Public Security: Simulations Need to Replace Conventional Wisdom

Kay Hamacher
Bioinformatics and Theoretical Biology Group
Technische Universität Darmstadt
kontakt@kay-hamacher.de

Stefan Katzenbeisser
Security Engineering Group
Technische Universität Darmstadt
skatzenbeisser@acm.org

ABSTRACT

Is more always better? Is conventional wisdom always the right guideline in the development of security policies that have large opportunity costs? Is the evaluation of security measures after their introduction the best way? In the past, these questions were frequently left unasked before the introduction of many public security measures. In this paper we put forward the new paradigm that agent-based simulations are an effective and most likely the only sustainable way for the evaluation of public security measures in a complex environment. As a case-study we provide a critical assessment of the power of Telecommunications Data Retention (TDR), which was introduced in most European countries, despite its huge impact on privacy. Up to now it is unknown whether TDR has any benefits in the identification of terrorist dark nets in the period before an attack. The results of our agent-based simulations suggest, contrary to conventional wisdom, that the current practice of acquiring more data may not necessarily yield higher identification rates.

Categories and Subject Descriptors: K.4.m [**Computing Milieux**]: Computers and Society—*Miscellaneous*

General Terms: Security

Keywords: simulations, data retention, privacy

The worst thing is to rush into action before the consequences have been properly debated.
–Perikles

1. INTRODUCTION

Conventional wisdom suggests that more is better. Arguably this assumption has led to major changes in the perception of anonymity, privacy, and access to data in the public and private spheres. In particular this became most evident in the aftermath of the 9/11 attacks: several data retention projects were introduced around the world in order to track terrorist activities. The Advance Passenger Information System (APIS) of the US Transportation Security Administration stores passenger name records of all airline travelers entering and leaving the US; the Terrorist Finance Tracking Program (TFTP) of the Federal Bureau of Investigation tracks financial transactions of all kinds; agencies were granted access to the records of financial transactions processed by the Society for Worldwide Interbank Financial Telecommunication (SWIFT); telecommunication data retention laws have been passed in many European states following a EU commission directive, requiring telecommunication providers to store and make available to law enforcement and other agencies call detail records (CDRs) of all attempted and successful communication events made over their infrastructure. These programs enable large-scale data analysis in order to search for patterns of current and future criminal and terrorist activities.

Despite their potentially huge privacy impact, the practical utility of most of the above-mentioned programs has rarely been critically assessed. If evaluations have been performed at all, they were done *after* the introduction of such programs, based on empirical data of the past. Conceptually, all these programs have large opportunity costs and therefore it is most desirable to judge on their effectiveness already before introduction.

In this paper we put forward the paradigm that public security policies, in particular ones that have a huge impact on privacy, should be *critically and experimentally assessed before their introduction through simulations*. In particular, for the pro-active analysis of the impact of security measures on complex dynamic systems, we propose to use agent-based simulations.

Historically, in biology [6], sociology [17], economics [9], and econophysics [11] agent-based simulations [10, 18] haven been very successful. The insight gained in this way turned out to be profoundly different from investigations on static data (such as traditional social network analysis) and from approaches of functional modeling via aggregates and average cases. Simulations outperform empirical studies in the sense that they allow to control external conditions, that are not realized *now* but might come into existence *in the future*. Therefore, they are the only sustainable evaluation paradigm to achieve *predictive power* on the efficiency and reliability of public security measures.

To illustrate the usefulness of the new paradigm, we apply as a case-study agent-based simulation to quantify the ability of Telecommunication Data Retention (see Section 2) to identify "terrorist cells" in a controlled environment of

NSPW'11, September 12–15, 2011, Marin County, CA, USA.

realistic communication dynamics. The results are indeed counter-intuitive, as they suggest that the naive paradigm of "more is better" can be violated, and that sometimes smaller amount of used data can be more informative.

2. TELECOMMUNICATIONS DATA RETENTION

Laws on Telecommunication Data Retention (TDR) have been passed in many European states over the past years in order to implement an EU directive [4]. Under this legislation, any telecommunication provider is required to keep a log of metadata of all calls made using its infrastructure. For example, for calls performed over a mobile phone, providers are required to log the phone numbers of caller and callee, as well the time of the call, its duration and the geographical location of the caller. This data is supposed to be stored for a period between sixth months and two years and can be accessed by law enforcement agencies. Initially targeted to combat organized crime and international terrorism, TDR logs soon became utilized in the prosecution of minor offenses [3], a tendency recently considered in violation of the German constitution by the Federal Constitutional Court of Germany [12]. At the same time European TDR regulations were ruled as unconstitutional by the Romanian constitutional court [5] and by the constitutional court of the Czech Republic [26].

The mandatory introduction of data retention of the communication behavior of the European population immediately sparked a lively discussion on its privacy impact. Supporters claim that it is an essential tool to fight crime and prevent future terroristic acts. Opponents doubt the usefulness of the data for terror prevention and point to the massive erosion of privacy of each citizen: the stored TDR data allows to generate detailed profiles of the "digital life" (or even movement profiles in case of mobile communication) of almost every European citizen [16].

Even though data retention went into effect in Europe in 2007 and was widely implemented in national legislation soon thereafter, there are up to now no scientific studies publicly available that evaluate the practical usefulness of TDR to combat crime. In this context, one has to distinguish between a *pre incident* and a *post incident* analysis: in the former, TDR records are used to predict future crimes or uncover yet unknown terrorist plans, while in the latter case TDR records are used as evidence in the prosecution of (past) crimes. Juridical practice shows that TDR may have positive effects if used in a *post incident* analysis: for example, a recent statement issued by the German government to the Federal Constitutional Court of Germany lists cases in which judges relied on TDR data as evidence [3], even though it remains open whether the cases could have been supported by other types of evidence. However, up to now it is still unknown whether automatic analysis of TDR records is useful in the *pre incident* case. Yet, this application has been used as a key argument in the political debate supporting the introduction of TDR legislation.

