ログ情報調査を支援する視覚化システム「見えログ」

電気通信大学大学院　情報システム学研究科

高田 哲司

zetaka@vogue.is.uec.ac.jp

発表概要

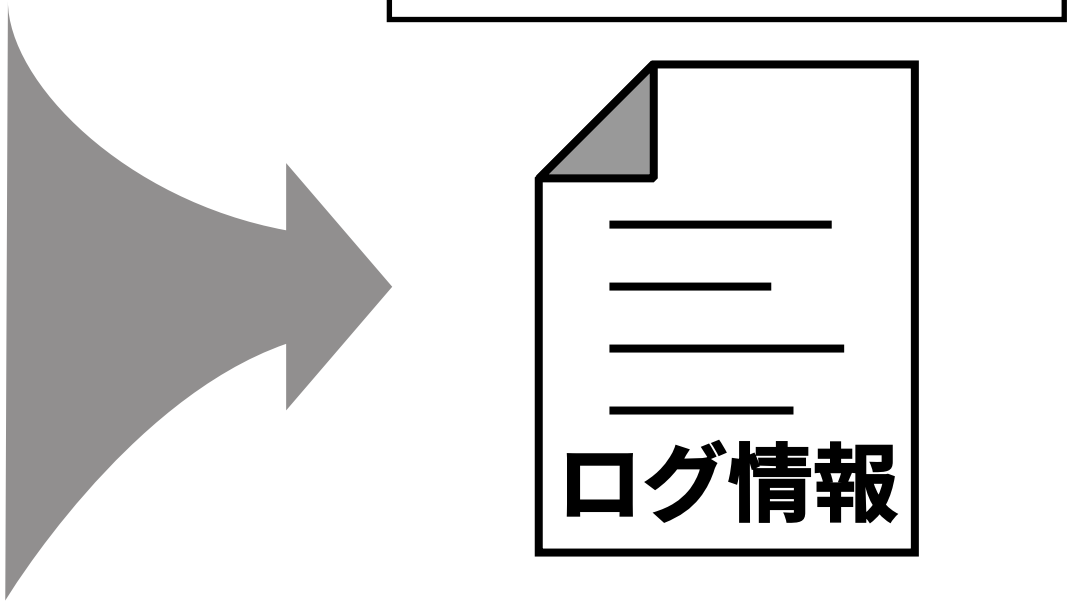
1. ログ情報調査の大切さと問題
2. 見えログ 表示法と仕組み
3. デモンストレーション
4. 質疑応答

ログ情報の大切さ

不正侵入対策

稼動状況把握

種々の異常検出



ログ情報の調査/監査は必要不可欠

ログ情報調査の問題点

なぜ作業が敬遠されるのか?

1. **文字情報**
2. **膨大な量**
3. **注目すべき情報の不明さ**
4. **多様なデータ形式と偏在性**

ログ情報調査の問題点(2)

