Privacy metrics and boundaries

Louis-François Pau

Rotterdam School of Management

June 2007

PRIVACY METRICS AND BOUNDARIES

ABSTRACT

This paper aims at defining a set of privacy metrics (quantitative and qualitative) in the case of the relation between a privacy protector ,and an information gatherer .The aims with such metrics are : -to allow to assess and compare different user scenarios and their differences ;for examples of scenarios see [4]; -to define a notion of privacy boundary, and design it to encompass the set of information , behaviours , actions and processes which the privacy protector can accept to expose to an information gathering under an agreement with said party ; everything outside the boundary is not acceptable and justifies not entering into the agreement ; -to characterize the contribution of privacy enhancing technologies (PET). A full case is given with the qualitative and quantitative privacy metrics determination and envelope, i.e. a Cisco Inc. privacy agreement.

ACM Categories and classification

K4.1 [Societal aspects] Privacy

ACM General terms

Security, Economics

Keywords

Privacy, Metrics, Set theory, Economics, Case, Privacy enhancing technologies

1. DEFINITIONS

The privacy protector is an individual , group or organization defending consciously a set of information , behaviours, values , processes and methods in terms of the full control and independence of these .This set is called the private information set.

The information gatherer is an individual, group, organization, machine or network , operating on the privacy protector's information set .However, in special cases , the information gatherer may also be "nature" that is the privacy protector's own environment .

The information gathering operations may be:
-either the result of an explicit agreement between the two above parties
-or the result of an implicit agreement between the two above parties , as linked to another type of agreement between them, such as a social or business agreement
-or may take place without the privacy protector's explicit or implicit acceptance

2. ORGANIZATIONAL AND SET THEORETICAL ASPECTS

Because of information networking effects, in real situations the privacy protector shares information fully or in part with other privacy protectors, each of those having different relations to information collectors.

Likewise, because of information networking effects, the information gatherer shares information fully or in part with other information gatherers, each of which having different relations to the privacy protectors. It is assumed that all members of

the information collector can access and share freely and equally the information gathered.

Because of organizational or decision hierarchies, a given privacy protector may be exposed to information gatherers at different levels, and vice versa as well, with different nested or linked private information sets.

The presentation below about privacy features is limited to the relation between one privacy protector and one information collector , although the nature of each can be very different .Needles to say, this relation is assymmetrical , meaning that the privacy protector is also sometimes the information gatherer , with two flows eventually existing and usually quite different .

Each privacy protector has defined a private information set, which nevertheless can be accessed and can flow freely between its members .This private information set is not restricted to static information, or dynamic information, but includes also process, sequence, tool and method information.

The information collector achieves a measure of success if it accesses (copy, transfer, delete, substitute) an element in the private information set of the privacy-collector, or achieves a change in the privacy protector's existence conditions (role, identity, risk exposure, etc...) .
The privacy features listed below are not organized by "vulnerabilities".

For a further formalism about the above, information games, as well as information warfare, offer useful frameworks.

This paper does NOT however take the approach of a cost-benefit analysis or quantitative game between the parties driven by their utility functions.

3. PRIVACY FEATURES

3.1. Copy feature

This feature exists in two instances:

-"Copy-quantitative" :how much information in the private information set is copied and available to the information collector ;if several information quantity measures exist, this feature exists for each
-"Copy-qualitative": what is the relative value, on a value scale ,of the private information copied and available to the information collector ; if several information value measures exist , this feature exists for each .

Besides, the feature can be assessed by either the privacy protector or the information collector; if there is a gap, there may be loss or a third party involved.

The different information types apply to a wide range of moral, capability, information asset, and other types.

3.2. Transfer, delete, and substitute features:

They are defined as in 3. 1. ,except that the copy process is replaced by the transfer , delete, substitute processes to the benefit of the information collector .If there is transfer, the information does not exist any more in the private information set.

3.3. Role change feature

This qualitative feature "Role-qualitative" indicates on a scale, how different the privacy protector's role in his previous context, has changed as a result of the information collection .Normally the scale go from vastly diminished contributing role in this context, to vastly enhanced .From the networking point of view it can be approximated by the number of links of influence or benefit this privacy protector gets as a result of the information collection.