In this paper we apply agent-based simulations in order to answer the question whether automatic analysis of TDR records helps to uncover criminal activities and terrorist plots, such as revealing a terrorist cell in a large network of communicating users. Since there are no large-scale real-world call records available for a scientific study, we set up the simulation in such a way that both the network topology and the call behavior of agents conform to statistical rules that are well-known from real-world telecommunication networks (this process is described in detail in Section 3). We add small cells of terrorists with known topology to the network. Subsequently, we investigate in Section 4 a set of measures, which assess the communication behavior of an agent based on its TDR logs and have the potential to distinguish ordinary users from terrorists. In particular, we found that a distinctive set of measures exists, that (at least in theory) can distinguish the communication behavior of terrorists and ordinary users. Some of these measures are stable under implementation details and different community sizes. Our results indicate that the fervid discussed length of the retention period has indeed influence on the usefulness of TDR.

3. SIMULATION METHODOLOGY

To empirically study the applicability of TDR in a *pre incident* analysis scenario, we perform agent-based simulations of networks of communicating users (called 'ordinary users' in the sequel), which we amend with small networks of 'terrorists'. During the simulation, we keep a log file of all communication events of all agents (that is, both terrorists and ordinary users) in the same way as telecommunication providers implement TDR.

The collected data is interpreted in order to answer the following questions:

1. What is the overall information that can be extracted from TDR records in an *ex post* analysis, which knows the division of agents into a set of ordinary users and a set of terrorists? In other words, how do the communication behavior and the implied, transient communication networks of ordinary users differ from that of terrorists?

2. What are suitable measures to distinguish ordinary users from terrorists in an automated way?

3. Are these measures stable under changing environmental conditions?

When designing the simulation study, one faces several problems: First, although there exists ample knowledge on aggregate communication *statistics*, no *individualized* empirical data on communication patterns is publicly available. Second, while the structures of terrorist networks from the last decade were successfully reconstructed, data on the *individual* communication behavior of real terrorists is not comprehensively available. Third, conceptually, there are myriad analysis procedures that can be applied to TDR data records in order to answer the above questions; in particular, their performance may differ substantially and a precise description is often not available.

In order to tackle these problems, we chose the following simulation strategy, which tries to match reality as closely as possible, given the limited public knowledge of empirical data.

Simulation approach. Our simulation is agent-based, where each ordinary user and terrorist is modeled as an agent, who autonomously decides on its activities. Before the simulation starts, we fix the topology of the communication

network in order to reflect well-known properties of communication networks. In particular, we construct a topology graph, consisting of nodes for each agent and edges indicating possible communication partners. The topology we use for the network of ordinary users is described below. To this network we add a small terrorist cell with a topology that conforms to a real-world terror network.

On the resulting network, we perform a large number of simulation steps, where every step corresponds to one communication event. We simulated an overall period of three consecutive years for each set-up. Each agent is autonomously allowed to choose its communication partner based on a randomized strategy (which, however, must conform to the chosen topology). Furthermore, he can individually determine the duration of the communication event (or the amount of information that is transmitted) and the waiting time until his next communication event takes place. The choice of these temporal characteristics is described in detail below. Subsequently, the agent waits until the next event needs to be performed.

The simulation itself records all communication events performed by all agents in order to establish a database of Call Data Records (CDRs) in accordance with the EU directive on TDR [4]. The database contains the identity of caller and recipient, the time of the call and its duration.

Topology of communication networks. While there is no empirical data on the CDR of large communities available for public research, the topological behavior of large communication networks has been studied extensively [14, 28, 23]. For example, it is known that many communication networks show "scale-free" behavior: the fraction of participants (nodes in the topology graph) who have k communication partners (i.e., have node degree k) follows a power law $k^{-\alpha}$ for some parameter $\alpha > 0$; this behaviour has been found in e-mail [14], landline phone [28, 24] and mobile phone networks [23]. The landline communication network was found to have an α of about 2.1, while mobile communication networks exhibit a significantly larger parameter $\alpha > 7.0$.[1] Furthermore, networks usually show an assortative behavior [21, 23]: well-connected nodes tend to connect to other well-connected nodes. That is, the average number of communication partners of a node of degree k is an increasing function of k.

In order to realistically simulate a large telecommunication network, we generated several random topology graphs of 50,000 to 1,000,000 communicating agents using the GraphTool[2] software for different values of α. All graphs were generated to closely resemble the empirical results described by [23]. In particular, the random growth strategy employed by GraphTool resulted in scale-free networks with an assortative nature.

Temporal characteristics. Similarly to the topology of networks, some knowledge on the temporal characteristics of communication networks is available in the public literature. Most importantly, the time t between subsequent communication events (e-mails or phone calls) is known to follow an exponential distribution $\mu e^{-\mu \cdot t}$, as described in [28, 29] for the case of a phone network. In our simulation, agents follow this rule.

The amount of data transferred in a single communication refers to the duration of a telephone call or the size of an Internet-based communication event. In a phone network we assume, without loss of generality, that the amount of data transfer is a monotonous function of the call duration. In our simulation, the amount of information conveyed within a communication event is randomly drawn from a log-normal distribution. This immediately leads to a higher information transfer of well-connected members of the community—which is the central, coordinating characteristic of "leaders" (or communication hubs) in civil and terrorist projects. Note that we did not model diurnal effects in the communication.

The temporal and topological properties of the communication network depend on each other. For example, [23] reports that the average total information transfer between two entities in a network (i.e., the sum of the information transferred in all communication events) is almost constant in the product of the node degrees of sender and recipient [22]. Our simulation satisfies this condition as well by implementing a special randomized strategy for the selection of communication partners.

Topology of terrorist network. The topology of terrorist networks has been intensively studied in the scientific literature. For example, [20, 19] depict the topology of several dark networks [31], among them the network of the 9/11 attacks or the Bali bombings. In our simulation, we use the topology of the networks described in [20] to model terrorist communication networks. Even though one may argue that these graphs are incomplete, they resemble the most precise data on the structure of "real" terrorist networks that are available in the public literature.

