smirnov test and log-likelihood values. It seems that Weibull (alpha=0.61, beta=5.46) and Gamma (alpha=0.43, beta=19.88) distributions fit the data best. These two passed Kolmogorov-Smirnov tests for levels of significance alpha=0.01 - 0.2.

	Parameters	Kolmogorov-Smirnov D	Log likehood
Poisson	λ=13.1	0.531	2.43×10^7
Lognormal	μ = 1.78, σ = 1.22	0.085	6.88×10^6

Tab. 1. Properties of best-fit Poisson and lognormal distribution.

The difference in log-likelihood values implies that the data sample is infinitely more likely to follow a lognormal distribution than a Poisson distribution. We must note, however, that even the lognormal distribution is not a perfect fit. Fig. 1 plots the cumulative distribution function (CDF) for inter-arrival times along with the fitted lognormal curve[1].

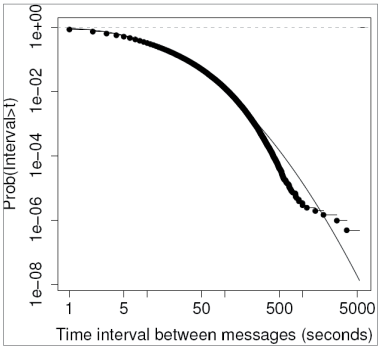

Fig. 1. CDF of inter-arrival times in seconds.

While this result contrasts sharply with existing assumptions made in the anonymity literature, similarly skewed inter-arrival time distributions have been observed, notably TCP packet inter-arrival times on routers [13].

Another important parameter for anonymity systems is the size of message blocks that are anonymized. When a message is larger than the block size, it must be split into parts that are forwarded (and anonymized) separately and the last block is padded with random bytes. This prevents traffic analysis based purely on message sizes going into and coming out of a mix. When the block size is chosen too low, messages will be split into many parts that (a) increase computational requirements and (b) decrease anonymity of users as many blocks will be routed from the same sender towards the same recipient. Oversize blocks require large padding, which unnecessarily increases the amount of data that must be transmitted.

Fig. 2 plots the distribution of message sizes for messages up to 35 KB. Just as for inter-arrival times, there is great disparity. The average size is 35.7 KB, while the

[1] Our last experiments show that Gamma distribution is not rejected by KS test and it fits the data even better.

median size is only 4 KB. However, the largest message is 35 MB. Unfortunately, we were unable to fit commonly used distribution function here.[2]

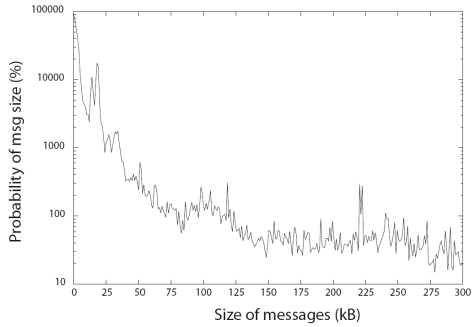

Fig. 2. Distribution of message sizes in KB.

Tab. 2 shows overhead in the terms of number of blocks to number of messages, i.e. the level of anonymity deterioration and communication overhead appropriate to different sizes of blocks. From the table, it appears that the exponential decrease of the number of blocks slows down between 30 kB and 50 kB block sizes.

Block size (KB)	1	10	20	30	40	50	100
Block overhead	53	5.9	3.3	2.5	2.1	1.9	1.4
Data overhead	1.0	1.1	1.3	1.4	1.6	1.8	4.6

Tab. 2. Deterioration of anonymity v communication overhead.

3.2 Time-Based Mixes: Adjusting Constant-Size Time Windows

Timed mixes (e.g., [10]) limit transmission delays by forwarding messages following the expiration of a time window. Hence, the rate of message transmission, along with its variance, affects the length the time window and the anonymity provided. We can explore how different values for the time window impact anonymity. Fig. 3 plots the probability distribution for the number of messages delivered to the SMTP server in time windows of 10, 20, and 60 minutes.