1. **手作業**

   人間は異常判断/認識能力は優秀だが、
   長時間の単調作業に正確性を求めるのは困難

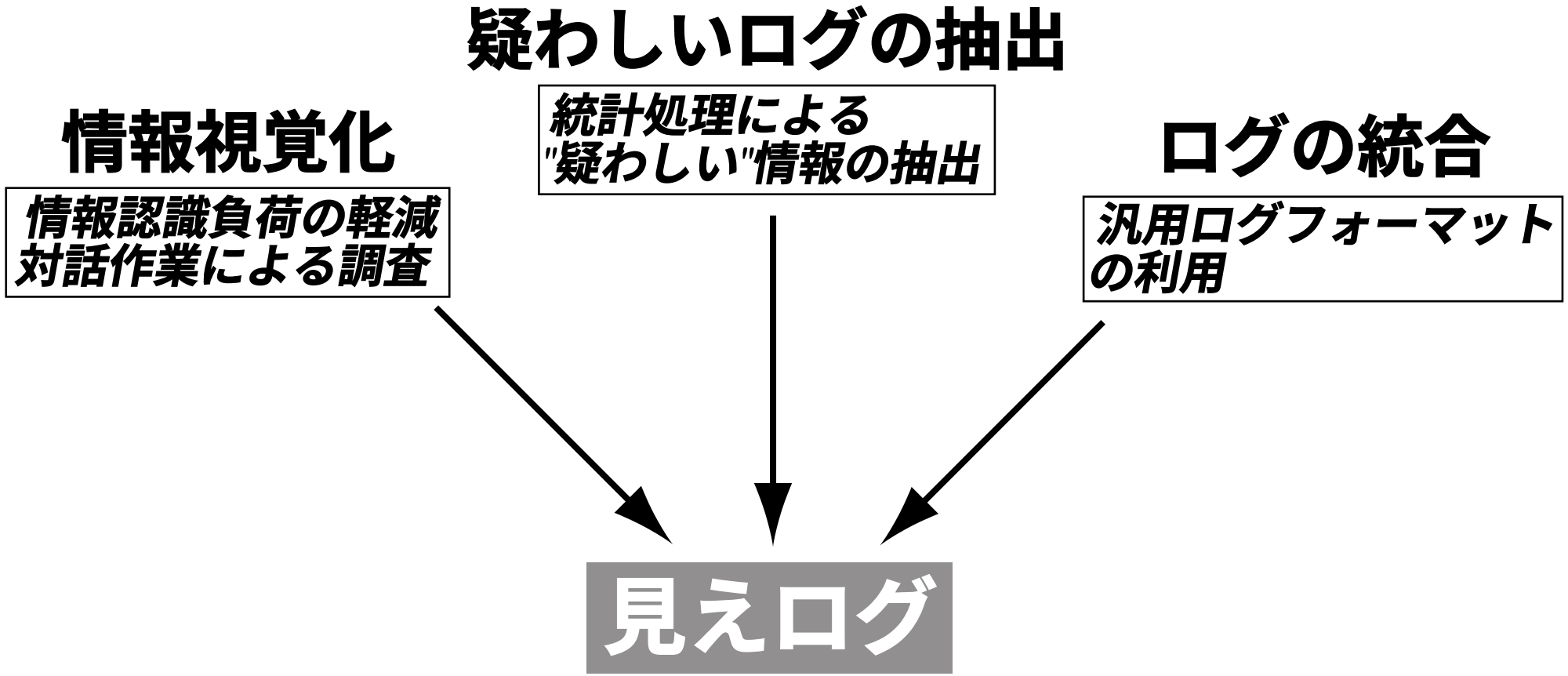
2. **自動化は難しい?**

   正規表現によるキーワード抽出等だけで
   大丈夫か?

現状は? (1)

1. やっていない (定期的に遂行 44%)
2. 地道に手作業
3. 既成/独自のスクリプトで処理
4. 既製のシステムを利用
   (swatch, syslog-ngなど)