We merged the terrorist cell with the network of ordinary users by connecting terrorist agents with other users by random links in order to simulate contacts of each member of the terror network to the "outside", reflecting that terrorists need to behave insuspiciously and to do normal business in every-day life.[3]

Temporal characteristics of terrorist traffic. Naturally there is no precise knowledge on the temporal characteristics (for example call and inter-call duration) of communication events between members of the terrorist network publicly available. Therefore we take the route of a worst-case scenario for the analyst: we assume in the simulation that the basic statistical models of temporal characteristics of the terrorist network are similar to that of ordinary users. More precisely, we assume a universal mean of this distribution, but allow for a broader or denser width of the distribution for the terrorist network to account for a different dynamics as a precursor of a hypothetical attack (e.g. the need for more communication to set up the attack or a potential "radio silence" beforehand). Technically, we model this effect by rescaling the standard deviation of the information amount transmitted within each communication event by a factor f. During the simulation, we varied f in the range of

[1] Even though some of the studies are quite old, we expect that the shapes of the distributions are stable, while the parameters vary.

[2] `http://projects.skewed.de/graph-tool/`

[3] A different approach would have been labeling some nodes of the random network as terrorist agents; however, this approach does not easily allow the use of known topologies of terrorist networks.

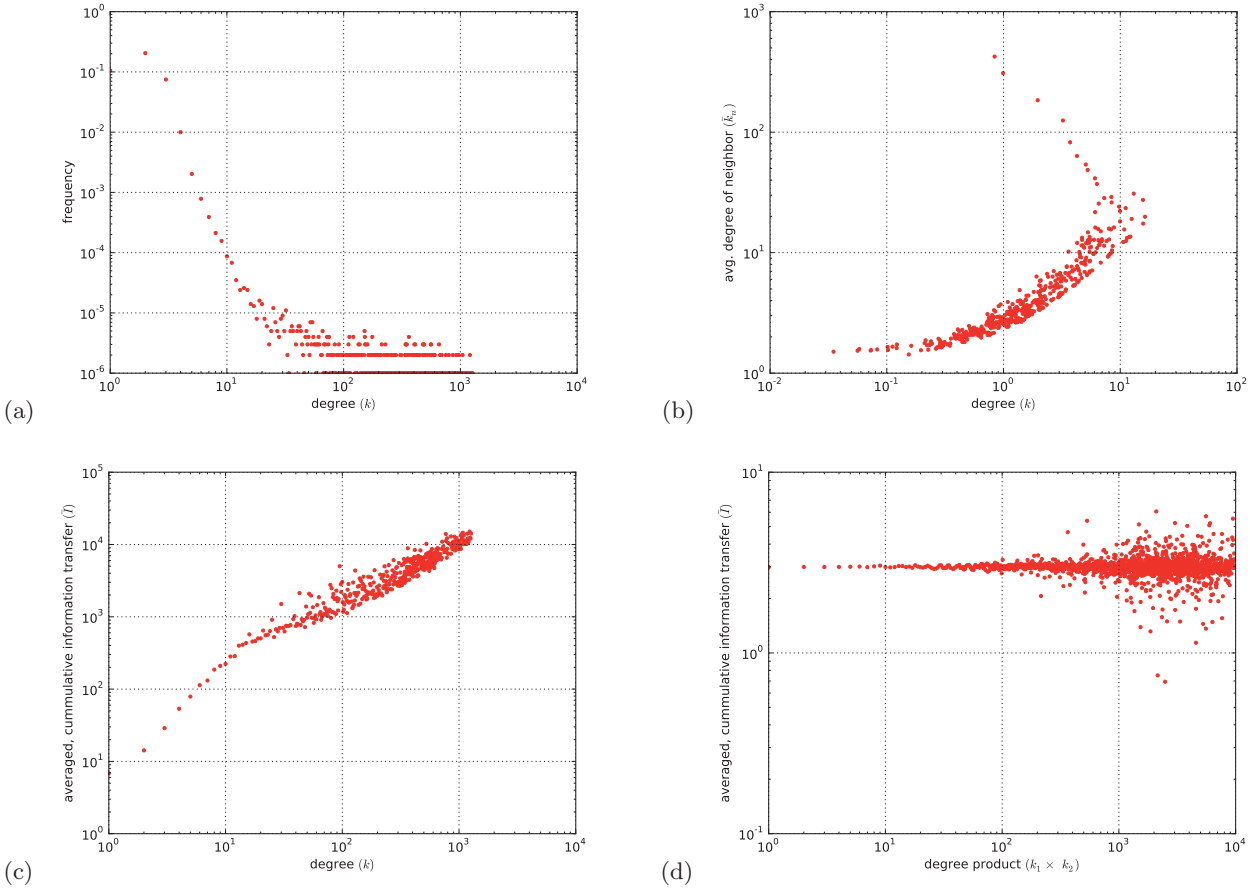

Figure 1: Statistical properties of the simulated CDR and topological features of the resulting transient communication networks; (a) degree distribution of the network; (b) distribution of the average neighbor degree with respect to the node degrees; (c) total information transfer with respect to the node degree; (d) average total information transfer depending on the product of the degrees of the communication partners.

0.6 to 1.4 and analyzed the influence of f on the stability of the proposed measures.

Validation of the simulation. In order to validate the above-mentioned simulation strategy, we compare the statistical properties of the resulting communication network to those of real communication networks reported in the open literature [14, 28, 23].

Figure 1 shows some statistics of a representative simulation run. In particular, Figure 1(a) shows that the simulated network is scale-free, i.e., the node degrees follow a power law, as expected. Figure 1(b) shows that the network is indeed assortative: highly connected nodes tend to connect to other highly connected nodes. That is, the average node degree of the neighbors of an agent increases depending on its own node degree. Both figures show that the networks of our setup closely resemble the empirical results of Onella et al. (both figures are very similar to Figures 4 and 7 of [23]). Figure 1(c) shows the total (cumulative) information transfer, depending on the node degree of the sending user; as expected, highly connected users tend to communicate more than low connected users. Finally, Figure 1(d) shows the total (cumulative) information sent between two specific users, depending on the product of the degrees of sender and recipient. The graph shows independence, which is exactly the behavior observed empirically by Onella et al. (compare Figure 10 of [23]).