For example, using a 10-minute window makes it very likely that fewer than 200 messages will be mixed, while using a 60-minute window makes 250 to 1000 messages more likely to be mixed. In other words, the graph reveals the probability distribution of the maximum achievable anonymity of users. This graph is again interesting for the fact that probability distribution cannot be represented with any known (to us) widely used distributions with sufficient level of certainty.

[2] The list of distributions we tried contains about dozen of the most frequently used ones

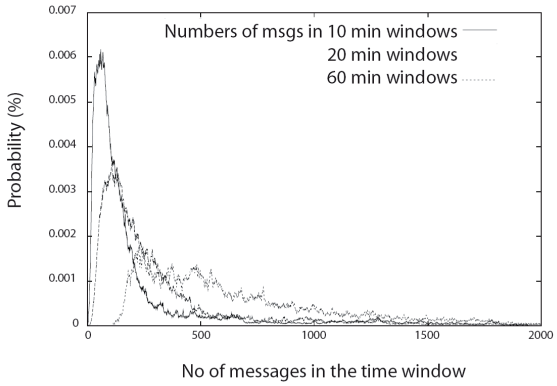

Fig. 3. Number of messages in time windows of 10, 20, and 60 minutes.

This information is important for users -- what would be their anonymity when using the mix if their behavior were alike behavior of average user (number of messages, times and days of sending emails, …).

We mentioned that anonymity systems must normalize the size of messages. This prevents linkability of sender and recipient based on size of messages. Fig. 4 plots variation of anonymity over time during an interval of four days.

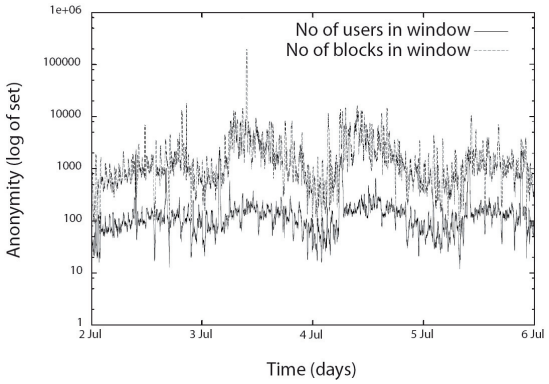

Fig. 4. Real anonymity of messages for incorrectly chosen block size (2 KB).

Note that the y-axis is logarithmic in scale! In addition to the maximum theoretical value of anonymity, the figure also plots the anonymity sets when messages are split into 2KB blocks (an admittedly unwise length). For 2KB blocks, the size of anonymity sets decreases very significantly (more than twenty fold). Anonymity is reduced by one third when expressed in bits, i.e., as the logarithm of anonymity set size.

3.3 Threshold Mixes: Varying Threshold Size to Measure Delay

A threshold mix is an important and popular class of anonymity system. A threshold mix keeps gathering email messages until a certain number of messages is reached. Time is irrelevant here and delivery time is not guaranteed at all. When the threshold is reached, messages are mixed and flushed to the next mix or to recipients.

Variance in message transmission manifests itself as changes in delay times, rather than reductions in anonymity as for timed mixes. Fig. 5 depicts time delays when the threshold is set to 128 (lower plot) or 1024 (upper plot) messages. Average delays are 880 seconds and 7090 seconds for thresholds set to 128 and 1024, respectively, when averaged over number of messages, and 1450 and 10400 seconds when averaged over time. We were pleasantly surprised by relatively short intervals when the delays are very low (well bellow 1000 seconds). These are followed by business hours when the time delays are somewhere between 3000-6000 seconds for the threshold of 1024 messages.

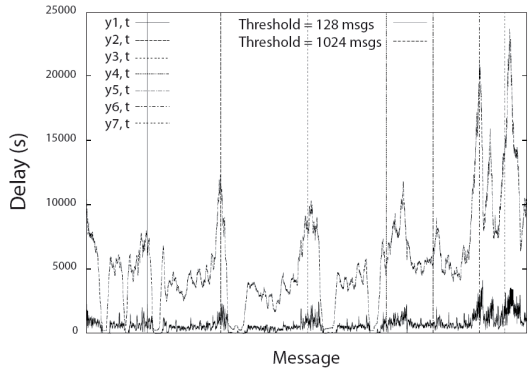

Fig. 5. Delay on a mix with threshold set to 128 or 1024 messages.