3.4. Identity change feature

This qualitative feature "Identity-qualitative" indicates on a scale if the information collection has destroyed or enhanced the visibility and independence of the privacy collector in his previous context as a result of the information collection ; this includes personal preferences, lifestyle, values being disclosed

3.5. Time feature

This quantitative feature “Time-quantitative” states the relative loss or gain of time of the privacy protector in his main time-dependent actions and their success, due to the information collection taking place

3.6. Risk feature

This quantitative feature “Risk-quantitative” states the rate of reduction or increase of the privacy protector in his main risk exposed actions and their success, due to the information collection taking place

3.7. Leaking feature

This quantitative feature “Leak-quantitative” defines the perceived probability, seen from the privacy protector, which if there is information collection, the information gathered will be copied to a third party due to the information collector’s networking facilities.

4. PRIVACY ACCOUNTABILITY

This term encompasses the set and process based protocols (in an information theoretic sense) whereby the privacy features are tracked and later negotiated .In quite many cases , the information collector will grant the privacy protector some rights ,services or goods , against his agreeing to disclosing some of the information in the private information set ,plus some other compensation . In many cases such a negotiation will operate on Min(Max) or Max(Min) values with a corresponding value logic, which works both on qualitative as well as quantitative values.

Privacy assurance is then the set of processes and physical measures taken by a privacy protector to safeguard his private information set.

5. PRIVACY FEATURE VALUE DISTRIBUTION

The assumption made here is that privacy is a mental perception for which there exists distributions of the extreme values , akin hazard and risk assessment about technical systems .Extremes of privacy features are perceived , from a human perception point of view, as stress ,for which bayesian priors may exist .The idea is to estimate the probability that the privacy holder is strong enough to overcome the information collection stress .These distributions are also assumed to exist over finite ranges ,although the end points may not be defined or reproducible .It is also assumed that the ranges have a sufficient granularity to allow for qualitative or quantitative values to be assessed over these ranges.

We do not make here any explicit assumptions such as the fuzzy set membership functions used by some approaches to risk and security using fuzzy sets ,and the resulting possibility expected values Likewise we do not specify or assume any structure in the privacy feature hierarchy, such as the tree based views in safety analysis (and fault trees) ; the reason for this last assumption is that the different privacy feature do not in general exhibit between them causal relations which can be formalized in this structured way .We do not either use formal approaches to information semantics which could eventually encapsulate privacy protector understandings ,as this is very difficult to characterize without proper knowledge representation and knowledge acquisition [5].

We do acknowledge that privacy features in different user scenario may be analyzed by Monte Carlo simulation about the whole usage or information collection simulation.

Thus, we conjecture here that the distribution of the extreme values of the privacy features each obey as a first approximation a probability distribution akin many extremes [1,2,6] ,that is f(x) measures the probability that the random feature value X is larger than x :

$$f(x) = 1 - k * \exp(-a * \exp(b * x))$$

where a , b , k are normalized constants ,and an order relation for x values exist which give low

ranks to values favourable to the privacy protector, and high values to those who are not .If order is reverse, formula changes. The constants are normalized so that f(x) is a true probability distribution function, that is:

f(0) or f(low) =1
f(infinite) or f(max) = 0

where “low” and “max” are end values of qualitative ranges.

The above approximation is essentially non-parametric as it does not assume parametric or specific distributions of the feature values .Some more complex wavelet based models exist which require the estimation of more parameters, which may be difficult to do in the context of this study. If enough data can be collected , the stress-strength models for reliability with non-parametric inference from a learning sequence , can give confidence bounds on the privacy features [3]; of course more precise estimates can be obtained if a prior distribution is assumed (Weibull, Poisson) and one can also utilize the eventual covariances between different privacy features (if known) .

Some prior research [7] has taken an information theoretical approach to the same issue of defining metrics on anonimity, but suffers from the need of prior distributions of information sets for both the privacy protector and the information gatherer.

6. PRIVACY ENVELOPE

We define now the privacy envelope as the set of states of the privacy protector (and his privacy assurance processes), which:
-either are inside tolerable values for the privacy protector, set explicitely for each feature by a privacy feature threshold value;
-or, correspond to one given probability p for all privacy feature values x to be less than the distribution of the extremes, such that p=f(x); in this case p becomes a privacy tolerance measure across all features.