Discussion. Due to the unavailability of empirical data at the level of individuals we took a "sampling route": we created communities of individuals in a way to fulfill the known statistical properties at the macroscopic level. Arguably, this does not fully capture reality; to cope with such a shortcoming, we decided to sample over several incarnations and analyze always the worst case for terrorists in these instances. Fortunately, detailed knowledge is available on the structure of darknets. Therefore, our modeling approach is empirically fully supported for them. For the integration of darknets into society and the temporal characteristics of terrorist traffic we were not able to obtain empirical data. Therefore, we modeled a worst-case scenario for the analyst: a) the terrorists' communications behavior is similar to average persons, but might only vary in intensity, and b) the connection of the terrorists to the outside world is governed by the same statistics as for normal citizens.

While the simulation results presented in Section 5 are influenced by the underlying model, we want to stress that, when more empirical data is made available, the simulation

can easily be adapted and the implication of this study may be revisited.

4. MEASURES

At the moment, analysis of dark networks is predominantly performed by manual, non-automated approaches using tools of data visualization[4], which allow to visualize the communication patterns, link it with external information and compute simple properties of the topology graph. Automated methods are—to the best of our knowledge—not reported in the open literature. Thus, one task was to identify measures which potentially allow to distinguish terrorists from ordinary users in an automated manner based on CDRs.

For the sake of notation, we let a communication event be denoted as $\epsilon = (t, i, j, d)$, where t denotes the initialization time of the communication event, i and j denote identities of sender and recipient, while d denotes the call duration (i.e., the amount of transmitted information). For a given communication event ϵ we will denote its components as $\epsilon[t]$, $\epsilon[i]$, $\epsilon[j]$ and $\epsilon[d]$. The total set of communication events will be called $\mathcal{T}$, while $\mathcal{T}|_{a\rightarrow} := \{(t', i', j', d') \in \mathcal{T} \,|\, i' = a\}$ denotes the set of all communication events that agent a initiates. As abbreviations, we write $IA_a := \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} \epsilon[d]$ for the total information amount an agent a is sending in all its communication events, while $calls_a := |\{\epsilon \,|\, \epsilon \in \mathcal{T}|_{a\rightarrow}\}|$ denotes the total number of calls and $k_a := |\{\epsilon[j] \,|\, \epsilon \in \mathcal{T}|_{a\rightarrow}\}|$ the total number of communication partners of agent a. Note, that this is an empirically measured number and does not necessarily conform to the degree of a in the static topology graph, typically discussed when all data is available. Rather, in our case it depends highly on the retention period.

Let $\epsilon \in \mathcal{T}$ be a communication event between sender $\epsilon[i]$ and recipient $\epsilon[j]$. We denote with $next_\epsilon$ the subsequent communication where $\epsilon[j]$ acts as sender. We further denote with τ the length of the retention period (in days) and with C_a the clustering coefficient of a node a. For a Boolean variable B we write $\langle B \rangle$ to denote the function $\langle B \rangle = 1$ if B is true and $\langle B \rangle = 0$ otherwise. Figure 2 lists the set of measures (having the potential to distinguish between terrorists and ordinary users) that we investigate in our study. Each measure assigns a single score to each user, based on the simulated CDR data. We distinguish between three types of measures:

- *Measures that depend only on the topological data or on the global communication behavior of a single agent.* Besides simple quantities (such as the average number of calls or communication partners per day) we consider the clustering coefficient of the node (which describes the connectedness of the neighbors of a single agent), possibly scaled by the average number of calls or communication partners. One may expect, due to the command structure of most terrorist networks, a larger connectedness of the cells, which results both in a larger clustering coefficient and higher communication behavior.

- *Measures that depend on all communication events of an agent.* In contrast to the above measures, which only depend on "global" properties of an agent, one can define measures that depend on each individual communication event performed during the retention period. For example, we may compute a sum of the information amounts transferred during all communication event of an agent, weighted by either the clustering coefficient or the node degree of sender or recipient. This way, messages sent by (or received by) highly connected agents get pronounced, which potentially again reflect the command structure within the network.

- *Measures that reflect the interdependencies between communication events.* Since some communication events (for example ones that contain commands) may trigger other communication events by peers, it may be worthwhile to trace pre-cursors of chains of communication by appropriate measures. To this end, we make the (probably not always realistic) assumption that an agent who received a command in a chain of events relays that command to one of his subordinates, without engaging in any other distracting communication event. Thus, given a communication event between agents a and b, it is worthwhile to study the statistical properties of the next communication event that b initiates with an agent c. For example, one can compute covariances between the clustering coefficients or node degrees of a, b and c or measure the total flow of information of a that is relayed via b to c. We furthermore consider a measure that quantifies the flow of information between a and c, conditioned over the event that b reports less information to c than received from a; this event occurs in a chain of command, if one agent receives its orders and assigns small sub-tasks to all its subordinates.

Theoretically, one can define much more involved measures, which combine properties of sequences of dependent communication events. However, measures correlating more and more persons or events are necessarily less significant due to statistics: the overall number of targeted persons (terrorists) eventually determines the maximal number of correlated entities. In a setting with just a few terrorists, a rule-of-thumb argument on statistical fluctuations increasing with the square root of the number of entities [27] prohibits the usage of more than two dependent communication events in a measure.

5. RESULTS AND ANALYSIS

We performed the simulations, as described in the last section, for varying parameter settings: in particular, we considered networks of sizes 50,000 to 1,000,000 agents, and varied the parameter f in the range of $[0.6, 1.4]$. Furthermore, we varied the data retention period to span $[7, 14, 30, 182, 365]$ days (i.e., only communication events within the respective retention period were used in the analysis). Unless otherwise mentioned, we show results averaged over a) these parameters, b) three different terrorist networks, and c) at least two independent runs for each of these incarnations. The rationale behind this procedure is the situation law enforcement officials face: as nothing is known on the existence, the size, and the behavior of potential terrorist cells, one needs to include (and therefore average over) a large number of potential threats. At the same time TDR

[4]Well-known software packages include Cytoscape [25], Gephi [7], Pajek [8], Sentinel Visualizer [1], and SocNetV [2]. See also [30].