Night hours impose the longest delays, but even these are only twice as long as those during business hours. The delays become very long for the two last days -- including incredible delays of over 20 000 seconds (5.5 hours). This is due to the fact that 7th and 8th July were bank holidays followed by a weekend. This graph has distorted time on x-axis as this represents (linearly) number of messages received by the mix. Vertical lines mark midnights.

Fig. 6 gives one more view to delays for threshold mixes: time delays during the course of a day. We are also interested in a similar graph for days of week but as the analyzed set was from the period of forty days, the results could get distorted by single extremes. Fig. 6 shows that while time of day does matter, it does not vary as widely as we were expecting.

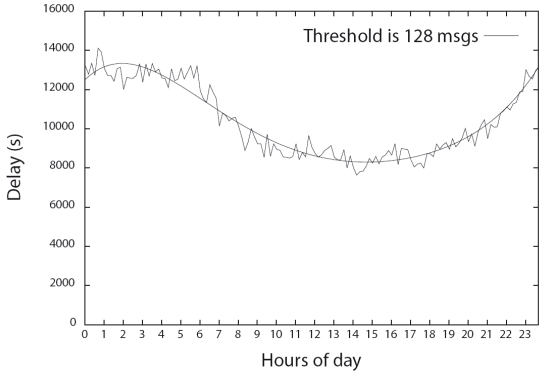

Fig. 6. Delay on a mix with threshold 1024 messages during a day.

3.4 User-Level Distribution of Recipients

Distribution of messages among recipients is the part of the analysis that has brought the biggest surprise so far.

We mentioned in the introduction Pareto principle commonly used in sociology and economics. This rule, if true, would state that e.g. four fifths of each user's messages are addressed to only one fifth of the user's communication partners. This was an expectation we were not able to confirm. What we have found instead, was a linear function mapping the proportion of recipients to the fraction of messages. This is not to say that all users are so balanced in disseminating their communications. While the average user respects linearity, a minority of users do concentrate their messages on a few recipients. Trying a bit harder, we were able to break linearity by carefully selecting a subset of users but we could never approach e.g. 80/20 rule.

Fig. 7 shows the distribution of messages among recipients, selecting only users who send three to five times more messages than their total number of recipients. At the same time, a minimum number of recipients are required. Even such a crafted selection ensures only a 67/33 or 60/40 rule.

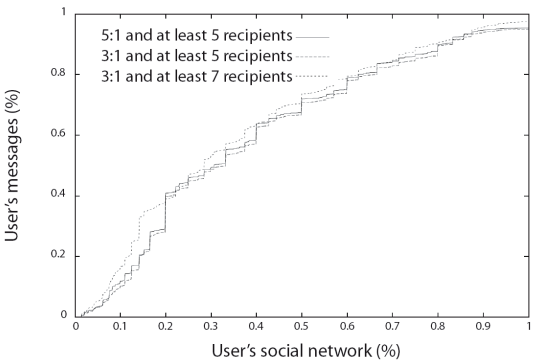

Fig. 7. User-level distribution of messages among recipients.

Notice the slow take off of the graph. This is due to small number of users sending messages to a high number of recipients so the aggregates for less than 10% of recipients are lower than expected.

Fig. 7 presents an aggregate measure of the distribution of messages to recipients, so it would be a misinterpretation to suggest that all users evenly distribute messages among recipients. We more closely examine behavior of particular users in Fig. 8. This graph shows the number of recipients per user for particular x fractions of top messages (x=0.2, 0.3, and 0.4) from Fig. 7. The leftmost line shows fraction of all emails that are addressed to 20% of number of recipients per user for particular fractions of top messages.