7. COMPUTATION AN ANALYSIS CASE

Appendix gives a full case with the full text of the terms of a service level agreement (SLA) , for a Privacy agreement proposed by Cisco to individuals seeking access to a Cisco Web site (see Figure 1) .The Appendix gives as well as an initial qualitative grading by the PrivacyHolder of the privacy attributes of this SLA. In this Section we give a full computation of the privacy envelope following the methodology and formulas of the previous Sections .We also determine quantitatively privacy equilibrium of the privacy features. Whereas the qualitative grading showed that the initial assessment of the Cisco SLA was actually quite unfavourable for the PrivacyHolder, the computation determines a better and more objective equilibrium, thus reflecting PrivacyHolder’s “privacy risk aversion” .The graphical presentation of the privacy envelope offers a visual way to show how privacy metrics cluster in favourable or unfavourable subsets to either party.

7.1. InformationGatherer (Cisco)

The InformationGatherer Cisco has three component demand functions for the service they supply commercially to third parties; the service relies on information inputs from PrivacyHolders. In general the demand is increasing with the quality of the information Cisco collects from the PrivacyHolders:
-demand grows exponentially, with a ceiling, vs. copied information quantity from PrivacyHolder
-demand decreases exponentially, with a floor, vs. role enhancement, visibility and support given to PrivacyHolder
-demand grows exponentially, with a ceiling, vs. leakage Cisco can make to third parties (includes Cisco’s customers)
Cisco’s utility function is a linear weighted sum of the above demand functions, and is to be maximized. It is assumed that the pricing is fixed and common for all customers of Cisco, thus revenue is proportional to supplied demand.

7.2. PrivacyHolder

PrivacyHolder is in this case not a commercial entity but owns know how and information assets, plus an identity and social role. For reasons of

simplification, all PrivacyHolders are assumed to behave identically.

The PrivacyHolder's information asset reserves are tapped by copy, delete, and substitute functions enabled under an SLA with Cisco, but are also accrued over time by the role enhancement of the PrivacyHolder which the SLA enables.

The information asset reserves of PrivacyHolder are a weighted linear combination of the past reserves, accrued (with the right signs) for transfer, delete , substitution and future role driven enhancements .The copy function does not deplete these assets .In the weights the maximum capacities allowed under the SLA are taken into account.

The utility for PrivacyHolder of his entering into an agreement with Cisco is proportionate to his information assets (which can be supplied in total or usually in parts) ,multiplied by the privacy utility of this supply , as measured by the privacy metrics distribution taken from the point of view of the PrivacyHolder (see Section 5) .However, as the PrivacyHolder is not acting for revenue maximisation but for privacy protection ,the privacy utility of this supply across all privacy metrics should be the largest (or eventually the median) ,but certainly not the smallest .

PrivacyHolder's utility function does not include in this case a price based component as the supply is on a free voluntary basis.

7.3. Privacy equilibrium

The determination or computation of the final equilibrium SLA driven privacy attribute values should result from a fair equilibrium between Cisco's utility function, and PrivacyHolder's utility function. The equilibrium should be computed across all privacy metric variables set by the PrivacyHolder (in this case: 13 see Table 1) .All these variables are to be determined within the bounds (upper and lower) on each attribute.
For lack of access to a MaxMin computation software with constraints, is computed here the minimum of the squared difference between the utility function values of Cisco and of the PrivacyHolder, subject to the set constraints on all privacy metric attribute values.

7.4. Privacy envelope

For each set of privacy attributes , is computed and visualized in a radial "star" form the computed thirteen privacy metrics using the formula in Section 5 (Figure 2) .The formula's parameters are however estimated and adjusted for computability.

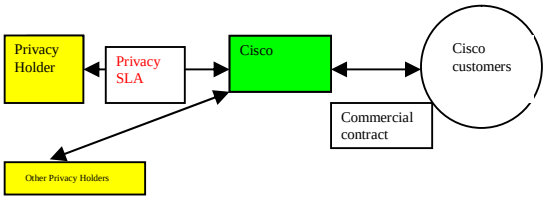

Figure 1: SLA and usage of information gathered

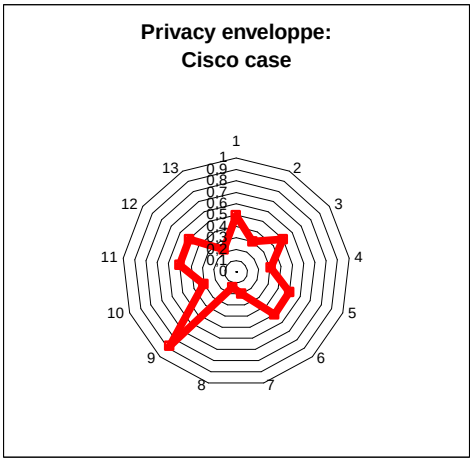

Figure 2: Privacy envelope (Cisco case)