Measures involving global properties of a single agent:

Average number of calls per day

$S_a = calls_a/\tau$

Average number of communication partners per day

$P_a = k_a/\tau$

Clustering coefficient C_a of agent

Clustering coefficient of agent, divided by the total number of outgoing calls

$C_a/calls_a$

Clustering coefficient divided by the total information amount transferred

C_a/IA_a

Measures involving all communication events of an agent:

Information transfer of agent, weighted by degree of recipient

$\kappa_a = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} \epsilon[d] k_{\epsilon[j]}$

Weighted Information transfer κ_a divided by total information transfer

κ_a/IA_a

Information transfer of agent, weighted by clustering coefficient of sender

$\gamma_a^{(1)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} \epsilon[d] C_{\epsilon[i]}$

Weighted Information transfer $\gamma_a^{(1)}$ divided by total information transfer

$\gamma_a^{(1)}/IA_a$

Information transfer of agent, weighted by clustering coefficient of recipient

$\gamma_a^{(2)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} \epsilon[d] C_{\epsilon[j]}$

Weighted Information transfer $\gamma_a^{(2)}$ divided by total information transfer

$\gamma_a^{(2)}/IA_a$

Measures involving properties of three communicating agents:

Covariance between the node degree of the receiver and the node degree of the receiver of the subsequent dependent communication event

$\mu_a^{(1)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} k_{\epsilon[j]} k_{next_\epsilon[j]}$

Covariance between the node degree of the sender and the node degree of the receiver of the subsequent dependent communication event

$\mu_a^{(2)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} k_a k_{next_\epsilon[j]}$

Covariance between the clustering coefficient of the receiver and the node degree of the receiver of the subsequent dependent communication event

$\nu_a^{(1)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} C_{\epsilon[j]} k_{next_\epsilon[j]}$

Covariance between the degree distribution of the receiver and the clustering coefficient of the receiver of the subsequent dependent communication event

$\nu_a^{(2)} = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} k_{\epsilon[j]} C_{next_\epsilon[j]}$

Sum of the information amount transferred by the subsequent dependent communication events of the receiver

$\alpha_a = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} next_\epsilon[d]$

Same as above, conditioned on the event that this information amount is smaller than the information originally transmitted, normalized by information amount

$\sigma_a/IA_a = \sum_{\epsilon \in \mathcal{T}|_{a\rightarrow}} next_\epsilon[d] \left\langle \epsilon[d] > next_\epsilon[d] \right\rangle / IA_a$

Figure 2: Collection of measures used to characterize all communication dynamics in the various simulations.

procedures need to be standardized, reproducible, and applicable to various communities and countries.

Identifying and selecting suitable measures. Given the set of measures defined in Section 4, we first identified those measures that are able to discriminate between terrorists and ordinary users. To this end, we needed to rank the measures. A suitable "quality function" is the mutual information $I(p, m)$ between an indicator variable $p \in \{0, 1\}$ for the agents' character (terrorist or ordinary user) and the particular measure m. The mutual information thus quantifies the amount of information an observer gains about the character p of a person, when the measure m is extracted from CDR databases. For law enforcement purposes in a *pre incident* analysis a high mutual information (MI) is obviously desirable.

All measures described above are real-valued and defined on a continuous domain. To derive required histograms of observables for the MI computation, we binned these values always into 10 bins. We ensured that this detail does not influence our overall results.

As the range and the occupancy of histogram bins for any two measures was found to vary considerably, the absolute value of the mutual information is not adequate to rank different measures against each other. In order to put all measures on an equal footing, we need to normalize by the entropy $H(p)$ of p and the entropy $H(m)$ of the particular measure m. Since for each run the entropy $H(p)$ stays constant and will therefore not influence the ranking of a measure, we therefore will deal with the normalized mutual information $I(p, m)/H(m)$ from here on.

Figure 3 shows the results in the form of bean plots [15], i.e., empirical probability distributions, where the bar indicates the mean value. Obviously, there are two distinguished sets of measures: one with an larger normalized MI, that therefore lend itself to distinguishing terrorists from normal users (σ normalized by the information amount, κ normalized by information amount, $\gamma^{(2)}$, the clustering coefficient of the agent, and $\gamma^{(1)}$ and $\gamma^{(2)}$, both normalized by the information amount). The other measures yield a rather low normalized MI and are thus not significant. Comparing sub-figures (a) and (b) and the bean plots for the different population sizes shows that the identification procedure is rather stable under varying environmental (community size) and regulatory conditions (retention period).

Confirming the identified and selected measures. To confirm the results, we ran an orthogonal analysis: for each measure m, we define the empirical distribution p_T over the set of all terrorists and over the set of all ordinary users p_O and compute the Kullback-Leibler divergences $D(p_T \,\|\, p_O)$. Important measures, i.e. ones which can discriminate best between terrorists and ordinary users, need to show high Kullback-Leibler divergence (DKL) values, according to the Chernoff-Stein-Lemma of hypothesis testing [13]. Figure 4 shows the results of this analysis; measures, which were identified as important by the first analysis method, are shaded in gray. The five measures with the highest DKL indeed correspond to the same measures found by the first analysis method, only $\gamma^{(2)}$ yields a lower DKL.

Stability of relevant measures. Ideally, any suitable measure to distinguish between ordinary users and terrorists is stable with respect to varying external conditions of their usage. For example, a good measure needs to be applicable to communication dynamics of any community size and any terrorists' cell size. Also small variations in the terrorists' behavior should not influence the overall significance of a detection mechanism. To illustrate this stability, we show in Figure 5 the results for varying community sizes and varying terrorists' behavior (modeled via f) for some of the important measures, identified above. It is evident that for community sizes larger than a small city these measures do not vary too much. For very small communities (less than 10^5 persons) there is some sensitivity. Such community sizes are, however, not the proposed scenario of TDR. Note, that the factor f also does not influence the value of the normalized MI to a large extend. We observed the same behavior for the other important measures (graphs are omitted due to space restrictions).