The graph contains distribution for users who sent at least 15 messages, which is a small portion (302 to be precise) of all users. One can see that there is only a small fraction (less than one fifth of those 302) of users whose

messages are spread really non-linearly among their recipients. For instance, for the middle 30% graph, around 20% of users distribute their top 30% of messages to at least 50% of their total recipients. Furthermore, only a few users (around 1%) distribute 20% of their messages among 80% of users. This graph demonstrates that while the average distribution of messages remains linear, a small number of users do exhibit the kind of concentrated message patterns which could be exploited by traffic analysis.

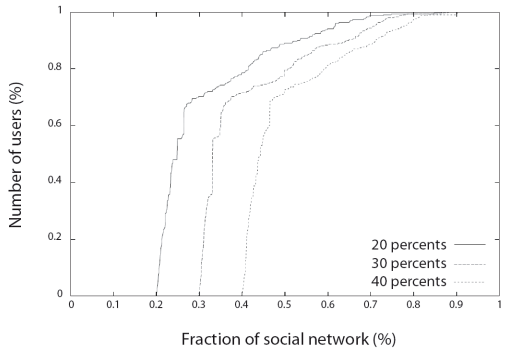

Fig. 8. Analysis of user behavior for addressing 20, 30, and 40 % of emails.

Fig. 8 is affected by rounding errors. When we use exact numbers then 174 users out of 302 from the plot sent exactly one message to each of their recipients (through the university SMTP server). It means that more than 50% of these users would still be immune to any intersection attacks after 40 days of observing all the email traffic on our SMTP server.

4. Impact on Anonymity Systems

We have introduced some interesting results of an email traffic data analysis. Let us now elaborate more on their importance for the design and implementation of anonymity systems.

4.1 Provided Anonymity

Each mix has a theoretical upper bound for the anonymity level provided. The difference between this upper bound and the real anonymity depends on the users' behavior.

The simplest aspect is the sizes of messages being sent through a mix and their comparison to the size of message blocks processed by the mix. We would like to see as few messages being split as possible, while limiting increase in the volume of the transmitted data. The result in section 3.1 shows that the number of blocks per messages decreases exponentially while the volume of the data increases linearly in the mix block size. This allows to increase the mix block size e.g. from 50 kB up to 100 kB with the data "overhead" going up from 80% to 170%.

Kesdogan et al. [10] prove optimality of exponential distribution for delay of messages in Stop-and-Go mixes assuming M/M/∞ queuing system with Poisson distribution of message arrivals. We show that this assumption is not true and as a result, such a mix will not mix the messages perfectly, and the anonymity provided will be lower than expected - the difference is however unclear in the moment.

We used the distribution of message sizes to compute the optimum mix block size, but there is more to be said. We have shown that although the total number of messages larger than a megabyte is very low, messages of up to tens of megabytes happen to appear regularly. Such a message can cause an effective (n-1) attack on the mix. Obviously, an attack may let such messages appear when most convenient.

Another interesting aspect affecting anonymity of users is the distribution of messages among users. The data set covering 40 days of email traffic contains messages to over 102 thousand recipients. However, only 7700 recipients received 10 email messages or more, covering almost 20% of all the traffic.

On the side of senders, only 2% of non-spam users sent at least 10 messages. This number increased to 8% in a subset of internal users (the data set contains all their email traffic). Local static attackers controlling "random" mixes would use the former number, while a dynamic local attacker controlling adaptively chosen mixes - closest to the selected victim - would use the latter one. In both cases, however, a large majority of users would have to sustain a long-term (several months) traffic analysis attack to lose their privacy.

The last finding in this section relates to the behavior of users. We have shown that distribution of recipients is far from the expected Pareto principle. When we analyzed behavior of users who sent more than 500 messages (108 in total), two thirds addressed their messages to only one recipient. The analysis further showed that the distribution of messages according to e.g 80/20 rule is far from reality. This again potentially influences results of statistical traffic analysis attacks.

4.2 Delivery Delay Variation

Timely delivery of messages is important factor from user's perspective, especially if the anonymity technologies should widespread. We show that there is not that much difference in the amount of traffic throughout the day, but the variation is very substantial between work days and weekends (particularly when combined with bank holidays). There seems to be another open question. If it is possible to sacrifice anonymity provided by mixes during low-traffic periods because of different traffic patterns?