8. CONCLUSION, CRITICISM AND ECONOMIC IMPACT

The formalism above assumes all privacy metrics to be calibrated or defined in terms of prior values or probabilities corresponding to an operating context

preceding the information collector's operations .This makes indeed absolute comparisons almost impossible, although the privacy envelopes determined by the same perceived p values may allow for comparison.
On the other hand, most humans and organizations do perceive stress in terms of reference frameworks which are subjective. Consequently, any case must specify the initial assumptions made.

The privacy ranges, and envelope, are of huge importance when implementing service level agreements (SLA's, defined in the Open Group's SLA Handbook) between the privacy protector and the information collector .This envelope, the thresholds, or the probability p, may be used in the SLA attributes to be negotiated. The economic impact of these privacy features has been researched in a companion report in the PRIME project which implements privacy SLA's with economic contract values.
Finally , this paper recognizes that privacy SLA's business value does not extend to all domains, while conversely there are application domains where their introduction would be a major incentive scheme to unleash controlled information and business interactions . To difficult areas belong health privacy (see e.g. EU IS/T project PRIDEH-GEN) ,and some security information. To the realistic areas belong telecommunications and mobile services, transport , information services , and logistics ; some calculation cases and even implementations are underway in some of these last areas .

REFERENCES

[1] Denny, Mand Gaines, S. , "Chance in biology. Using probability to explore nature", Princeton, NJ: Princeton University Press. xiii, 291 p. ,ISBN 0-691-09494-2
[2] Gumbel E.J., "Statistics of Extremes", Columbia University Press, 1958
[3] Johnson, R.A., " Stress-strength models for reliability", in P.R.Krisnaiah, C.R.Rao (Eds) "Handbook of statistics", Vol 7 , 1988, 27-54, North Holland ,Amsterdam
[4] Kumagai J., and , Cherry,S., "Sensor nation" :Sensors and sensibility, IEEE Spectrum, Vol 41, no 7, July 2004, 18-40
[5] Pau L-F , "Survey of expert systems for risk assessment, test generation , and maintenance", in : A.Kandel, E. Avni (Eds) , "Engineering risk and hazard assessment" , Vol II, CRC Press, Boca Raton,FL, ISBN 0-8493-4657-6, pp117-135
[6] "Privacy policy Colorado" , http://sciencepolicy.colorado.edu/admin/publication_files/resourse-14-2003.37.pdf
[7] Danezis, Andrei Serjantov ; George , Towards an information theoretic metric for anonymity, in :Paul Syverson , Roger Dingledine, (Eds),and "Privacy Enhancing Technologies - PET '02". Springer Verlag, LNCS, 2002., URL: http://petworkshop.org/2004/talks/rump/serjantov1.ppt

ACKNOWLEDGMENTS

This paper is part of the publicly approved reviewed deliverables of the European Union's Framework 6 project "PRIME" on Privacy enhancing technologies (Economics subtask) www.prime-project.eu.org

APPENDIX : A CASE : CISCO's PRIVACY POLICY

This is an illustration of the values a privacy protector exposed to this "privacy policy" could set , facing the information collection process by Cisco which this protector may have agreed to if he has agreed upon the terms of this service level agreement (SLA)(Tables 1 and 2) .The text of the Cisco policy is reproduced below ((C) Cisco Inc) (Table 3)

In the qualitative attribute range, "Average"/"Same" stands for the conditions existing prior to entering into an agreement .In the quantitative attribute range, values depend on individual's client configuration .

The distributions of extremes is not reported here ,although it can easily be derived from the privacy feature values (and range specifications) below- A

simple colour coding approximation has been done, that is :
-orange:if privacy feature value is at threshold for privacy protector
-green : ””””””””””””””””””””””””””””””””” is better than threshold “”””””””””””””””
-red : “””””””””””””””””””””””””””””””””””” exceeds threshold “”””””””””””””””””””””””””

It can be seen that ,by and large, and yet to be confirmed by a single aggregate probability of reaching extreme, the relationship between the privacy protector and the information collector formalized by the agreement underneath, is not favourable to privacy protector .