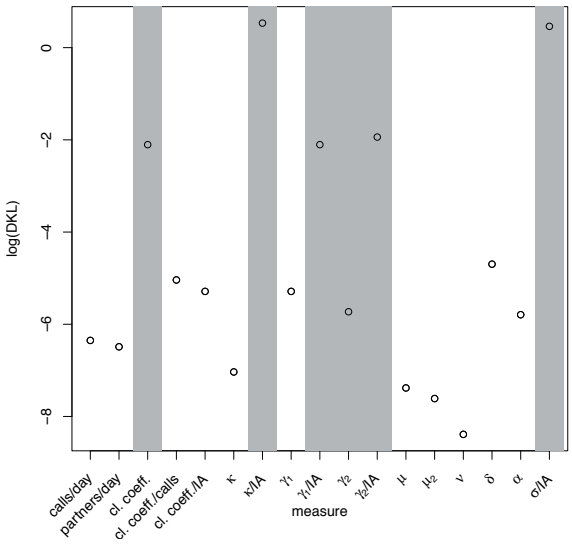

Figure 4: Kullback-Leibler divergence (DKL) between empirical histograms for all measures. The DKL is computed between the histograms of measures restricted to terrorists and to ordinary agents; measures shaded in gray were identified as relevant in the first step (see Figure 3).

We have thus established a set of measures that might be applicable cross-country- and cross-jurisdiction-wise. These measures are capable of signaling important behavior and their *relative* importance—as expressed by their respective ranking in comparison to other, unsuitable measures—is maintained.

Influence of the data retention period. One central question in the context of TDR implementation is the optimal data retention period. The longer the period, the more communication events are stored and analyzed, which allows in turn to uncover a larger set of contacts for each individual agent. On the downside, this aggravates the privacy problems and potentially increases the noise level within the data, leading to a lower classification rate.

To investigate the influence of the data retention period, we show in Figure 6 the results, where we differentiate between different communication behavior of the terrorists given by the parameter f. Clearly and unsurprisingly, the information extractable from a given measure grows with

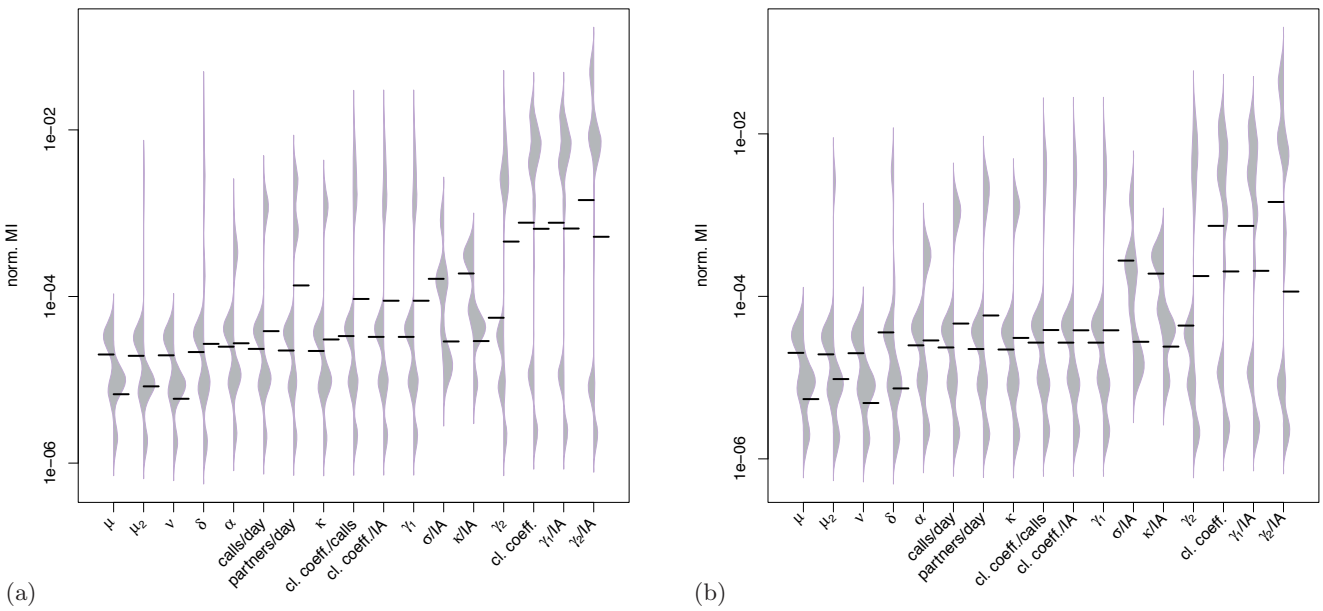

Figure 3: Bean plots of the normalized mutual information $I(p,m)/H(m)$ for all measures of Figure 2 for (a) 90 days and (b) 365 days retention period. A bean plot shows the empirical distribution function (upwards) of a measure. For each measure, the left portion of the gray shaded area shows the results for a population size of 150,000 agents, the right part for a size of 600,000 agents, thus eventually comparing these different set-ups. The black bar indicates the average of the respective distribution.

the number of days for which CDRs are available. However, for a retention period larger than 90 days our results suggest a declining quality of all predictive measures. Interestingly, this holds for all measures and all communication behavior. The terrorists actually do not have to adapt their behavior to circumvent detection, but would rather be better off in a regulatory setting with a retention period of half or a full year. The only exception is the sixth measure (σ/IA), where for $f = 1$ we found a more or less monotonous increase in the predictive power. However, here terrorists can adapt, and destroy the signal all together by switching to a modified dynamics. This comes as no surprise as σ/IA measures the flow of "hierarchical information" in a chain. If, however, terrorists start to communicate very small pieces or do the opposite (collecting information and then communicating in few, long events), these hierarchies are not as transparent as for $f = 1$.