見えログ: その特徴



ログ情報の調査を支援

情報視覚化

情報を抽象化し、図として提示

1. **情報把握の容易化、高速化**
2. **多くの情報を一度に提示可能**
3. **図との対話作業が可能**

ログ情報の統合

汎用ログフォーマットの利用

時刻	タグ情報	メッセージ

変換例）Nov 2 00:35:36 milan.ac.jp Xsession: root login

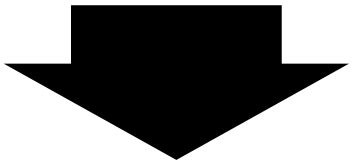
ログ情報の統合 ⇒ **調査負担の軽減**
ログ情報間の関連の明確化

疑わしいログの抽出

1. **既知のキーワードによる方法**

2. **統計処理による方法**
   仮定

疑わしい情報は大量の正常な情報内に埋もれている

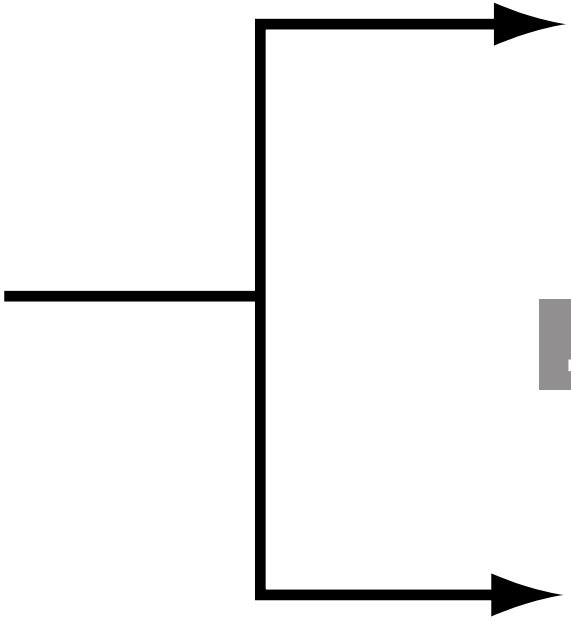


頻度情報に基づく抽出

頻度情報

ログ情報

connect from tokyo
connect from ueno
connect from tokyo
connect from tokyo
connect from ueno
connect refused ito



単語別出現頻度

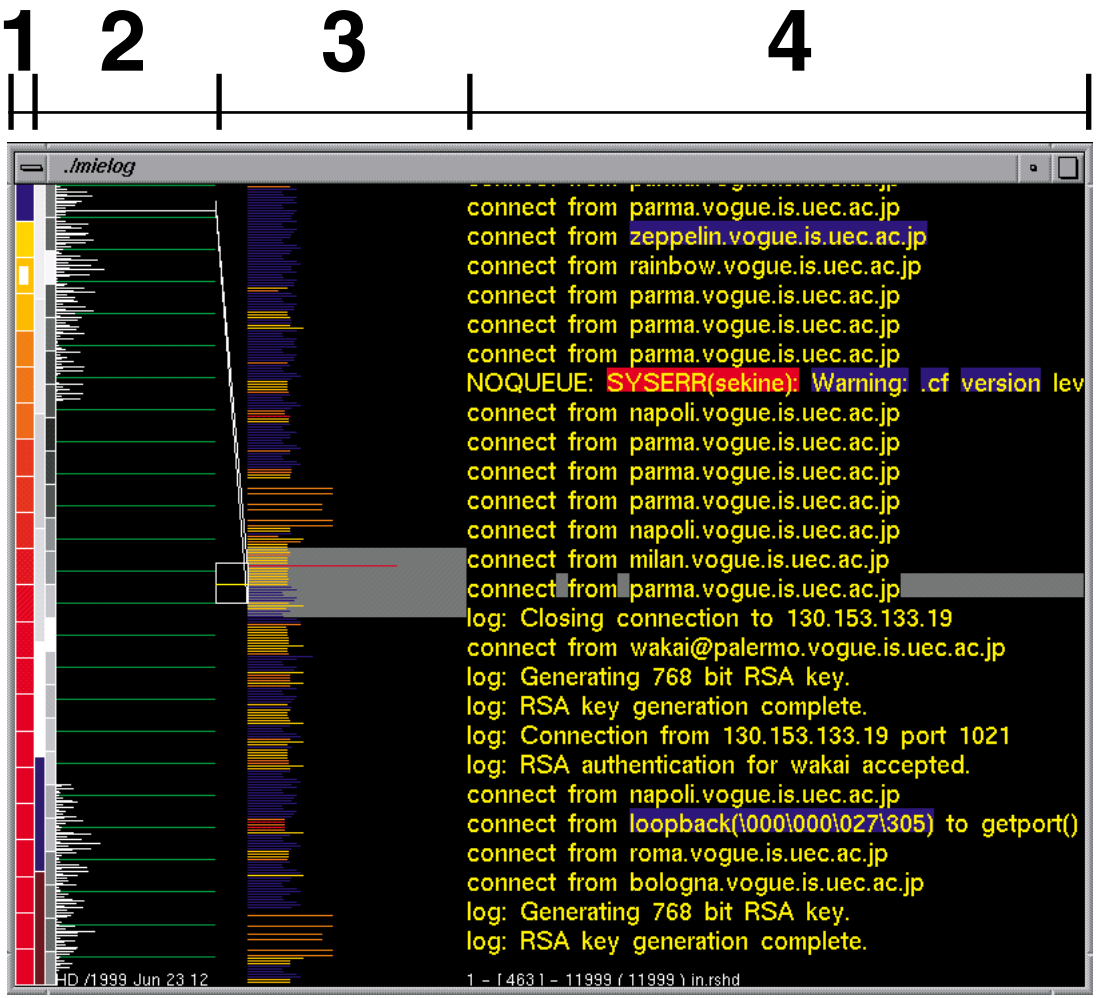
6	connect
5	from
3	tokyo
2	ueno
1	ito
1	refused