Table 1 : INITIAL ASSUMPTIONS AND PRIOR's

CONTEXT	PRIVACY PROTECTOR	INFORMATION GATHERER
Identity	Individual in his work environment and for his learning	Cisco
Explicit agreement	Individual releases some basic information on identity,roles, responsibilities information needs, IT system, use of Cisco systems , against promise by Cisco to send a newsletter and give access to a Cisco product site via a login/password	Cisco collects information about individual and his professional needs and environment, as well as about the Web traffic generated by individual
Implicit agreement	Cisco may have an agreement with individual's employer or IT systems platform operator	Cisco has agreements with unnamed third parties
Non authorized agreement	N/A	N/A

TABLE 2: Privacy feature ranges , thresholds and values for Case (Privacy protector's view)

PRIVACY FEATURE	RANGE/UNITS	ORDER (Increased risk/Diminishing risk)	VALUE	THRESHOLD	VALUE vs. THRESHOLD
Copy-quantitative	0-100 MB	I	10 k	30 k	
Copy-qualitative	Useless-Minor-Average-Critical(individual)-Business critical(employer)	I	Critical (individual)	Average (no anonymous browsing possible)	
Transfer-quantitative	0-100 MB	I	0	0	
Transfer-qualitative	Useless-Minor-Average-Major-Business critical(individual)-Business critical(employer)	I	Major , due to unspecified nature of cookies and attached trojan horses, and unspecified use	Average	

			of log files		
Delete-quantitative	0-100 MB	I	0	0	
Delete-qualitative	Useless-Minor-Average-Major-Business critical(individual)-Business critical(employer)	I	Average	Average	
Substitute-quantitative	0-100 MB	I	10 MB downloadable information files or executable files; no protection against disruptions	0	
Substitute-qualitative	Useless-Minor-Average-Major-Business critical(individual)-Business critical(employer)	I	Business critical(employer)	Average	
Role change	Much reduced-Reduced-Same-Enhanced-Much enhanced/new role	D	Same as individual's contributions are not recognized by Cisco	Enhanced	
Identity change	Destroyed-reduced-Same-Enhanced-Significantly strengthened	D	Reduced, as personal usage and preferences are disclosed and logged ,to access any self-selected information	Same	
Time	Blocked-Reduced-Same-Enhanced-Competitive advantage	D	Enhanced, due to timely access to information maybe not available otherwise	Enhanced	
Risk	Significantly larger-Larger-Same-Reduced-Significantly less	D	Reduced	Reduced	
Leakage	Reduced-Same-Increased-Major leakage	I	Increased , due to Cisco controlled transfer to partners and others ,as well as unspecified data content of cookies	Same	

Privacy Statement

Table 3 : CITATION (C) Cisco Inc

Cisco Systems, Inc. Online Privacy Statement

Cisco respects your privacy and is committed to protect the personal information that you share with us. Generally, you can browse through our website without giving us any information about yourself. When we do need your personal information to provide services that you request or when you choose to provide us with your personal information, this policy describes how we collect and use your personal information.

Information Collection

Personal information means any information that may be used to identify an individual, including, but not limited to, a first and last name, email address, a home, postal or other physical address, other contact information, title, birth date, gender, occupation, industry, personal interests, and other information when needed to provide a service you requested.

When you browse our website, you do so anonymously, unless you have previously indicated that you wish Cisco to remember your login and password. We don't automatically collect personal information, including your email address. We do log your IP address (the Internet address of your computer) to give us an idea of which part of our website you visit and how long you spend there. But we do not link your IP address to any personal information unless you have logged in to our website. Like many other commercial websites, the Cisco website may use a standard technology called a "cookie" to collect information about how you use the site. Please go to "Cookies and Tracking Information" below for more information.

Cisco collects personal information when you register with Cisco for a Cisco account, when you use certain Cisco products or services, when you register to attend a seminar or participate in an online survey, when you ask to be included in an email or other mailing list, or you submit an entry for a sweepstakes or other promotions, or when you submit your personal information to Cisco for any other reason. From time to time, Cisco receives personal information from business partners and vendors. Cisco only uses such information if it has been collected in accordance with acceptable privacy practices consistent with this Policy and applicable laws.