Clearly, one could argue that—even if data is stored for a longer period of time—one can perform an analysis over smaller time periods to reduce the above-mentioned adverse effects (thereby effectively throwing away observations). However, given the huge privacy impact this is highly undesirable: the "need-to-know" principle demands that only the minimal required information should be collected and utilized.

6. CONCLUSION

In this paper we have put forward the paradigm that the effectiveness of new public security measures should be evaluated experimentally *by agent-based simulation*. Due to the complexity of society and the involved networks of human interactions, simulation is the analysis tool of choice—even allowing for predictions. A *predictive* assessment is particularly important in those cases, where steps are undertaken that induce high opportunity costs (such as gross damages to individuals' privacy), while showing unclear effectiveness.

As a case study we reported results of an agent-based simulation tailored towards answering the question whether TDR is useful in a *pre incident* analysis, where call data records are used to uncover a dark-net of terrorists embedded in a large telecommunication network of ordinary citizens. In particular, our results indicate that there is indeed a distinctive set of measures that can provide information on the character of an individual (terrorist/ordinary citizen) based on his call records. The chosen measures are stable under implementation details and environmental conditions such as community size or call durations. However, contrary to conventional wisdom, large retention periods can spoil the detection accuracy of the measures used in our study, questioning the wisdom of "the more data, the better".

Acknowledgements. We thank Tyler Moore and the NSPW attendees for their valuable feedback.

7. REFERENCES

[1] `http://www.fmsasg.com/Products/SentinelVisualizer`.
[2] `http://socnetv.sourceforge.net`.
[3] Bundesministerium der Justiz, Antwort der Bundesregierung zum Fragenkatalog des 1. Senats des Bundesverfassungsgerichts in den Verfahren 1 BvR 256/08, 1 BvR 263/08, 1BvR 586/08, `http://www.vorratsdatenspeicherung.de/images/StN_BMJ_2009-06-02.pdf`, 2009.
[4] Directive 2006/24/EC of the European Parliament and of the Council of 03/15/2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

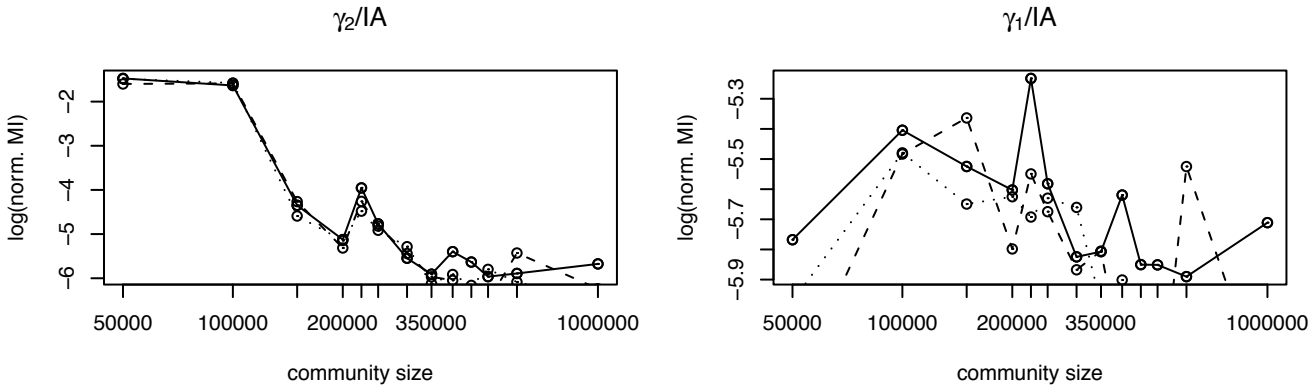

Figure 5: Dependency of the normalized MI on the community size and varying communication behavior of terrorists f. Note the small scale of variation in the right sub-figure.

[5] Curtea Constituţională a României. decision (Decizia Nr.1.258) on 10/09/2009 regarding laws no. 298/2008 and no. 506/2004.

[6] G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz. Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(2):159–171, 2009.

[7] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.

[8] V. Batagelj and A. Mrvar. Pajek - analysis and visualization of large networks. In *Graph Drawing Software*, pages 77–103. Springer, Berlin, 2003.

[9] C. Bianchi, P. Cirillo, M. Gallegati, and P. Vagliasindi. Validating and calibrating agent-based models: A case study. *Computational Economics*, 30:245–264, 2007.

[10] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Nat. Acad. Sci.*, 99(Suppl 3):7280–7287, 2002.

[11] M. Buchanan. Economics: Meltdown modelling. *Nature*, 460:680–682, 2009.

[12] Bundesverfassungsgericht. case 1 BvR 256/08 as of 2.3.2010, 1 - 345.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, Hoboken, 2. edition, 2006.

[14] H. Ebel, L-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Phys. Rev. E*, 66:035103(R), 2002.

[15] P. Kampstra. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets*, 28(1):1–9, 11 2008.

[16] C. Kurz and F. Rieger. Stellungnahme des Chaos Computer Clubs zur Vorratsdatenspeicherung, 2009. 256/08, 1 BvR 263/08 und 1 BvR 586/08 `http://www.ccc.de/updates/2009/vds-gutachten`.

[17] X. Li, W. Mao, D. Zeng, and F.-Y. Wang. Agent-based social simulation and modeling in social computing. In C. Yang, H. Chen, M. Chau, K. Chang, S.-D. Lang, P. Chen, R. Hsieh, D. Zeng, F.-Y. Wang, K. Carley, W. Mao, and J. Zhan, editors, *Intelligence and Security Informatics*, volume 5075 of *Lecture Notes in Computer Science*, pages 401–412. Springer, Berlin, Heidelberg, 2008.

[18] C.M. Macal and M.J. North. Tutorial on agent-based modeling and simulation. In *Proc. 37th Conf. Winter Sim.*, WSC '05, pages 2–15, 2005.

[19] N. Memon, D. Hicks, D. Muhammad Akbar Hussain, and H. L. Larsen. Practical algorithms and mathematical models for destabilizing terrorist networks. In *IEEE Military Communications Conference (MILCOM 2007), Proceedings*, pages 1–7. IEEE Press, 2007.