二連結単語別出現頻度

5	connect	from
3	from	tokyo
2	from	ueno
1	connect	refused
1	refused	ito

頻度情報を求めることで
疑わしい情報を抽出する

画面構成



4つの表示領域

1 タグ格子領域

2 時間別頻度領域

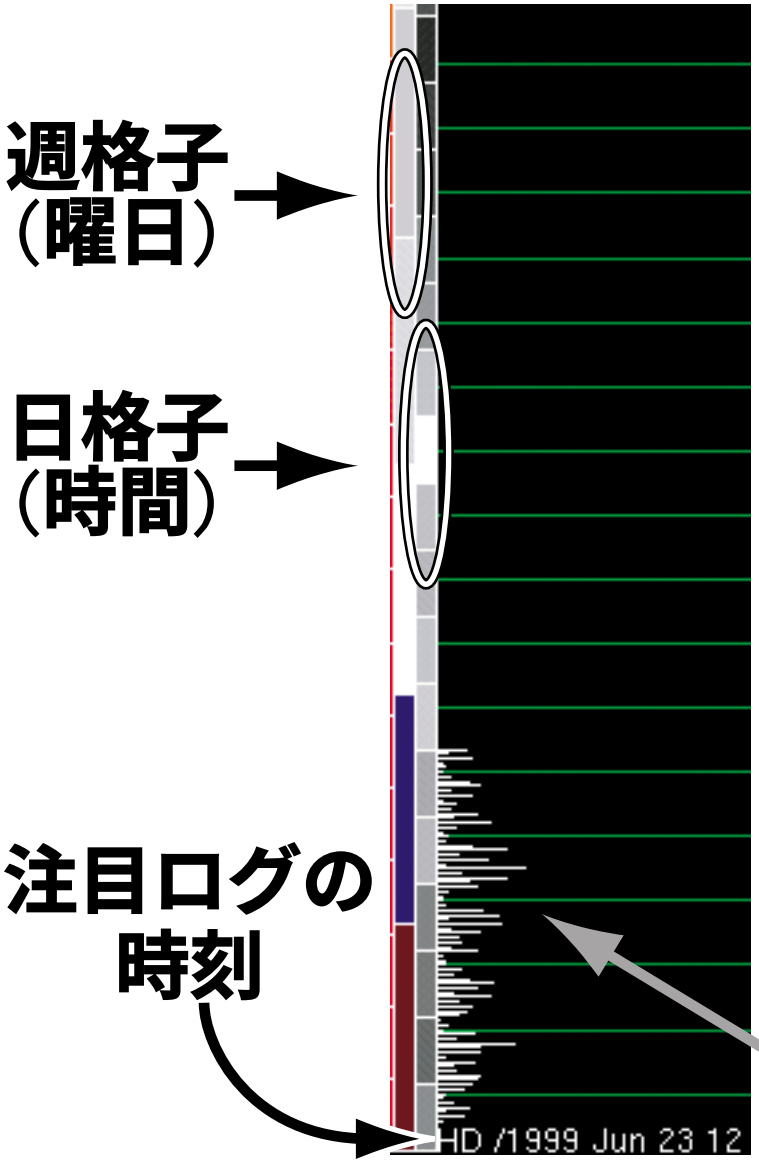