4.3 Statistical Disclosure

Statistical disclosure attacks are very simple but also very powerful attacks against privacy technologies. There

are several interesting aspects influencing efficiency of these attacks that we have identified during our analysis.

We have not found, for a large portion of users, any analytical statistical distribution that would describe distribution of messages onto recipients.

- Many users send messages in batches - especially users with higher number of different recipients. The average number of messages within one mixing window of 100 messages went easily over 5. The reason seems to be a use of off-line emailing. We believe that it is another aspect of users' behavior whose effects on statistical attacks are worth a further research.
- Even though the assumptions of uniform message distribution by Serjantov et al [15] do not hold in real-world traffic, it is very easy to create a general traffic profile that can be successfully used for automated statistical disclosure attacks.
- Several experiments we have conducted indicate that the false positives start appearing much faster than false negatives. In other words, the attacker can substantially limit the set of recipients with a high probability that actual recipients will be picked.

We have run several simple experiments to verify results for statistical disclosure attack by Danezis [3] (later extended by Dingledine et al. [11]). The basic equation estimating the amount of data for a successful attack against "Alice" is (95% confidence):

$$t > \left[2.m\left(\sqrt{\frac{N-1}{N^2}(b-1)}+\sqrt{\frac{N-1}{N^2}(b-1)+\frac{m-1}{m^2}}\right)\right]^2$$

where N is the number of recipients, m the number of the Alice's recipients, b the size of the mix's pool, and t is the number of mix rounds that must be collected.

The simple experiments we did were using mix of the size b=100. N was than in the region of 5000, and m lower than 50. The large N and relatively small b allows us to simplify the equation:

$$t > \left[2.m\left(\sqrt{\frac{b}{N}}+\sqrt{\frac{b}{N}+\frac{m-1}{m^2}}\right)\right]^2$$

The first square root is around 0.15, while the second is 0.15 for just one recipient (which does not sound right), and then monotonically decreases from 0.72 down to 0.14 (0.22 is for 40 recipients) - the sum is therefore between 0.86 and 0.28, and the evaluation of the whole equation is in Tab. 3.

m	1	2	3	4	5	6	7	8	9	13	20
t	1	12	14	22	35	45	57	72	84	142	295

Tab. 3. Number of messages needed for a successful attack.

The experiments we have carried out suggest that the attack complexity is undervalued for low numbers of recipients, while over valued when more recipients is to be identified. The non-uniform distribution of messages also causes different anonymity levels for differently "popular" recipients. We believe that there the attack can be elaborated and made more efficient with the use of real-world data.

4.4 Open Questions

Parts of this paper's analysis elaborate on performance properties of mixes. These mixes were not applied on the data in real time but we have instead written a simulator of a mix system.

The analysis assumes that there is a mix server deployed on the SMTP server that has produced the traffic data. We have analyzed two basic constructions of mixes: with a fixed time delay and with a fixed pool size.

There are few open questions which we can't answer yet. The first is question about emails written as replies to previous emails. Deployment of a mix would change time line of this part of traffic. The significance of this problem is very hard to estimate as SMTP servers do not store ``In-reply-to" headers and it is unclear how would delay of the responses influence results of our analysis.

Another question is relevance of data. Traffic from university may suffer from oscillations like holidays, weekends, and so on. Obviously, it would be slightly more interesting to analyze data about email traffic of common users but it is very hard to get these. On the other hand, we have been lucky in a sense, because the collected data come from non-engineering departments and the users reflect general population much better than would users with IT background.

The last question is related to the validity of the data set because of an incredibly high number of users who sent very few messages. This was a surprise for us as well and we have verified this on a data set covering a fourteen months' period. The distribution of messages among users was invariant.

5. Conclusions

The paper presents some interesting properties related to behavior of users of electronic mail. We believe that although the analyzed data set is not final, it can help inform how to set several parameters of anonymity systems for email traffic. We have also explored how variation in message inter-arrival times and sizes impact delays for threshold mixes and variation of anonymity provided by timed mixes. We anticipated there would be substantial differences in data traffic throughout a day or a week. We show that variation exists but the differences are quite comparable with random variations that appear regardless on typical user behavior (especially with smaller thresholds and shorter intervals for collecting messages).