[20] N. Memon, K. C. Kristoffersen, D. L. Hicks, and H. L. Larsen. Detecting critical regions in covert networks: A case study of 9/11 terrorists network. In *Proc. 2nd Int. Conf. on Availability, Reliability and Security, ARES 2007*, pages 861–870. IEEE Comp. Soc., 2007.

[21] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(30):208701–1, 2002.

[22] J. Onella, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertéz, and A. Barabási. Structure and tie strengths in mobile communication networks. *Proc. Nat. Acad. Sci.*, 104(18):7332–7336, 2007.

[23] J. Onella, J. Saramäki, J. Hyvönen, G. Szabó, M. Menezes, K. Kaski, A. Barabási, and J. Kertéz. Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys.*, 9:179, 2007.

[24] L. Schintler, S. Gorman, A. Reggiani, R. Patuelli, and P. Nijkamp. Scale-free phenomena in communication networks: A cross-atlantic comparison. In *43rd European Congress of the Regional Science Association*, 2003.

[25] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, 2003.

[26] Ústavní soud České republiky. decision (PR 22/11) on 03/22/201.

[27] W.H. Press et al. *Numerical Recipies in C*. Cambridge University Press, Cambridge, 1995.

[28] Y. Xia, C. Tse, W. Tam, F. Lau, and M. Small. Scale-free user-network approach to telephone network traffic analysis. *Phys. Rev. E*, 72:026116–1, 2005.

[29] Y. Xia, C. K. Tse, W. M. Tam, F. C. M. Lau, and M. Small. Telephone traffic analysis based on a scale-free user network. In *International Symposium on Nonlinear Theory and Its Applications (NOLTA'05*, pages 110–113, 2005.

[30] J. Xu and H. Chen. Criminal network analysis and visualization. *Commun. ACM*, 48:100–107, June 2005.

[31] J. Xu and H. Chen. The topology of dark networks. *Commun. ACM*, 51(10):58–65, 2008.

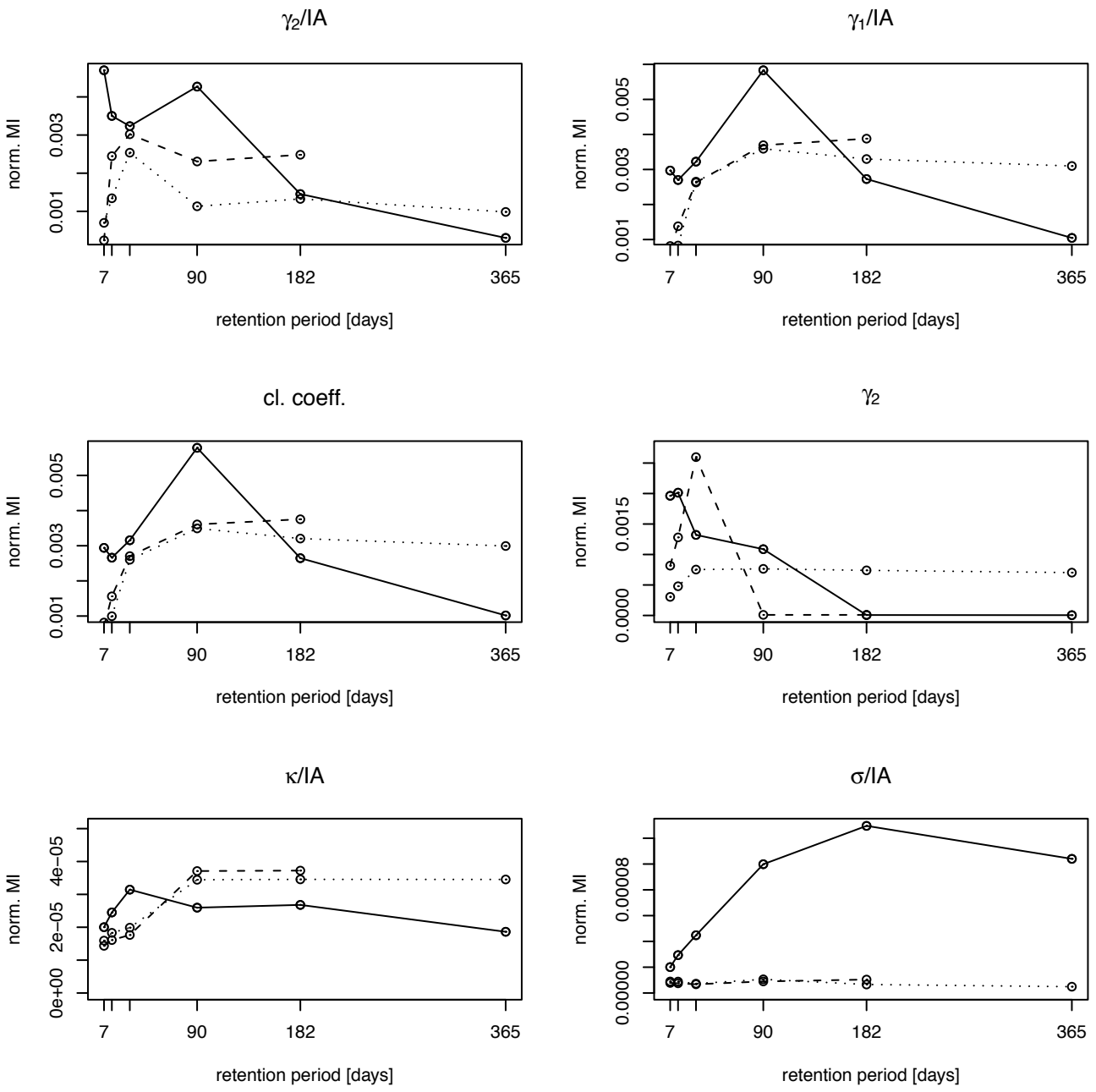

Figure 6: Normalized mutual information for all suitable measures as a function of the retention period; the individual lines identify different communication behavior of terrorists (solid lines $f = 0.6$, dashed lines $f = 1.0$, and dotted lines $f = 1.4$).