3 アウトライン領域

4 文字表示領域

タグ格子領域

- タグ情報を頻度表として格子状に表示
- 頻度値を色で表す
- タグ名は文字表示領域に表示

時間別頻度領域



ログ情報の時間分布を表示

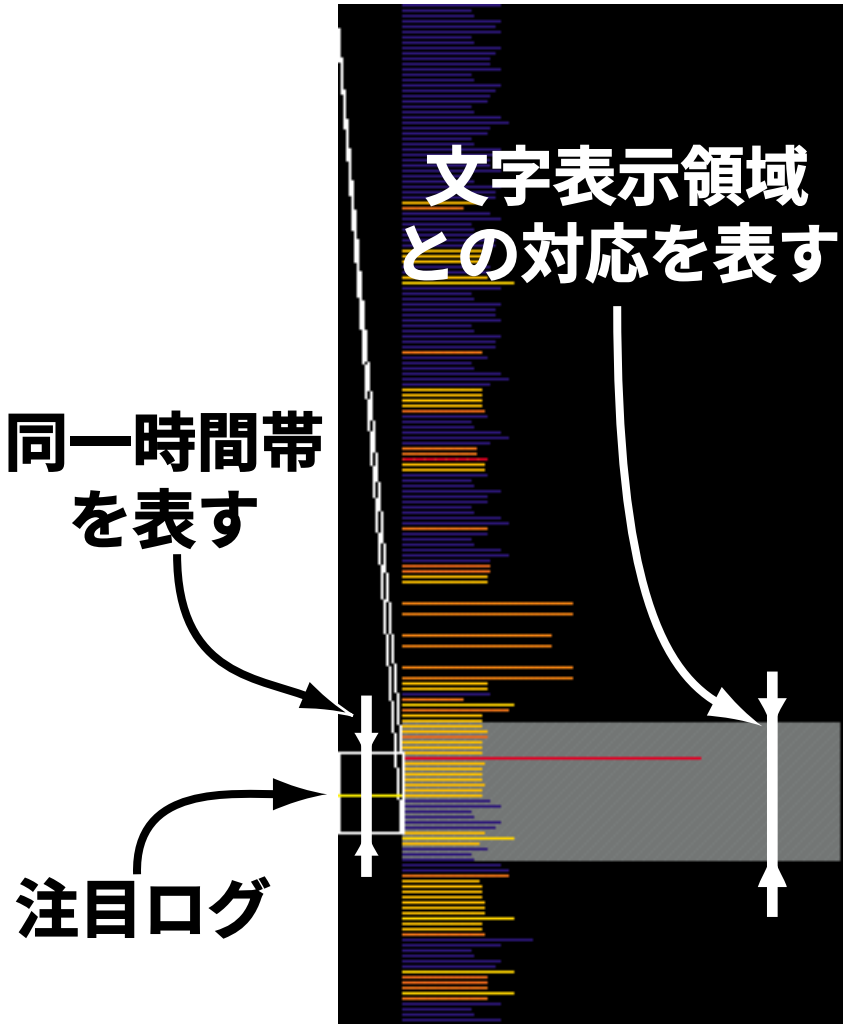
格子表示

周期的な側面(曜日/時単位)