Acknowledgements

This research was supported by the Research Plan No. MSM, 0021630528 -- Security-Oriented Research in Information Technology.

References

[1] BERTHOLD, O., FEDERRATH, H., KOPSELL, S. Web MIXes: A System for anonymous and unobservable internet access. In *Workshop on Design Issues in Anonymity and Unobservability*, 2000, p. 115--129.

[2] CHAUM, D. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of ACM*, 1981, vol. 24, no. 2, p. 84-88.

[3] DANEZIS, G. Stastistical disclosure attacks: Traffic confirmation in open environments. In *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, Gritzalis, Vimercati, Samarati, and Katsikas (Eds.). Athens: Kluwer, May 2003, p. 421 - 426.

[4] DANEZIS, G. The traffic analysis of continuous-time mixes. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2004)*. Springer-Verlag, LNCS 3424, 2004.

[5] DANEZIS, G., DINGLEDINE, R., MATHEWSON, N. Mixminion: Design of a type III anonymous remailer protocol. In *Proceedings of 2003 Symposium on Security and Privacy*. IEEE Computer Society, May 11-14, 2003, p. 2-15.

[6] DÍAZ, C., SASSAMAN, L., DEWITTE, E. Comparison between two practical mix designs. In *Computer Security (ESORICS'04)*, P. Samarati et al. (Eds.), Springer-Verlag, LNCS 3193, p. 141 - 159, 2004.

[7] DÍAZ, C., SEYS, S., CLAESSENS, J., PRENEEL, B. Towards measuring anonymity. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, 2002.

[8] DINGLEDINE, R., MATHEWSON, N., SYVERSON, P. F. Tor: The second-generation onion router. In *USENIX Security Symposium*, 2004, p. 303--320.

[9] FREEDMAN, M. J., MORRIS, R. A peer-to-peer anonymizing network layer. In *SIGSAC: 9th ACM Conference on Computer and Communications Security*. ACM SIGSAC, 2002.

[10] KESDOGAN, D., EGNER, J., BUSCHKES, R. *Stop-and-Go MIXes Providing Probabilistic Anonymity in an Open System*. In Springer-Verlag, LNCS 1525, p. 83--98, 1998.

[11] MATHEWSON, N., DINGLEDINE, R. Practical traffic analysis: extending and resisting statistical disclosure. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2004)*, LNCS 3424, 2004.

[12] NEWMAN, R. E., MOSKOWITZ, I. S., SYVERSON, P., SERJANTOV, A. Metrics for traffic analysis prevention. In *Proceedings of Privacy Enhancing Technologies workshop (PET 2003)*. Springer-Verlag, LNCS 2760, 2003.

[13] PAXSON, V., FLOYD, S. Wide-Area Traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking,* 1995, vol. 3, no. 3, p. 226 - 244.

[14] RACKOFF, C., SIMON, D. R. Cryptographic defense against traffic analysis. In *Proceedings of {ACM} Symposium on Theory of Computing*, 1993, p. 672 - 681.

[15] SERJANTOV, A., DANEZIS, G. Towards an information theoretic metric for anonymity. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, 2002.

[16] SERJANTOV, A., DINGLEDINE, R., SYVERSON, P., PETITCOLAS, F. A. P. From a Trickle to a Flood: Active Attacks on Several Mix Types. In *Information Hiding 2002 (IH 2002)*. Springer-Verlag, LNCS 2578, 2002, p. 36 - 52.

About Authors ...

Kamil MALINKA was born in Valtice. He received his M.Sc. from Masaryk University in 2005. His research interests include security of anonymity systems, biometric authentication etc.

Petr HANÁČEK is an associate professor at the Faculty of Information Technology, Brno University of Technology. He concerns with information system security, risk analysis, applied cryptography, and electronic payment systems for more than ten years.

Dan CVRČEK is an associate professor at the Faculty of Information Technology, Brno University of Technology. He has been a researcher at the University of Cambridge, UK and focused on IT security issues related to reputation, privacy, cryptographic protocols, and wireless sensor networks.