Access to certain Cisco web pages require a login and a password. The use of those web pages, and the information or programs downloadable from those sites, may be governed by a written agreement between your employer and Cisco. Unless you request deletion of your personal information as specified below, your personal information may be retained by Cisco to verify compliance with the agreement, log software licenses granted, to track software downloaded from those pages, or track usage of other applications available on those pages.

Notice

When personal information is collected, we will inform you at the point of collection the purpose for the collection. Cisco will not transfer your personal information to third parties without your consent, except under the limited conditions described under the discussion entitled "Information Sharing and Disclosure" below.

> If you choose to provide us with your personal information, we may only transfer that information, within Cisco or to Cisco's third party service providers with your permission. Upon receiving your permission, we may transfer your information across borders and from your country or jurisdiction to other countries or jurisdictions around the world.

We will always give you the opportunity to "opt out" of receiving direct marketing or market research information. This means we assume you have given us your consent to collect and use your information in accordance with this Policy unless you take affirmative action to indicate that you do not consent, for instance by clicking or checking the appropriate option or box at the point of

collection. In some cases, when applicable, we will provide you with the opportunity to "opt in." This means we will require your affirmative action to indicate your consent before we use your information for purposes other than the purpose for which it was submitted.

Cookies and Tracking Technology

A cookie is a small data file that certain Web sites write to your hard drive when you visit them. A cookie file can contain information such as a user ID that the site uses to track the pages you've visited, but the only personal information a cookie can contain is information you supply yourself. A cookie can't read data off your hard disk or read cookie files created by other sites. Some parts of Cisco's website use cookies to track user traffic patterns. We do this in order to determine the usefulness of our website information to our users and to see how effective our navigational structure is in helping users reach that information.

If you prefer not to receive cookies while browsing our website, you can set your browser to warn you before accepting cookies and refuse the cookie when your browser alerts you to its presence. You can also refuse all cookies by turning them off in your browser, although you may not be able to take full advantage of Cisco's website if you do so. In particular, you may be required to accept cookies in order to complete certain actions on our website. You do not need to have cookies turned on, however, to use/navigate through many parts of our website, except access to certain of Cisco's web pages may require a login and password.

How We Use Information Collected

Cisco uses information for several general purposes: to fulfill your requests for certain products and services, to personalize your experience on our website, to keep you up to date on the latest product announcements, software updates, special offers or other information we think you'd like to hear about either from us or from our business partners, and to better understand your needs and provide you with better services. We may also use your information to send you, or to have our business partners send you, direct marketing information or contact you for market research.

Information Sharing and Disclosure

Because Cisco is a global company, your personal information may be shared with other Cisco offices or subsidiaries around the world. All such entities are governed by this Privacy Policy or are bound by the appropriate confidentiality and data transfer agreements.

Your personal information is never shared outside Cisco without your permission, except under conditions explained below. Inside Cisco, data is stored in controlled servers with limited access. Your information may be stored and processed in the United States or any other country where Cisco, its subsidiaries, affiliates or agents are located.

Cisco may send your personal information to other companies or people under any of the following circumstances: when we have your consent to share the information; we need to share your information to provide the product or service you have requested; we need to send the information to companies who work on behalf of Cisco to provide a product or service to you (we will only provide those companies the information they need to deliver the service, and they are prohibited from using that information for any other purpose); or we want to keep you up to date on the latest product announcements, software updates, special offers or other information we think you'd like to hear about either from us or from our business partners (unless you have opted out of these types of communications). We will also disclose your personal information if required to do so by law, to enforce our Terms of Use, or in urgent circumstances, to protect personal safety, the public or our websites.

Your Ability to Review and Delete Your Account and Information

If you are a registered CCO user, you can review your personal information by accessing http://tools.cisco.com/RPF/profile/profile_management.do. You may also request deletion of your Cisco account or any of your personal information held by us by sending an email to

privacy@cisco.com.

Data Security

Your Cisco account information is password-protected for your privacy and security. Cisco safeguards the security of the data you send us with physical, electronic, and managerial procedures. In certain areas of our websites, Cisco uses industry-standard SSL-encryption to enhance the security of data transmissions. While we strive to protect your personal information, we cannot ensure the security of the information you transmit to us, and so we urge you to take every precaution to protect your personal data when you are on the Internet. Change your passwords often, use a combination of letters and numbers, and make sure you use a secure browser.

Questions or Suggestions

If you have questions or concerns about our collection, use, or disclosure of your personal information, please email us at privacy@cisco.com.