ヒストグラム表示

時間別出力頻度数の概略

アウトライン領域



ログ情報の概略図

connect from aaa

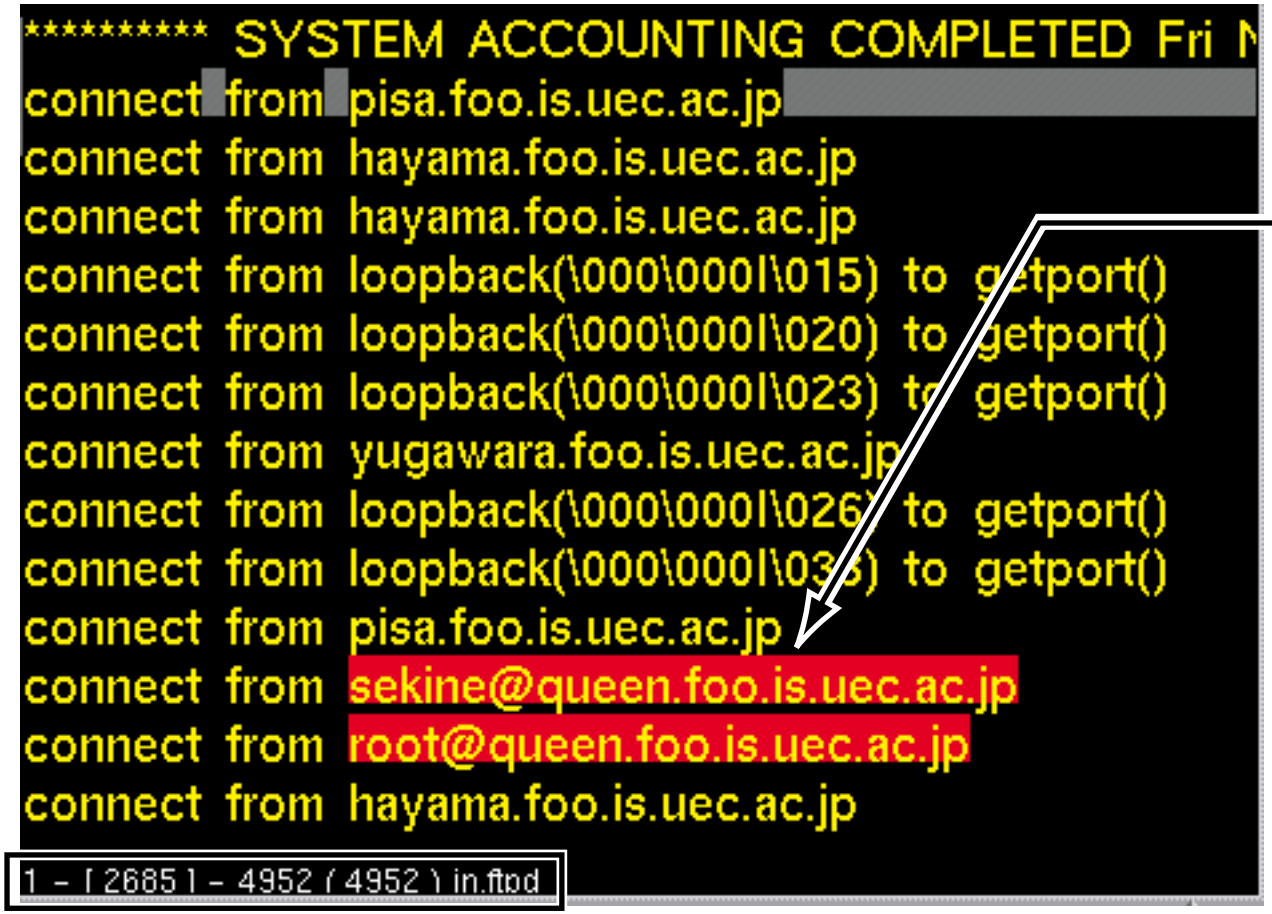
yp_all error

fatal: connection failed

- 注目ログを中心に
  その前後の概要を表示
- 色はタグ別頻度領域と同一

文字表示領域

注目ログ前後のログ情報をありのままに表示



注目ログ

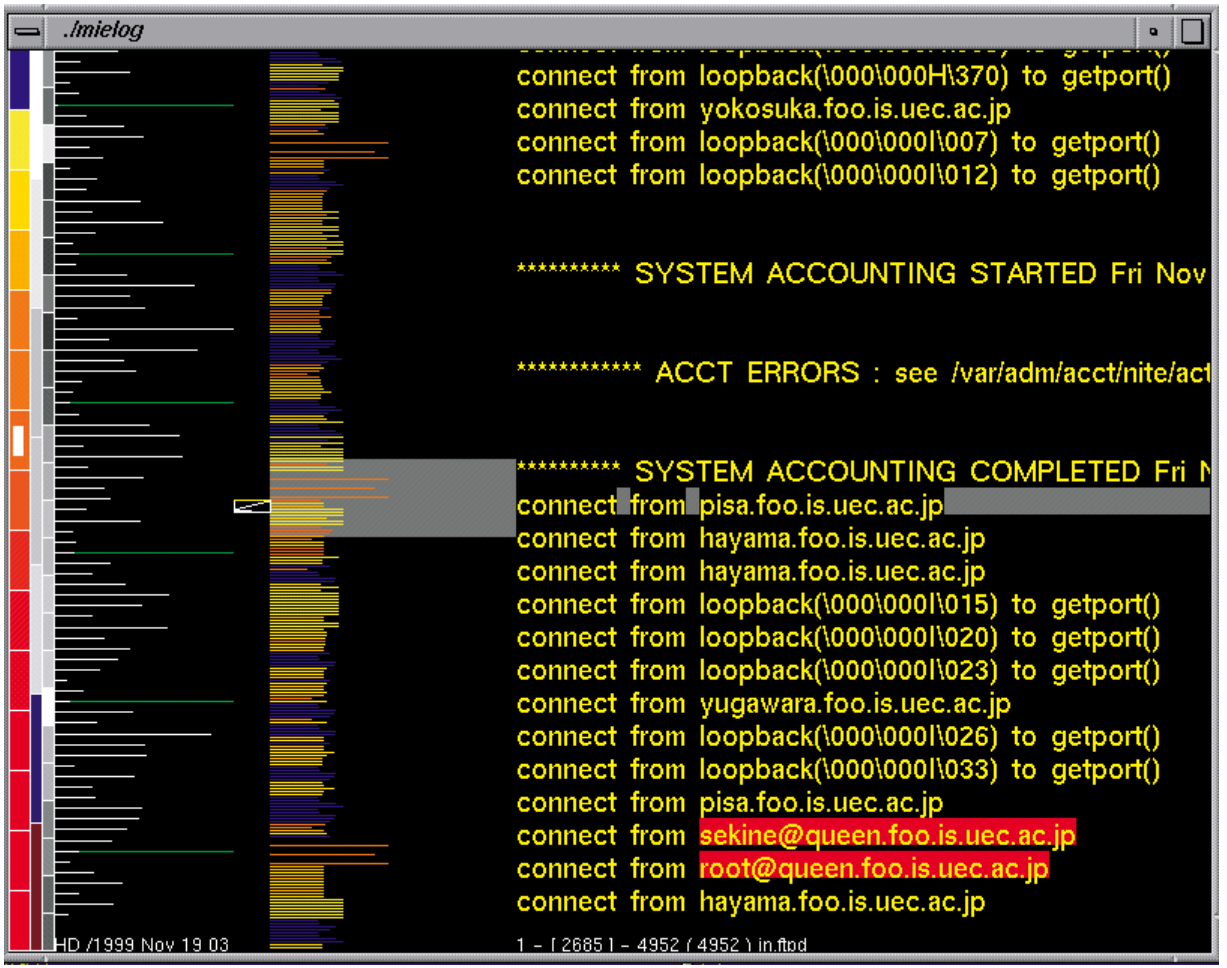
注目すべきキーワード

赤 ユーザ指定のキーワード

青 統計処理により抽出されたキーワード

注目ログの詳細情報

デモンストレーション


./mielog
connect from loopback(\000\000H\370) to getport()
connect from yokosuka.foo.is.uec.ac.jp
connect from loopback(\000\000I\007) to getport()
connect from loopback(\000\000I\012) to getport()
********** SYSTEM ACCOUNTING STARTED Fri Nov
************ ACCT ERRORS : see /var/adm/acct/nite/act
********** SYSTEM ACCOUNTING COMPLETED Fri N
connect from pisa.foo.is.uec.ac.jp
connect from hayama.foo.is.uec.ac.jp
connect from hayama.foo.is.uec.ac.jp
connect from loopback(\000\000I\015) to getport()
connect from loopback(\000\000I\020) to getport()
connect from loopback(\000\000I\023) to getport()
connect from yugawara.foo.is.uec.ac.jp
connect from loopback(\000\000I\026) to getport()
connect from loopback(\000\000I\033) to getport()
connect from pisa.foo.is.uec.ac.jp
connect from sekine@queen.foo.is.uec.ac.jp
connect from root@queen.foo.is.uec.ac.jp
connect from hayama.foo.is.uec.ac.jp
HD /1999 Nov 19 03
1 - [2685] - 4952 (4952) in.ftpd

今後の課題

疑わしいログの抽出に対する評価

「抽出された情報 = 調査すべき情報」か?

作業における自由度の高さ

定型処理の構築、処理の自動化が必要

おわりに

はたして、有用なシステムだろうか?

プロの意見を頂きたい

- 定型処理として普段行っている調査作業方法
- 図として提示すべき情報、不要な情報
- 提案、文句、賛辞?など

MieLog Web page:

http://www.vogue.is.uec.ac.jp/~zetaka/Public/kenkyu/proj/mielog.html